Journal of Research High School

Vol. 2023 (2) 1 – 5
ISSN 2688-3651 [online]

The Relationship Between Maternal Immune Activation and Autism Spectrum Disorder

Avery Erwin[1] *

[1]North Shore Country Day School, Winnetka, IL, USA
*Corresponding Author: avery.b.erwin@gmail.com

Advisor: Courtney Rivet-Noor, Crr4tz@virginia.edu

Received November 27, 2022; Revised May 21, 2023; Accepted, June 21, 2023

Abstract

Maternal immune activation (MIA) has been linked to higher incidence of autism spectrum disorder (ASD). As maternal inflammation has been associated with viral infection and environmental stressors, much research has been dedicated to studying these connections. However, even with the research dedicated to understanding the links between MIA and ASD, how inflammation causes ASD remains unclear. This review aims to summarize the research focused on understanding the links between autism and maternal inflammation during pregnancy. We will focus on defining maternal immune inflammation, understanding how it develops, and how it can lead to ASD. Finally, we will explore new research objectives for prevention and therapeutic intervention in MIA induced ASD. Bringing these ideas into a central article will highlight common themes in MIA/ASD research as well as bring new experimental avenues to the forefront of researcher's minds.

Keywords: Behavioral and Social Sciences; Neuroscience; Inflammation; Cytokines; Autism Spectrum Disorder; Maternal Immune Activation

1. Introduction

There is no current treatment for ASD, emphasizing the importance of exploring and suggesting new avenues of work for such. As more data is collected regarding ASD, we move closer to discovering an effective treatment. Accumulating evidence suggests that maternal immune activation (MIA) while pregnant puts offspring at higher risk to develop autism spectrum disorder (ASD). If MIA, defined as the reaction of the maternal immune system of a pregnant person to a pathogen, occurs during crucial stages of neurodevelopment, the offspring is more susceptible to developing a neuropsychiatric illness, such as ASD (Boulanger-Bertolus et al., 2018). MIA has been observed in various mammals ranging from humans to mice (Boulanger-Bertolus et al., 2018) (Choi et al., 2016), demonstrating the conserved nature of the phenomenon and allowing for in depth investigation into its causes. After countless studies of these MIA models, the scientific field has gained a better understanding of how MIA relates to ASD, how modern science can be used to target MIA and how to prevent ASD in the future. Here we will summarize the main theories behind MIA!s induction and it's connection to ASD.

2. Discussion

Much of the data surrounding maternal immune activation comes from human studies. MIA occurs in humans when an expecting mother is exposed to one or more pathogens (a disease-causing organism) (Boulanger-Bertolus, et al., 2018). When a pathogen is detected, the infected space becomes inflamed in an attempt to kill the identified threat. Inflammation is the body's reaction to an irritant. Inflammation usually includes redness, swelling and aching. The goal of inflammation is to eliminate all of the foreign substance (the pathogen) and to return the body to normal

Vol. 2023 (2) 1 – 5
ISSN 2688-3651 [online]

health. However, during this process, the body scans for all possible foreign substances, which could result in the fetus being accidentally targeted by a strong immune response (Boulanger-Bertolus et al., 2018). Considering the danger inflammation can pose to developing offspring, it is important to understand the types of exposures that can lead to MIA.

One exposure that has been found to induce significant levels of inflammation is viral infection. Some viruses commonly associated with MIA are Influenza, Coronavirus and Zika (CDC, 2022) (Marteleto, 2021). These viruses can produce serious symptoms such as fevers, rashes, vomiting, ultimately resulting in inflammation throughout various parts of the body. This type of viral-induced inflammation is thought to be related to changes in neurodevelopment. For example, between 2015 and 2017, the Zika outbreak was at its peak. Interestingly, at this time there was also a spike in fetal developmental deficits. In Brazil, almost 4,000 babies were born with congenita malformations such as difficulties talking, seeing, walking and eating. This suggests that Zika, a mosquito born disease, can cause numerous birth defects if caught during pregnancy. A more recent example of a maternal viral infection is COVID-19. A few months into the pandemic, pregnancy was added to the list of at- risk health conditions by the U.S. Centers for Disease Control and Prevention (Marteleto, 2021). Finally, the flu can also be damaging for a developing fetus. Similar to Zika and Coronaviruses, the flu can induce widespread inflammation that could result in neurodevelopmental damage (CDC, 2022). The inflammation present when contracting these three viruses is deemed dangerous for the fetus because of its possibility to lead to a developmental issue, and more specifically, ASD.

While viruses have been shown to induce MIA, other types of environmental factors can also induce inflammation. For instance, exposure to stress has been shown to induce inflammation. A study conducted at University of California Irvine (UCI) focused on the environment of the expecting mother and how it might relate to psychiatric disorders in the offspring (Boulanger-Bertolus et al., 2018). Teenage mothers in particular seemed prone to elevated inflammation levels, which in turn correlated with brain structure alterations. It was hypothesized that teenage mothers may be more vulnerable to stress (e.g., due to economic instability, unplanned nature of pregnancy, stigma), which could partly explain their increased levels of inflammation. The study showed that stress exposure directly correlated with neurodevelopmental issues (Boulanger-Bertolus et al., 2018).

The body's initial reaction to a detected foreign substance is called innate immunity. To initiate an antiviral immune response, the pathogens need to be detected and considered a threat (Lim, 2013). Toll-like receptors (TLRs) are a type of pattern recognition receptor (PRRs) whose primary purpose is to identify pathogen-associated molecular patterns (PAMPs) and trigger a defense response (Lim, 2013) (Sameer and Nissar, 2021). After identifying the pathogens, these membrane proteins induce inflammatory cytokines that will recruit immune cells to try to kill the virus (Lim, 2013). During pregnancy, a strong immune system is especially important because without it, the pathogen can destroy the fetus. A one study demonstrates, mice lacking TLR-4, a key PRR, were much significantly more susceptible to fetal loss and cognitive defects (Chan et al., 2021). These data suggest that TLR4 is essential for fetal development because without it, the fetus has lost an important mechanism of immune protection and becomes more exposed to various pathogens. However, as the immune system proceeds to attack the pathogens, the fetus may be affected because it might be mistakenly identified as a foreign substance in the body.

Despite inflammation being beneficial in infected regions, inflammation can also have long-term negative impacts on the brain of the offspring. Damien Fair and his research team conducted a study connecting inflammation to an increased risk of neurodevelopmental and psychiatric disorders (Brawley, 2018). The study included assessments of the memory and executive functioning skills in two-year old children that had been exposed to MIA in utero. Throughout each trimester of the pregnancy, Fair collected blood samples from each of the exacting mothers from which they measured the quantity of an inflammatory marker (Brawley, 2018). The results showed that less inflammation correlated with better memory scores and brain communication. This suggests that inflammation directly affected the neurodevelopment in the fetus.

While general inflammation has been found to be detrimental to neurodevelopment, this effect is primarily driven by cytokines. A cytokine is a protein that either triggers the immune system or slows it down, with each variation possessing a different job (Vinicius et al., 2018). To illustrate this, Fair's study highlighted the importance of IL-6 in negative inflammatory responses in the brain (Gabay, 2006). IL-6 is an interleukin cytokine released by macrophages, a cell type commonly associated with MIA. Interleukin cytokines specifically work to stimulate the immune system

Vol. 2023 (2) 1 – 5
ISSN 2688-3651 [online]

when necessary. Because of this, interleukins are commonly investigated in studies of MIA.

Another aspect of MIA and inflammatory responses that is under investigation is the environments that lead to IL-6 production. IL-6 is associated with both increased stress and neurodevelopmental problems. This was demonstrated in both Fair's study and the study examining teenage mothers. In both of these works, elevated IL-6 levels were associated with higher levels of stress exposure (Boulanger-Bertolus et al., 2018). The teenage mothers were shown to possess both higher psychosocial stress levels and IL-6 levels (Boulanger-Bertolus et al., 2018).

Thus, as mentioned above, the mothers with the most IL-6 exposure (mostly the teenage mothers) corresponded with the fetuses with more neurodevelopmental issues (Boulanger-Bertolus et al., 2018).

Another cytokine commonly associated with MIA is Interleukin-17 (IL-17), a cytokine released by Th17 cells. Similar to IL-6, this cytokine induces inflammation in response to invading pathogens. Neuroscientist Gloria Choi conducted a study on MIA in mice that examined how different exposures to pathogens and inflammation in utero impacted the developing offspring (Choi et al., 2016). Choi's study consisted of two groups of pregnant mice: a group injected with IL-17 inhibitors or a control group without IL-17 inhibitors (Choi, et al., 2016). The inflammatory levels present in each mouse was also recorded (Choi et al., 2016). After birth, the offspring of both groups were tested for behavioral changes that mimic symptoms of ASD such as abnormal communication, repetitive behaviors and social deficits (Choi et al., 2016). It was found that the pups born from mice injected with displayed in the mice without the IL-17 inhibitor injection, rather than with it (Choi et al., 2016). These results suggest that, like IL-6, IL-17 can also induce inflammation in a way that impacts ASD in offspring.

As important as cytokines have been shown to be in maintaining health, it is clear that in times of pregnancy, they can increase the risk of neurodevelopmental issues in utero. Since the interleukins are the cause of inflammation, they are the considered key contributors to MIA and a risk factor for ASD. The placenta, an organ that develops to provide nutrients to the fetus, acts as a shield for the developing child, however only to an extent (Boulanger-Bertolus et al., 2018). For instance, when cytokines induce inflammation in an expecting mother (MIA), they can pass through the placenta that is protecting the fetus (Han et al., 2021). After passing through the placenta, the exposure to the cytokine can alter the neurodevelopment of the fetus. As presented in Choi's study, the inflammation can cause cortical changes in regions of the brain controlling behavioral and cognitive abilities, possibly resulting in ASD (Figure 1) (Choi et al., 2016).

Considering the strong relationship between inflammation and neurodevelopmental issues, it is reasonable to consider cytokines and inflammation as a viable target for preventative ASD therapies. One possible approach to reducing inflammation is producing pregnancy safe immunosuppressants. Immunosuppressants are commonly used for organ transplants, autoimmune diseases, and other scenarios where an active immune system is detrimental (Han et al., 2021). The immune system's job is to keep the body safe and remove anything that is foreign. However, in cases of transplants or pregnancy, the immune system can do more harm than good. In the case of MIA, taking an immunosuppressant would reduce the strength of the immune system, allowing the fetus to be more protected from inflammatory cytokines.

Although an immunosuppressant may seem like a good solution, like most medications, there are possible risk factors. The most obvious issue regarding an immunosuppressant during pregnancy is whether or not it is pregnancy safe. Expecting mothers are cautioned to consider what medications they're taking in order to ensure safety for themselves and the fetus. Current research has limited information regarding safety for pregnant women, so immunosuppressants are not commonly prescribed during pregnancy (Han et al., 2021). However, some types of this drug, such as hydroxychloroquine can cross the placenta with no

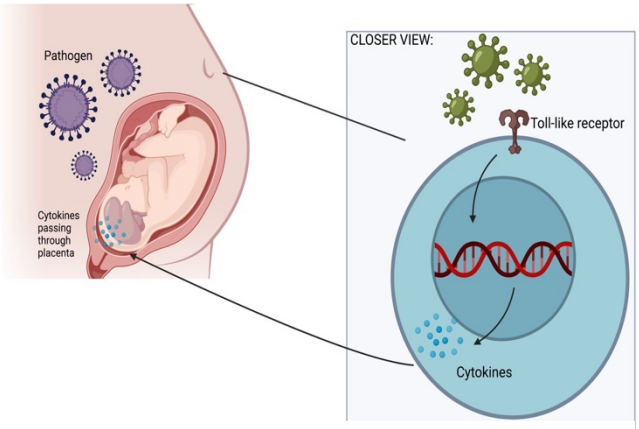


Figure 1. Graphical representation of maternal immune activation inducing cytokine expression through toll- like receptors. Figure made in Biorender.

Vol. 2023 (2) 1 – 5
ISSN 2688-3651 [online]

evidence of neurodevelopmental defects.

Even if the immunosuppressant is proven to be pregnancy safe, it might not be the best option for someone with a virus or autoimmune disease. The point of an immunosuppressant is not to lessen the strength of a specific action of the immune system, but to weaken the entirety of the immune system. This raises more health risks for the expecting mother because it also is reducing her ability to fight off any virus she may contract. A more specific approach is a cytokine inhibitor. A cytokine inhibitor can target specific cytokines and limit their ability to function. If this was directed towards an inflammatory cytokine associated with neurodevelopment impacts, such as IL-6 or IL-17, it could possibly reduce inflammation that could harm the fetus, while allowing the rest of the immune system to function properly.

3. Conclusion

Maternal immune activation in an expecting mother puts the fetus at higher risk for neurodevelopmental and neuropsychiatric disorders, such as autism. Some common MIA inducers are viral infections and stress. When these instances are detected in the body, inflammatory cytokines are released to help the body return to normal health. Although the inflammation may be helping the mother, it could also be attacking the fetus. Once MIA is induced, cytokines are capable of passing through the placenta and damaging the neurodevelopmental process of the fetus. However, MIA can be responsible for more neuropsychiatric disorders than just autism. For example, inflammation has also been associated with schizophrenia. Thus, work investigating the benefits of inhibiting detrimental inflammation during pregnancy could have a positive impact on a wide variety of disorders. Better understanding of the causes and outcomes of MIA on neuropsychiatric disorders is vital to the future of neurodevelopmental health. With the creation and utilization of new therapies targeting inflammation, science and medicine can continue to develop to better help and prevent neurodevelopmental disorders such as ASD.

Acknowledgments

I would like to thank my mentor Courtney Rivet-Noor for guiding me through the writing process and teaching me so much along the way.

References

Boulanger-Bertolus, J., Pancaro, C., & Mashour, G. A. (2018). Increasing role of maternal immune activation in neurodevelopmental disorders. *Frontiers in Behavioral Neuroscience*, *12*. https://doi.org/10.3389/fnbeh.2018.00230

Brawley, T. (2018, April 9). *Study confirms that inflammation during pregnancy is linked to Baby's brain*. OHSU News. https://news.ohsu.edu/2018/04/09/study-confirms-that-inflammation-during-pregnancy-is-linked-to- babys-brain

Centers for Disease Control and Prevention. (2023, June 7). *Flu & pregnancy*. Centers for Disease Control and Prevention. https://www.cdc.gov/flu/highrisk/pregnant.htm

Chan, H. Y., Moldenhauer, L. M., Groome, H. M., Schjenken, J. E., & Robertson, S. A. (2021). Toll-like receptor-4 null mutation causes fetal loss and fetal growth restriction associated with impaired maternal immune tolerance in mice. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-95213-1

Choi, G. B., Yim, Y. S., Wong, H., Kim, S., Kim, H., Kim, S. V., Hoeffer, C. A., Littman, D. R., & Huh, J. R. (2016). The maternal interleukin-17a pathway in mice promotes autism-like phenotypes in offspring. *Science*, *351*(6276), 933–939. https://doi.org/10.1126/science.aad0314

Gabay, C. (2006). Interleukin-6 and chronic inflammation. *Arthritis Research & Therapy*, *8*(Suppl 2). https://doi.org/10.1186/ar1917

Vol. 2023 (2) xxx – xxx
ISSN 2688-3651 [online]

Han, V. X., Patel, S., Jones, H. F., & Dale, R. C. (2021). Maternal immune activation and neuroinflammation in human neurodevelopmental disorders. *Nature Reviews Neurology*, *17*(9), 564–579. https://doi.org/10.1038/s41582-021-00530-8

L. Ferreira, V., H.L. Borba, H., de F. Bonetti, A., P. Leonart, L., & Pontarolo, R. (2019). Cytokines and interferons: Types and functions. *Autoantibodies and Cytokines*. https://doi.org/10.5772/intechopen.74550

Lim, K.-H., & Staudt, L. M. (1970, January 1). *Toll-Like Receptor Signaling*. Cold Spring Harbor Perspectives in Biology. https://cshperspectives.cshlp.org/content/5/1/a011247.short

Lim, K.-H., & Staudt, L. M. (1970, January 1). *Toll-Like Receptor Signaling*. Cold Spring Harbor Perspectives in Biology. https://cshperspectives.cshlp.org/content/5/1/a011247.short

Rudolph, M. D., Graham, A. M., Feczko, E., Miranda-Dominguez, O., Rasmussen, J. M., Nardos, R., Entringer, S., Wadhwa, P. D., Buss, C., & Fair, D. A. (2018). Maternal IL-6 during pregnancy can be estimated from newborn brain connectivity and predicts future working memory in offspring. *Nature Neuroscience*, *21*(5), 765–772. https://doi.org/10.1038/s41593-018-0128-y

Sameer, A. S., & Nissar, S. (2021). Toll-like receptors (tlrs): Structure, functions, signaling, and role of their polymorphisms in colorectal cancer susceptibility. *BioMed Research International*, *2021*, 1–14. https://doi.org/10.1155/2021/1157023

Journal of Research High School

Vol. 2023 (1) 6 – 12
ISSN 2688-3651 [online]

The Emergence of Losers in the South African Economy as a Result of Trade Liberalization

Sriya Bandyopadhyay[1] *
[1]The American School of Dubai, Dubai, UAE
*Corresponding Author: sriyabandyopadhyay@gmail.com

Advisor: Carl McPherson, cjjmcpherson@gmail.com

Received Appril 23, 2023; Revised May 23, 2023; Accepted, June 21, 2023

Abstract

Trade liberalization is the process of reducing government-imposed restrictions on international trade. In South Africa, trade liberalization has been a key aspect of economic policy since the end of Apartheid in the 1900s. The nation has sought to increase its competitiveness by removing trade barriers, implementing more liberal trade policy frameworks, and anticipating the demands of South Africans for foreign goods and of foreign countries for South African goods. The results of trade liberalization have been mixed, with some industries benefiting greatly from this liberalization while others have struggled due to a lack of government support. Despite these challenges, South Africa continues to pursue business that its natural reserves and technological advancements can accommodate and is experiencing rising demand. This paper analyzes the mining, vehicle manufacturing, and agricultural industries in South Africa. Recent policy actions and trade transactions demonstrate that all three industries are experiencing growth. However, too much liberalization (lack of government support) is posing a threat to the extraction of certain precious metals, while the inconsistent distribution of foreign investment in vehicle manufacturing is causing the emergence of losers among a small number of small and medium-sized enterprises (SMEs).

Keywords: Economics, South Africa, Trade liberalization, Apartheid, International Trade

1. Introduction

South Africa's prospective to contribute to the global economy is significant and has the potential to grow if it can make the most efficient and tactical use of its resources, which starts with knowing where productivity is lacking as a result of its recent trade liberalization measurers. With a young and growing population, a favorable geographic location acting as a gateway to the African continent, and new involvement in international organizations such as the World Trade Organization (WTO), the Group of 20 (G20), and Brazil, Russia, India, China, and South Africa (BRICS), the nation will have more opportunities to engage in global trade negotiations, promote investment, and collaborate with other nations on economic and development issues. Even so, the country must identify the losers in the South African economy. In this paper, the term "losers" refers to individuals, groups, or industries that experience negative consequences or hardships due to various economic factors or changes, which in this case, is trade liberalization. The term "losers" does not imply personal failure or a lack of effort but rather highlights the adverse impacts certain individuals or sectors may face. By recognizing the losers, South African policymakers can design targeted policies, interventions, or support mechanisms to address their concerns, mitigate negative impacts, and promote a more inclusive and equitable economic environment.

South Africa has a long history of trade liberalization, which began in the 1990s as the country transitioned from apartheid to a democratic government. The government of South Africa has implemented a number of policies aimed at reducing trade barriers and encouraging foreign investment. These policies have led to increased trade and economic growth but have also had some negative consequences, such as increased competition for domestic industries and a growing trade deficit. Despite these challenges, the South African government continues to pursue trade liberalization

Vol. 2023 (1) 6 – 12
ISSN 2688-3651 [online]

as a key strategy for economic development. As such, the following analysis demonstrates that there are three groups that will be experiencing losses due to trade liberalization in South Africa: the gold-mining industry; vehicle-manufacturing firms that are unable to develop advanced technologies; and agricultural industries that are overdependent on remittances.

Prior to implementing trade liberalization, South Africa pursued an industrialization strategy of import substitution (ISI). Reduced reliance on imports and the development of domestically owned, self-sustaining industries were the two objectives of ISI. This strategy had some detrimental effects, such as lowering customer choice in terms of commodities, decreasing economic efficiency, and delaying the growth of the export industry. South Africa started moving away from ISI in the 1990s and toward trade liberalization. Over the last decade, South Africa's trade policy has changed significantly, including the elimination of tariffs and subsidies through WTO commitments, the signing of free trade agreements with other nations, and the negotiation of upcoming liberalization commitments at the multilateral and regional levels. However, the average unweighted tariff on imported goods in South Africa is still fairly high at around 11% (du Venage, 2021). South Africa's overall commerce (imports plus exports) grew from 37% to 60% of GDP between 1994 and 2018 (Flatters and Stern, 2007). As a result, South Africa's exports have greatly increased, with minerals, metals, and automobile items among its top exports. Due to their great contributions to the nation's GDP, this paper focuses on the changes experienced in the Metal Mining, Vehicle Manufacturing, and Agricultural industries.

However, in times when the nation is facing difficulties with adopting liberalized trade policies, two industrial policies are ensuring some stability in the country. The South African government created the National Industrial Policy Framework (NIPF) as a framework for policy to direct the industrial growth of the nation. The NIPF has provided support for small and medium-sized enterprises (SMEs), which are an important source of employment and economic growth in South Africa (Mather and Greenberg, 2010). The second industrial trade policy that has become essential to the nation is the Industrial Policy Action Plan (IPAP). The IPAP (Mather and Greenberg, 2010) addresses a multitude of factors that affect the South African economy, such as their unstable currency, the slow rate of their capital expenditure, the anti-competitive behavior of competitive firms, the weak skills system, the poor sector-level productive integration, and the scarcity and high cost of capital. Another important policy framework to note are Black Economic Empowerment (BEE) policies, which set targets for the participation of black-owned and black-managed companies in the industry, and the creation of special economic zones that offer tax incentives and other support for businesses owned by historically disadvantaged groups. BEE policies are essential when considering trade because they establish standards that companies need to meet in terms of equity ownership, and change the composition of ownership in businesses and therefore the firm's trade partners.

1.1 Mining Metals

The metal mining industry in South Africa has undergone significant changes in employment in recent years. The industry has been affected by a combination of factors, including global economic conditions, competition from cheaper imports, and technological advancements. The industry has been negatively impacted by the lack of government support, power shortages, and other infrastructural problems.

Gold and uranium ore are significant parts of the South African economy (S&P Global, 2019). The gold market in South Africa has experienced a number of changes in recent years. Only two years in the past 20 years have seen a positive annual increase in gold production for gold mining businesses, with South Africa generating 87% less gold in January 2015 than it did in the same month in 1980 (Republic of South Africa Statistics Department, 2015). Given that South Africa, according to estimates from the US Geological Survey, has the second-largest gold reserves in the world, the steep fall in gold production is surprising. As such, politicians who are relying on the growth of this profitable sector are upset about the status of the sector. According to Kevin Mileham, the opposition's spokesman for energy and natural resources, mining is dying "not because the mineral resources are running out, but because of government incompetence, bad policy decisions, and militant labor unions," he said in a July speech to parliament. These operational obstacles and a lack of government support are slowing down the mining process of precious metals such as gold and preventing South Africa from being able to compete with foreign competitors. The uranium market

Vol. 2023 (1) 6 – 12
ISSN 2688-3651 [online]

in South Africa, on the other hand, has seen an increase in demand (Consolidated Uranium Inc., 2023), driven by the growing use of nuclear energy worldwide. However, this demand has not translated into significant growth in production, as South Africa has relatively limited reserves of uranium ore. Uranium deposits were found in the Namaqualand region (surficial deposits), the Karoo Uranium Province (sandstone-hosted deposits), and the Springbok Flats Basin between 1967 and 1976. (carbonaceous shale and coal-hosted deposits). Recent exploration has boosted the potential for resources in these deposits, with the exploratory activities concentrating on drilling to confidently identify resources. Therefore, while there have definitely been setbacks in the mining of these specific metals, there are opportunities that the South African people are trying to take advantage of.

1.2 Vehicle Manufacturing

The South African automobile component trade has undergone significant changes in recent years. In recent times, the industry has shifted towards exporting components to other countries in order to maintain competitiveness and growth. Factors that affect this include labor costs, taxes, and the overall cost of living. South Africa's cost competitiveness has been negatively affected in the recent past by factors such as high unemployment, low GDP growth, and political instability. The total contribution of the automotive industry to the GDP in 2020 was 4.9% (2,8% manufacturing and 2.1% retail), down from 6.4% in 2019, which was due to the severe effects of COVID-19 on automotive manufacturing and retail as a result of the country lockdown restrictions that were in place during the year (NAAMSA, The Automotive Business Council, 2023). The manufacturing of vehicles and automotive components, the country's largest manufacturing sector, contributed a significant 18.7% of the value added to the domestic manufacturing output, maintaining the industry's position as a major participant in South Africa's industrialization process. According to the National Association of Automotive Component and Allied Manufacturers (NAACAM), the industry employs over 130, 000 people in South Africa, and the export of components is worth around R56 billion ($3.7 billion) annually. The NAAACAM provides services to its members, such as market information, training, and networking opportunities, and also works to improve the competitiveness of the industry through the promotion of technology transfer, research, and development. Additionally, the South African government has implemented supportive programs such as the Automotive Production and Development Program (APDP) and the Automotive Industry Support Program (AISP) that provide financial assistance to the industry. So, the vehicle manufacturing industry has managed to stay stable and provide opportunities for employment and increased integration into the domestic economy.

1.3 Agriculture

The agricultural industry in South Africa has undergone significant changes in employment in recent years. The industry has been affected by a combination of factors, including changes in government policies, globalization, and advancements in technology.

One of the most prosperous sectors of the South African agricultural industry is the citrus market. The Citrus Exchange (Mather and Greenberg, 2010) in South Africa is a platform for the trade of citrus fruits, such as oranges, lemons, and grapefruits. The majority of South African citrus exports are destined for Europe, with other key markets including the Middle East, the Far East, and North America. The industry is governed by various trade agreements, such as the European Union-South Africa Trade Agreement and the Southern African Development Community (SADC) Trade Protocol. These agreements aim to facilitate trade and ensure fair competition in the global citrus market.

However, another significant factor in the agriculture economy in South Africa is the corn market. The corn market is a huge part of labor and remittances (Crush and Caesar, 2017) and an important factor in the chronic poverty in the nation. While the lowering of trade barriers tends to indicate more economic integration, opportunities, and hence economic growth, South Africa has reacted in the opposite way. Studies have shown that substantial resource transfers from migrants to South Africa haven't significantly reduced chronic poverty (Booth, 2021). Where remittances are used to create money, they generally replace revenue streams that have been lost as a result of the

Vol. 2023 (1) 6 – 12
ISSN 2688-3651 [online]

country's general economic downturn. This situation is paradoxical because there are few incentives for migrants to remit where networks are weak, such as in places with more recent mobility patterns. These findings have significant ramifications for our understanding of the purpose of remittances and our expectations of how they can advance social protection and reduce poverty.

2. Materials and Methods

The specific industries identified in this research are metal manufacturing and mining, transport, and agriculture. Once the industry or industries have been identified, gather relevant data on employment within those industries. This includes data on the number of employees in each of the respective industries. The data for all graphs has been collected from the Integrated Public Use Microdata Series, International (IPUMS-International). The data for general industry participation is from the "INDGEN" variable in the database. The INDGEN variable records the industrial classifications of all the various samples into twelve groups that can be fairly consistently identified across all available samples on the IPUMS-International database. Generally speaking, the groups follow the International Standard Industrial Classification (ISIC). Important information about the service sectors that could not be reliably differentiated in all samples is retained in the third digit of INDGEN. The activity or output of the company or industry where a person worked is referred to as the "industry." The data was then adjusted to accurately represent the number of people working in each industry by examining the "PERWT" variable. This variable calculates the actual part of the population that has been considered for data in each respective year.

3. Resutls

The relationship between trade liberalization and employment in South Africa is mixed and depends on the specific industry. On the one hand, trade liberalization has allowed South African businesses to access larger markets, attract foreign investment, and promote export-oriented industries. This has had a particularly beneficial impact on the manufacturing industry, as employment in that sector has increased from 1996 to 2007. On the other hand, trade liberalization has also exposed some sectors, as increased competition from imports has led to a decrease in employment in the mining and extraction and agricultural industries. The country's historical inequalities and structural challenges, especially those relating to skill gaps and limited labor market flexibility, have contributed to declining domestic mining and extraction and agricultural industries.

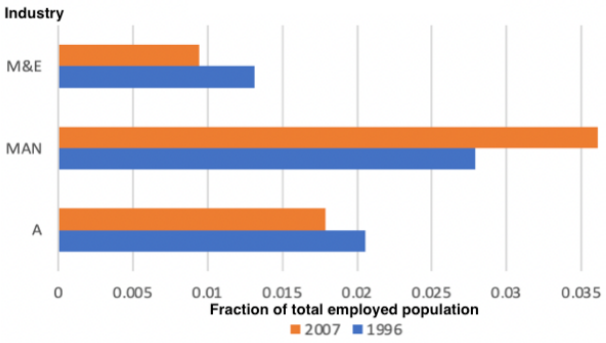


Figure 1. Illustrates the fraction of the total employed population employed in the Mining & Extraction (M&E), Manufacturing (MAN), and Agricultural (A) industries in 1996 and 2007.

Note. IPUMS-International. (1996, 2007). Harmonized International Census Data for Social Science and Health Research. *https://international.ipums.org/international/index.shtml*

4. Discussion

Knowing a country's leading sector is important because it can help inform government policies and decision-making in areas such as taxation, infrastructure development, and trade. It can also provide insight into where to focus investment and resources to promote economic growth. Additionally, understanding the leading sector can help identify potential areas of expertise and competitiveness in the global market.

Vol. 2023 (1) 6 – 12
ISSN 2688-3651 [online]

Based on Figure 1, it is suspected that the mining industry experiences a decline in overall employment, and this is confirmed by Table 1, which depicts the decrease in nominal value of workers in the sector. This clearly indicates a decrease in the demand for jobs in the mining industry, and this aligns with what was mentioned in the literature review. Currently, there is only one metal that South Africa can export, uranium, so that already narrows down the avenues South Africans can take in terms of employment. Furthermore, it confirms that the insufficient government support to allow for efficient extractions of the ore is limiting the number of firms that can actually find this metal and trade with foreign consumers. The data conveys that trade liberalization has indirectly allowed the metal mining industry to focus and prioritize extracting metals that are needed more urgently and are considered to be harder to find.

Table 1. States the number of employees in the Mining & Extraction (M&E), Manufacturing (MAN), and Agricultural (A) industries in 1996 and 2007.

Number of Employees by Industry		
Industry	1996	2007
A	834428.23	843395.13
MAN	1132228.11	1704152.66
M&E	531707.14	444158.24
Total	40578356.9	47173595.1

Note. IPUMS-International. (1996, 2007). Harmonized International Census Data for Social Science and Health Research. *https://international.ipums.org/international/index.shtml.*

The vehicle manufacturing industry is enjoying an expansion of employment, as evident from all three figures and tables. During Apartheid, trade regulations were incredibly strict; hence, South Africans could either choose to buy a car or a motorcycle from a local firm or not at all. As such, trade liberalization has clearly revolutionized the vehicle manufacturing industry for three reasons. Firstly, there are many emerging associations and councils that work to protect the rights of these companies. Secondly, employee rights are being considered more heavily. Thirdly, there is an increase in direct investment from foreign manufacturing companies. This means more South Africans are needed to work at these newly available companies. However, something to consider is that, in order to keep up with global demands, South Africa will need to adopt advanced mechanization processes to produce vehicles more efficiently. This will mean that eventually there should be a decrease in employment in this industry, but it isn't apparent yet.

Table 2. States the fraction of employment (out of total national employed population) in the Mining & Extraction (M&E), Manufacturing (MAN), and Agricultural (A) industries in 1996 and 2007.

Fraction of Total Employed Population by Industry		
Industry	1996	2007
A	0.02056338	0.01787854
MAN	0.02790227	0.03612514
M&E	0.01310322	0.0094154

Note. IPUMS-International. (1996, 2007). Harmonized International Census Data for Social Science and Health Research. *https://international.ipums.org/international/index.shtml*

The agricultural industry poses an interesting viewpoint because, while Figure 1 depicts the decline in fraction-based employment, there has been an increase in the nominal value of employees in the sector. This means that the industry is growing at a considerably slower rate. This expresses the true extent of the impact of an increased reliance on remittances after trade liberalization. South Africa is a significant exporter of citrus fruits and corn, but due to the lower barriers to international integration, many South Africans are moving to work at agricultural posts in other countries. This is causing an increase in the remittance inflow to the nation, but it presents the potential situation where these workers will stop sending money back to South Africa, negatively impacting the nation's economy.

5. Conclusion

This paper examines the impact of trade liberalization on three industries in South Africa: metal mining, car manufacturing, and agriculture. These are three industries that contribute most significantly to the nation's exports. In relation to the metal mining industry, small and medium sized enterprises have been hit the worst by trade liberalization in South Africa. Small and Medium Enterprises (SMEs) in the metal industry mining gold in South Africa have been affected in a number of ways. Beginning with the fact that there has been a decrease in demand for precious metals, fewer firms are being incentivized to reap the benefits of this by entering the gold mining industry. The issues surrounding the gold industry include declining grades of ore, increased costs, and increased regulations.

Vol. 2023 (1) 6 – 12
ISSN 2688-3651 [online]

However, the outlier of this trend is Uranium, which is experiencing increasing demand, making it a pull-factor for South Africans in search of jobs. In the vehicle manufacturing industry, it is clear that while car manufacturing holds the most significant position in the South African GDP, the losers from trade in this market are the South African car manufacturers who do not have the appropriate technology to adapt to electric vehicles. In this way, these local manufacturers are not able to export their products as other nations increasingly want eco-friendly forms of transport. But, it is still experiencing gradual increases in employment since there is an increase in direct foreign investment for these firms. The agricultural industry seems to be the only South Africa where there is no clear "loser". The nation has rich resources allowing them to satisfy the demand of a growing population and chronic poverty in many rural areas. Out of all the markets in the agricultural sector it is clear, however, that the citrus and corn markets are what carry the heaviest load of the GDP and the South African resources. While the industry is experiencing slow growth, it is the only one that is expected to grow consistently due to trade liberalization. Although, it should not be overlooked that corn-producing-firms in the agriculture industry are vulnerable to becoming losers if predictions about the increased remittances come to pass. As a general trend, trade liberalization decreased the price of the goods examined in the industries analyzed in this paper and increased their quality. As such, it could be further researched whether this a tendency that exists in all African countries, or whether it is unique in South Africa due to its history of Apartheid. It could even be expanded to other regions that may have undergone racial segregation and inequality to a similar scale.

Acknowledgments

Thank you for the guidance of Mr. Carl McPherson from the University of California, Berkeley in the development of this research paper.

References

Barnes, J., & Kaplinsky, R. (2000). Globalization and the death of the local firm? The automobile components sector in South Africa. Regional Studies, 34(9), 797–812. https://doi.org/10.1080/00343400020002949

Booth, A. (2021). *The reason we're running out of farmers*. BBC. Retrieved April 23, 2023, from https://www.bbc.com/future/bespoke/follow-the-food/the-reason-we-are-running-out-of-farmers/

Caesar, M. (2017). Why food remittances matter: rural-urban linkages and food security in Africa. Fao.org. https://www.fao.org/urban-food-actions/resources/resources-detail/en/c/1043625/

Consolidated Uranium Inc. (n.d.). Consolidated Uranium Inc. Retrieved April 21, 2023, from https://consolidateduranium.com/

du Venage, G. (2021, December 22). How mass theft of railway infrastructure is derailing South Africa's economy. The National. https://www.thenationalnews.com/business/2022/01/04/how-mass-theft-of-railway-infrastructure-is-derailing-south-africas-economy/

Dunne, P. (2006). Trade, Technology and Employment: A Case Study of South Africa. http://carecon.org.uk/DPs/0602.pdf

Republic of South Africa Statistics Department. (2015, March 12). https://www.statssa.gov.za/?p=4252. Retrieved December 23, 2022, from Statistics South Africa website: https://www.statssa.gov.za/?p=4252

Flatters, F., & Stern, M. (2007, June). *Trade and Trade Policy in South Africa: Recent Trends and Future Prospects*. Trade and Trade Policy in South Africa - DNA Economics. Retrieved April 23, 2023, from http://www.dnaeconomics.com/assets/dlas/FILE_063020080109_FILE_101820070217_Trade_and_Trade_Policy_in_South_Africa.pdf

Vol. 2023 (1) 6 – 12
ISSN 2688-3651 [online]

Mather, C., & Greenberg, S. (2010, August 4). Market Liberalisation in Post-Apartheid South Africa: The Restructuring of Citrus Exports after 'Deregulation'. *Journal of Southern African Studies*, *29*(2), 393-412. Taylor & Francis Online. https://doi.org/10.1080/03057070306210

NAAMSA. (2023). *naamsa | Environment 2021*. Naamsa. Retrieved April 23, 2023, from https://naamsa.net/environment-2021/

NAAMSA. (2023). *naamsa | Industry Overview 2021*. Naamsa. Retrieved April 23, 2023, from https://naamsa.net/industry-overview-2021/

Spar, D. L. (2006, Summer). Continuity and Change in the International Diamond Market. *Journal of Economic Perspectives*, *20*(3), 195-208. American Economic Association. 10.1257/jep.20.3.195

S&P Global. (2019, September 1). *South African gold loses its shine*. S&P Global. Retrieved April 23, 2023, from https://www.spglobal.com/en/research-insights/featured/special-editorial/south-african-gold-loses-its-shine

Journal of Research High School

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

Development of a Conceptual Design for a Future Mars Rover Mission

Rishabh N. Srinivas[1] *

[1]Sharon High School, Sharon, MA, USA
*Corresponding Author: rishabhnsrinivas@gmail.com

Advisor: Ramy K. Imam, ramyeladli7@gmail.com

Received September 8, 2022; Revised April 27, 2023; Accepted, May 25, 2023

Abstract

Mars is a widely unexplored planet with few rovers able to gather data on Mars. To hasten the progress of interplanetary research, we as humans need to develop the technology to explore other planets. The current solution is to send vehicles to planets such as Mars. Space Agencies such as NASA have been sending rovers and landers to mars for years. The scientists' effort has brought forth a lot of data about the red planet, and without them, these results could not be possible, and more innovation is needed to continue. Rovers are very complex but can simplify tools within subsystems within the entire rover system. It's detrimental to the function of the rover that every tool performs perfectly. With the development of a conceptual design for this rover, we could significantly decrease the time it would take to uncover the mysteries of Mars. In this paper, I researched details about previous rover missions to figure out a new plan to send a rover to mars. This outline includes the rocket for taking the rover to mars, the landing sequence, the landing location, the path the rover should take, and a basic understanding of what the rover would include and how it would appear. In addition to finalizing a landing location and identifying all working systems, the rover design (shown in the figures below) passed the computational tests with minimal stresses when applied when simulating forces present on Mars.

Keywords: Rover, Mars Rover, Space Exploration, Mars Mission, Mars Exploration, Space Rover, Planetary Rover

1. Introduction

1.1 Problems (Problems of Current Mars Exploration)

As the current situation is now, it would take humans many more years to even set foot on Mars with the numerous constraints placed on space programs worldwide. The feat of controlling three concurrent Mars rovers working as of today (August 3rd, 2022) is extraordinary, but setting foot on Mars within this era is impossible unless more unmanned vehicles are launched. Mars rovers are specifically crafted to work on its terrain but have sacrificed speed to reach that level of precision in their movement. The solution to the sluggish rate of exploration is to send more rovers to mars --covering more area within the same amount of time. The caveat to this solution is the years of planning and dozens of years' worth of funding needed.

Mars is possibly the single alternative planet in our solar system that could support life due to its position in our solar system, capability to support organisms, proof that water was once on Mars, etc. That was all in the past. Mars, now, is a barren wasteland and sports a toxic atmosphere. This atmosphere causes Mars to have countless dust storms and other phenomena similar to earth. These characteristics of Mars only make the planet harder to traverse with a vehicle. Since Mars is 140 million miles/225 million kilometers away from Earth, there is only one chance to send a rover there before returning to the drawing board.

1.2 Background

Mars has a very light atmosphere, approximately 1% of Earth's atmospheric density, yet it is very toxic as it is composed mainly of carbon dioxide. Because of its low atmospheric density, the environment on Mars is very extreme in terms of pressure, temperature, and wind. The temperature of Mars fluctuates throughout the day by 40 degrees Celsius on average: the coldest temperature being -107 degrees celsius in winter and -18 degrees celsius in summer. This vast temperature difference is due to the perihelion and aphelion distances. Because of the temperature differences between seasons, carbon dioxide at the poles condenses during winter and sublimes during summer. The

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

pressure changes diurnally are unique to its environment. Because of the pressure variation, the wind speed is also very extreme. Wind speed varies between 2 and 7m/s in summer to 5-10m/s during winter. Similar to tornadoes on Earth, Mars has dust storms during which the wind speeds up to 30m/s.

The topography of each hemisphere of Mars is also vastly different. The southern hemisphere consists of heavily cratered Highlands. In contrast, the northern hemisphere is smooth and has clear indicators showing separations between sedimentary, volcanic, and Aeolian materials. This difference impacts the planning of the rover's mobility system when driving on mars.

History and Innovation of Rovers

To understand how to improve the design of rover missions, further inspection of past rover missions is needed. The first rover sent to Mars was the Pathfinder rover in 1996. It completed basic tasks: taking pictures, performing chemical analyses of rocks and soil, and collecting data on winds and other weather factors.

The second and third rovers sent to Mars were the Spirit and Opportunity rovers in 2003. They performed the same tasks as the Pathfinder rover except with updated technology to take higher resolution pictures and find more data to support the previous theory that Mars used to have a habitable atmosphere. In addition, it also analyzed the chemical and mineralogical makeup of rocks and soil and examined the interior of rocks.

The next rover sent to Mars was the Curiosity rover in 2011. This robot was jokingly said to be a lab as it has many more scientific tools to aid in data collection. Specifically, it has 17 cameras, a laser to vaporize and study small pinpoint spots on rocks (at a distance), and a drill to collect powdered rock samples. It hunts for clasts that may have formed in water or have signs of having organic material in them. As the previous rovers have indicated, it analyzes the powdered rock samples to find the chemical fingerprints in different rocks and regoliths to determine the composition and its history (past presence of water).

The next rover launched was the Mars 2020 Perseverance rover. Because of past findings from previous rovers, the Perseverance rover's main job is similar to the Curiosity rover, so there was no need to update its overall design, with only its technology needing to be updated.

The most significant innovations between rovers are the power and mobility systems. From the Spirit and Opportunity mission to the Curiosity rover, the power delivery system has changed from solar power to nuclear generators. The solar panels are prone to environmental problems like dust tornadoes that could have caused the rover to lose energy because the dust-covered solar panels cannot produce electricity while it's covered. Sending another rover with solar panels after they realized the mistake would only be more problematic since they would have to figure out how to clean the solar panels. The mobility system improved from a simple 6-box wheel drive train to a 6-wheel rocker-bogie suspension system to aid the rover's travels. The other difference between rovers is the scientific tools onboard. With each iteration, the rovers housed improved cameras and geological analyzers based on previous missions; it also formatted its instruments specific to each expedition. For example, the Pathfinder rover got the first pictures of Mars on the surface and analyzed the planet's makeup. Every rover after the Pathfinder rover attempted to confirm the Pathfinder's findings and explore new theories: oxygen content on mars and signs of water in minerals. (Dubov, 1997)

Table 1. Comparing the 3 current mars rovers (Beck, 2021; Serveveda, 2022; Liu, 2021;Liu, 2022; Mars.nasa.gov, 2022; Owlapps, 2022)

Names of Current Mars Rovers	Curiosity	Perseverance	Zhu Rong
Launch Date	November 06, 2011	July 30, 2020	July 23, 2020
Date landed	August 05, 2012	February 18, 2021	15 May 2021
Coordinates landed at	4.5895 degrees south latitude and 137.4417 degrees east longitude.	18.44°N 77.45°E / 18.44; 77.45 (Octavia E. Butler Landing)	109.925° E, 25.066° N at an elevation of −4,099.4 m
Location Name	Gale crater	Octavia E. Butler Landing	
Operation time (SOLS)	3512	477	Till today
Distance traveled	16.2 miles	7.32 miles	1500 kilometers (as of Jan 31, 2022)
Number of Parameters measured	4	3	5

The designs of rovers have drastically improved since the first rover on Mars. The rovers increased in size,

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

weight, tools aboard the rover, and overall life span. The first rovers on Mars, the Pathfinder and Sojourner rovers sent in 1996, were 65 cm long vehicles, 48 cm wide, 30 cm tall, and weighing 10.5 kg. It was also solar-powered, so its functionality was limited. The distance it was allowed to travel was also limited because it would not work outside their lander's vicinity. They both only operated for 83 sols(85 earth days) and 92 sols (95 earth days), respectively. A sol is approximately 24 hours and 37 minutes on earth. The two newest rovers from NASA, Curiosity and Perseverance rover, are tank-like compared to the first rovers, both sizing at 3 meters in length, 2.7 meters in width, and 2.2 meters in height and weighing in at 1025 Kg and 899 Kg, respectively. Both rovers are still operational after landing on August 05, 2012, and February 18, 2021, respectively. Their long lifetime is thanks to the switch from solar power to nuclear power because it would work during long-lasting dust storms that would leave solar panels powerless. Another notable innovation of the rovers is the mobility system, which has now changed from a 6-wheeled system into a rocker-bogie system because of its ability to traverse many different terrains. (Mars.nasa.gov, 2019).

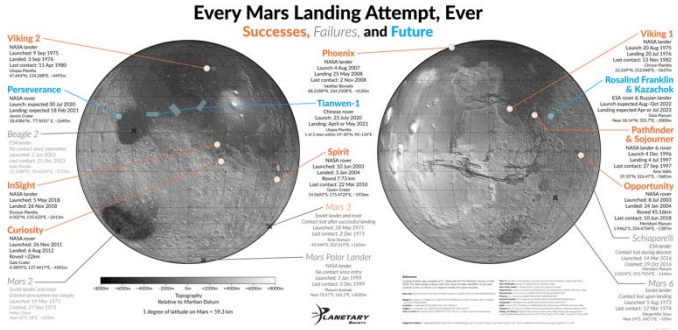


Figure 1. This figure shows where missions were attempted to land on Mars, successful missions are the boxes, the failed missions are the x's, and the future missions are circles (Lakdawalla & McGovern, 2020).

Understanding where and why other Mars missions failed is crucial to bettering future attempts. Looking at figure 1, there are many failed missions meaning many reasons for failing. One of the failed missions, The Beagle 2 lander (launched on June 2, 2003), lost connection to mission control as it entered Mars' atmosphere. Scientists have speculated several causes: an abnormally thin atmosphere over the landing site, electronic glitches, a gas bag puncture, damage to a heat shield, a broken communications antenna, and a collision with an unforeseen object. NASA found the lander on the surface of Mars 10 years later with its Reconnaissance Orbiter in 2013. Controlling as many variables as possible during the mission is crucial because there is only one chance to send that rover into space. Curiosity and Perseverance rovers have been successful because Nasa engineers could control more variables to minimize risk and improve the precision with advanced technology.

Subsystems of Rovers

A working rover has many subsystems. The subsystems are:

- Electrical Power: This subsystem controls power delivery during the entire mission. It includes powering all the other subsystems and maintaining a certain power level to ensure no failures happen. The rover cannot survive long using batteries, so a generator is needed to supply power. Most rovers to this day use nuclear generators.
- Structure: The rover's structure has to be manufactured to complete its mission. It has to be able to carry all the data-collecting instruments. The robot must also be able to protect its scientific instruments despite facing a partially unknown topography.
- Data Handling: This subsystem makes sure whatever data is stored and communicated quickly because of the volatile environment around the rover. The chance of losing valuable data is worse than losing the rover. This subsystem is directly tied to the telecommunications subsystem.
- Telecommunications: The telecommunications system is a two-way wireless UHF(Ultra handling frequency) radio link between a transmitter (lander or satellite) and receiver (rover). This link sends signals directly by first sending them to the proxy. Essentially it works like a walkie-talkie radio system between the lander/satellite and the rover.
- Temperature control: The rover uses a warm electronics box (WEB) to make sure that the robot's electronics can be used in the freezing conditions of Mars. This box is made out of aerogel (essentially a super insulator with a density of 1/50th that of water) to make sure the electronics can function while keeping it between -40 deg C and +40 deg C.
- Environmental Protection: Due to the cold atmosphere of Mars, teams at Nasa made several innovations to ensure its functionality, including low-temperature actuators and lightweight aluminum structures. This specific subsystem is also closely tied to the structure because the exterior has to withstand external conditions while maintaining a safe environment internally.

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

- Mobility: The mobility subsystem has to aid the main body by allowing it to maneuver around mars freely while keeping itself intact. For instance, with the Rocker-bogie suspension system used in the previous 2 NASA rovers, the rover can move freely and over obstacles without springs (because of their fragility).

Design Schematics of Current Rover Systems and Their Scientific Tool

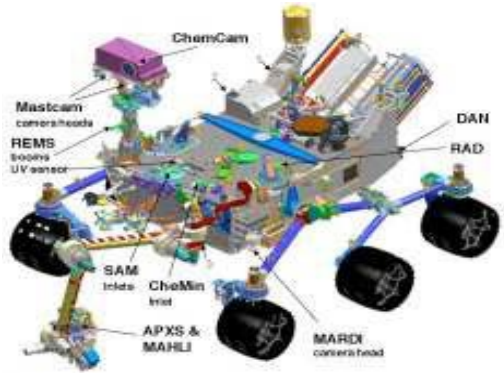


Figure 2. This figure is a schematic of the Curiosity rover with its scientific instruments pointed out (NASA/JPL-Caltech, 2011).

The Curiosity rover, shown in Figure 2, has countless scientific tools on rovers using a schematic of the rover itself. Some are the Chem Cam, DAN, RAD, MARDI camera head, APXS & MAHLI, CheMin, SAM, REMS, Mastcam, and MEDLI. The ChemCam (Chemistry and camera) is "An instrument that first uses a laser to vaporize materials then later analyzes their elemental composition using an onboard spectrograph. The DAN (Dynamic Albedo of Neutrons) is a pulsing neutron generator used to detect water content as low as 1/10th of 1 percent and resolve layers of water and ice beneath the surface. The RAD (Radiation Assessment Detector) is an instrument that measures and identifies all high-energy radiation on the Martian surface: protons, energetic ions or various elements, neutrons, and gamma rays. The MARDI is a camera that took color video during the rover's descent towards the surface, providing an "Astronaut's view" of the local environment. The APXS(Alpha Particle X-Ray Spectrometer) is a spectrometer that measures the abundance of chemical elements in rocks and soil. The MAHLI (Mars Hand Lens Imager) is a camera that provides earthbound scientists with a close-up view of the minerals, textures, and structures in Martian rocks and the surface layer of rocky debris and dust. The CheMin (Chemistry & mineralogy X-ray diffraction) is an instrument that identifies and measures the abundance of various minerals on Mars. The SAM (Samples analysis at Mars instrument suite) is a suite of scientific instruments that searches for compounds of the element carbon associated with life and explores ways in which they are generated and destroyed in the martian ecosystem. The REMS (Rover Environmental Monitoring Station) is an instrument that measures and provides daily and seasonal reports on Martian weather. The MastCam is a camera that takes color images and video footage of the Martian terrain. The instrument is also used to study the Martian landscape and support the driving and sampling operations of the rover. The MEDLI (Mars Science Laboratory Entry Descent and Landing Instrument) is an instrument that collected engineering data during the spacecraft's high-speed and hot entry into the Martian Atmosphere. (Mars.nasa.gov, 2019)

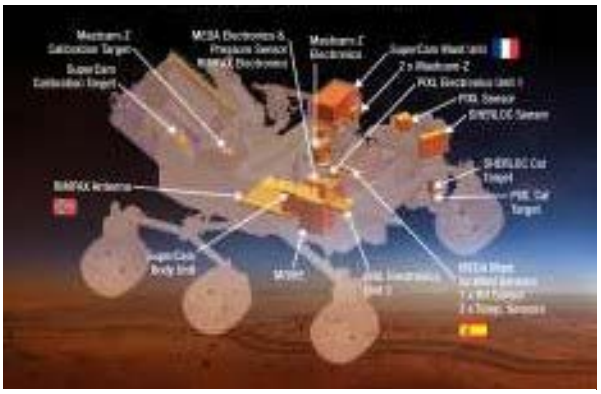


Figure 3. This figure is a schematic pointing at each of the tools atop the Perseverance Rover. (The anagrams are anagrams for the tools and will be explained in the following paragraph). (NASA/JPL-Caltech, 2020)

The Perseverance rover, shown in Figure 3, has many different scientific tools and different tools compared to the Perseverance rover's predecessor, the Curiosity rover. Some of the instruments on the Perseverance rover are the same technology on the Curiosity rover but innovated to increase quality and efficiency. The Perseverance rover still has a Mastcam abroad which is now called Mastcam-z. The new technology includes the MEDA (Mars Environmental Dynamics Analyzer), a set of sensors that will provide measurements of temperature, wind speed, direction, pressure, relative humidity, dust size, and shape. Another new technology is the MOXIE (Mars Oxygen ISRU experiment), an exploration technology investigation that will produce oxygen from Martian Carbon Dioxide. The PIXL is an X-ray fluorescence spectrometer that will also contain an imager with high resolution to determine the fine-scale elemental composition of martian surface material. It will also provide capabilities that permit enhanced detection and analysis of chemical elements than ever before. RIMFAX is also a new technology aboard the rover. It is a ground penetrating radar that will provide centimeter-scale resolution of the geologic structure of the subsurface. The SHERLOC, a spectrometer that provides fine-scale imaging and uses an ultraviolet laser to determine fine-scale mineralogy and detect organic compounds, also boards the Perseverance rover. SHERLOC will be able to be the first UV Raman

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

spectrometer to fly to the surface of Mars and will provide complimentary measurements with other instruments in the payload. The newest camera, the SuperCam, provides imagery, chemical composition analysis, and mineralogy. This instrument can also detect the presence of organic compounds in rocks and regolith from a distance.

Landing locations

NASA's high-interest places, shown in Figure 4, are the SW Melas basin, Holden crater, Eberswalde Crater, Mawrth Vallis, Mili Posse, North East Syrtis Major (NE Syrtis), Jezero crater, and Columbia Hills (Gusev Crater). The criteria for choosing a spot to land has a lot of conditions, for example, "can the rover achieve all the missions' scientific objectives at this place?" or "does this area even have a high variety of soil for the rover to rest and could one of the different soils on this land have supported microbial life before?" etc. (Mars.nasa.gov, 2020).

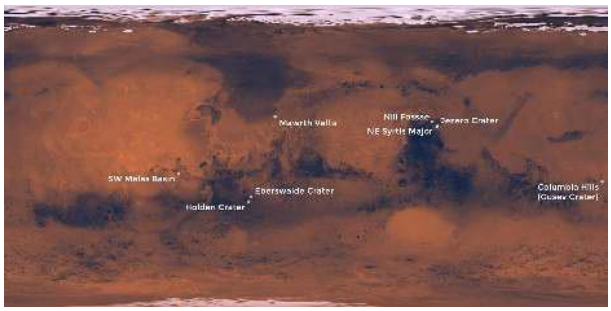


Figure 4. Nasa's 8 highest priority potential landing locations on Mars (NASA/JPL-Caltech, 2015).

Descent of Rovers

The landing sequence, shown in Figure 5, shows the steps that both the Perseverance and Curiosity rovers took during their descent to Mars. The landing sequence of the Perseverance rover is a perfect copy of the landing sequence of the Curiosity rover showing it was the best sequence for transporting rovers. The stages are as follows: after flying through space and entering the Mars atmosphere, the transportation module holding the rover first deploys a parachute to slow down its initial speed going into the atmosphere. After reaching a safe speed, the main shell separates from its heat shield and activates its radar and mobility boosters to ensure it's going the right way. After that stage, it separates from its backshell. After careful maneuvering, it touches down, and the device used to let it touch down flys away.

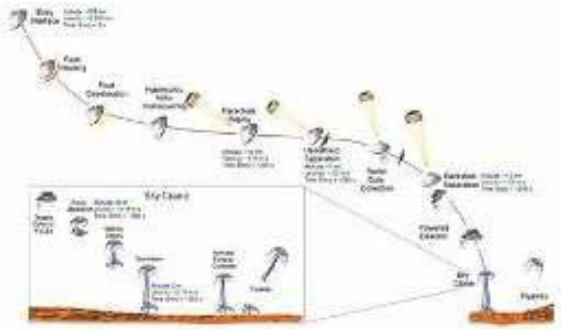


Figure 5. This graphic portrays the sequence of events in August 2012 from when NASA's Curiosity rover, enters the Martian atmosphere to a moment after it touched down on the surface. (NASA/JPL-Caltech, 2010)

Literature Review of Similar Papers

Within the intersection of Aerospace engineering and Robotics, researchers and scientists published hundreds of papers and patents to innovate on current systems. Looking at these ideas, understanding why engineers made specific choices becomes clearer.

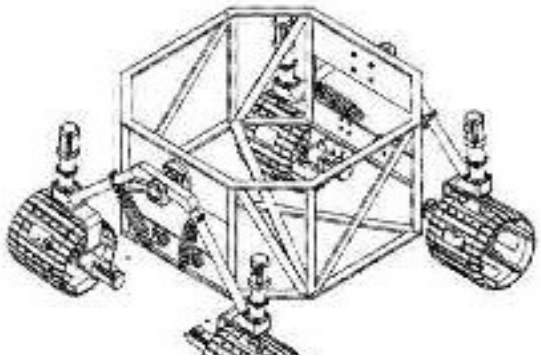


Figure 6. This is a picture of the patent by the researchers at Chongqing University (刘 et al., 2017)

A paper from NASA Glenn Research Center called Exploration Rover Concepts and Development Challenges presented an outline of exploration rover concepts and the various development challenges associated with each as they apply to exploration objectives and requirements for missions on the Moon and Mars. It also provided a brief description of the rover concept proposals since the initial development of the Apollo-era lunar rover, along with a comparison of their relative benefits and limitations. This paper also evaluated developmental challenges such as environment, atmosphere, and design obstacles to ensure the vehicle's protection. Finally, this paper details a modular concept for lunar and martian rovers launched as remote-controlled rovers shifting into astronaut-driven vehicles. This paper gave background information to analyze rover protection systems and put Mars's destructive weather and challenging atmosphere into perspective (Zakrajsek et al., 2005).

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

Another paper, 3DROV: A Planetary Rover System Design, Simulation, and Verification Tool by scientists at the European space agency, builds upon current simulation technology to simulate an entire rover mission, covering everything from the planet's environment to the rover's electrical systems. The development of this technology benefits everyone, from amateurs to professionals, to simulate rover missions in real time versus separate computational tests. (Poulakis, 2008)

The unconventional patent which let me think outside the box for the rover design was a patent by researchers at Chongqing University.

The Chinese researchers' design, shown in Figure 6, although being just a shell, is unlike all the other Martian rovers. This rover has an octagonal chassis base and uses springs for suspension. This rover utilizes swerve drive, a mechanism granting free mobility, to traverse the surface, causing more stress on the battery than a 6-motor chassis. Without proper dimensions, determining this rover's payload capacity becomes difficult. In any case, it is not by any means a bad design. The design is different and requires testing to finalize the results (刘飞et al., 2017).

2. Materials and Methods

2.1 Research on Previous Missions

To understand how to make a new mission, we first must identify what all previous missions did to make it happen. First needing to identify all the challenges of each rover on the planet. In our case, it would be on Mars. Mars has a very light atmosphere which causes the environment to be a lot more extreme in terms of pressure, temperature, and wind. It also has a lot of variation in the land from the southern hemisphere having highlands and heavily created to the northern hemisphere having very smooth land and indicators showing sedimentary, volcanic, and aeolian material. Using this information, we know that the rover has to be able to overcome severe temperature fluctuations, so the necessity of an insulating box around the electronics is key to prolonging the life of the onboard computers/electronics. We also know that the rover has to not be able to be blown in the wind of Mars. To test, a design would need to be placed into a wind tunnel to simulate the extreme wind. The rover itself has to be able to traverse many different terrains because of the vastly different topography of Mars.

The missions of the new martian rovers are fundamentally the same: looking for past signs of life by attempting to identify environments capable of supporting microbial life. It will also seek signs of possible microbial life in rocks and minerals known to preserve organisms over time. The missions also include collecting core rock and “soil” samples and storing them on the Martian surface. By preserving samples on Mars, humans will be better prepared for testing oxygen production from the Martian atmosphere.

2.2 Payload Capacity of the Rocket

Before developing a rover's design, it's crucial to establish a plan to carry the rovers to mars. For the last ten years, the Atlas V 541 rocket took the two rovers from NASA sent to Mars because it has a maximum payload capacity of 6.28 metric tons in GTO Payload 1500 m/s to GEO configuration, and up to 8.29 metric tons in GTO Payload 1800 m/s to GEO configuration. There are other things to consider when landing the rover referenced in the landing section of the background, so these limits are crucial even though the rover is half of the limited weight.

2.3 Plan for the Mission

After finding how the rover will get up to the red planet, we have to figure out where it will land on the red planet. The next step is finding a location for the rover to land. From the Landing locations portion of the introduction (1.2.4), it's known that NASA currently has 8 high priority landing locations. The Jezero crater has one of the highest chances of completing all the missions because of its history of having fluctuating levels of water inside it. Because of the draining and refilling cycles, clay deposits were likely to be formed and may exhibit signs of past microbial life. NE Syrtis is the second best place to land because it has shown volcanic activity before. Another supporting factor for NE Syrtis is that it had ice around it, and that ice would melt and would have been a paradise for microbial life. Some organisms don't even need sunlight to survive, but they would rely on the energy of

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

volcanoes instead.

The way to figure out how to traverse those areas is to sketch an initial path for the rover and have drivers and scientists use the topography of a map of Mars to find obstacles. The initial route aims to try to scout the most area for the least amount of distance traveled.

Before sketching a path, the distance the rover can travel within a period has to be established. After looking at previous rovers' timelines, the rover stays stagnant for long periods to give more power to processing and communicating data. For the Curiosity rover, the average drive length was 28.9 meters after 622 completed out of 738 attempted drives, according to a paper from NASA. The rover's movement took place over 2488 sols. The total distance driven by the rover is 21318.5 meters. Since this is the longest-running, currently operational rover, the ratio of traveled distance to the number of sols run provides an extensive trial of information that can be applied in other rover applications to extrapolate what distance the new rover can travel within 100 sols. (Rankin, 2020)

2.4 What is Needed to Make the Rover Work

Many tools and electronics are needed to make the rover work, and each instrument and device falls under a unique subsystem that would work almost independently of the other subsystems. For example, the power supply subsystem would need a battery, generator, and power splitter to make it work. For rovers, we need a way to get power, move around, and protect against the environment. Understanding the array of tools is crucial to complete all the different missions for the rover. A functionality matrix can track the tools within each subsystem to provide a holistic view of the interactions between subsystems.

2.5 Conceptual Design for the Rover

With the knowledge of what the rover should have inside of it and how it will move around, assembling the subassemblies should be easy. Since the Perseverance rover launched recently, an insignificant amount of mechanical innovations are possible, so the rover would look almost identical to the Perseverance rover. Although the rovers would be mechanically identical, the scientific tools within the rover would be more advanced.

Specifically, the rover needs to have a rectangular body with a rocker-bogie mobility system, a stationary camera on top of it, and a robotic arm that can drill into the surface and extract samples. It also needs to be able to leave mineral samples behind to not slow down the rover over time.

The outside of the body of the rover has to be covered in sensors to determine its location and move accordingly. After all, many environmental variables could impact the encoder feedback from the wheels.

For most of the complex parts of the rover, grabcad.com was used to find the more complicated designs that would take a long time to design by hand (this includes the rocker-bogie chassis, wheels, etc.). Although an extensive website, grabcad.com does not have all the parts needed for the rover, NASA has enough drawings and models to extrapolate a basic shape from the complicated parts. At this point, Autodesk Inventor was used to designing each of those parts by hand to undergo the stress analysis (the rover arm, the outer casing of the rover, the robotic arm, etc.). (Croston T. 2014)

2.6 Tools on Rover

Many things are not essential in the CAD design but need to be included in the rover, like the scientific tools (because it is housed within the main body of the rover). The essential instruments a rover would need are a camera and microphones to know more about Mars' topography and the sounds of the environment around it. The rover would also need tools to extrapolate atmospheric data, so there would be a suite of weather-related instruments to understand the temperature, wind speed, and other things. Another goal of the rover is to understand the mineralogical makeup of the surface of mars, so the rover would need to include a chemical analyzer (a different one for the soil and one for the air), an x-ray or microscope to look closer at the composition and to see if the soil has bacteria in it, and a drill to collect the samples.

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

2.7 Forces Acting on the Rover

Previously covered was the force of the wind on the rover. The strong winds on mars could make the rover susceptible to being blown away if the center of gravity of the rover is not low enough. Another force acting on the rover is gravity. When designing the rover, these forces need to be taken into consideration. To analyze its structural integrity, computational analysis within a CAD program to find a structure that could withstand the forces of Mars. The rover needs to be manufactured from low-density and high-strength materials to withstand the forces while being light enough to be transported.

2.8 Future Fixes to the Design

The plan is rudimentary, not delving deep into each topic but analyzing information on a surface level. For example, the landing location data came from only a few sources. The decision matrix was made using qualitative judgment based on research done by other universities and interpreted as quantitative data based on the advantages and disadvantages of each site. The rover paths were also sketched from judgment on Google Earth's map of mars based on where they should move to complete each mission, cover as much surface area as possible, and avoid elevation changes. With a more detailed map, these paths can be further advanced. Another future improvement would be the design of the rover, but because of time restraints, many aspects of the rover could not be modeled and were generalized. A crucial fix is to add joints and fix inefficiencies before assembling all subsystems. Other changes would be insignificant in comparison but important nonetheless. For instance, changing the computational design would be to include fasteners in as many places as possible and run more simulations with different forces to understand the stress points on the rover.

3. Results

3.1 Decision Matrix Results

This decision matrix was filled out based on research of other studies and information on NASA's website. Table 2 shows how I interpreted the information that I have seen. In this case, the location with the higher total number is the most favorable since it would be able to complete each mission question the most as a rover can do.

Table 2. This decision matrix was used to decide the optimum place to land a rover. It uses the mission questions as criteria; means that the location can barely fulfill that question; means that it fulfills the questions' requirements; means that it excels in this question's criteria.

Place Name (points = pnts)	NE Syrtis	Jezero Crater
Does the area show signs in the rock record that it once had the right environmental conditions to support past microbial life? (7 pnts)	5	7
Does the area have a variety of rocks and "soils" (regolith), including those from an ancient time when Mars could have supported life?(7 pnts)	7	3
Did different geologic and environmental processes, including interactions with water, alter these rocks through time?(7 pnts)	7	5
Are the rock types at the site able to preserve physical, chemical, mineral, or molecular signs of past life?(7 pnts)	7	7
Is the potential high for scientists to make fundamental discoveries with the samples cached by the rover, if potentially returned to Earth someday? (7 pnts)	5	5
Does the landing site have water resources (water ice and/or water-bearing minerals) that the rover could study to understand their potential use by future human explorers? (7 pnts)	7	7

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

3.2 Rover Pathing

$$\frac{21318.5}{2488} = \frac{D}{S} \quad (1)$$

Equation 1 shows how far a rover can move in a certain amount of sols. Where 'S' is the number of sols traveling and 'D' is the total distance traveled in meters.

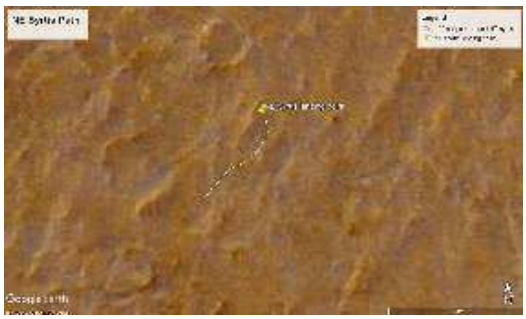


Figure 7. This Figure shows a path plotted for the 100 Sol mission in NE Syrtis. (Adapted from Google Earth)

Figure 8. This is a path plotted for the 200 Sol mission in NE Syrtis. (Adapted from Google Earth)

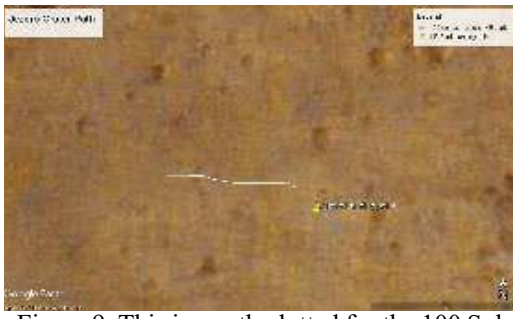


Figure 9. This is a path plotted for the 100 Sol mission in Jezero Crater. (Adapted from Google Earth)

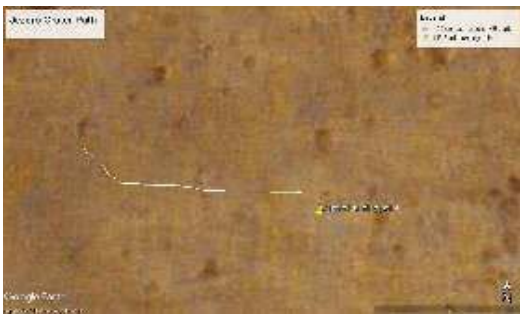


Figure 10. This is a path plotted for the 200 Sol mission in Jezero Crater. (Adapted from Google Earth)

Using Equation 1, in the case of a 100-sol trip, we can solve the distance to be approximately 856.85 meters Figures 7 and 9 show a path that allows the rover directly to get to the nearest change in the soil while staying in the range of 856.85 meters. Both paths have the lowest change in elevation to eliminate all factors that would hurt the rover. Figures 8 and 10 show paths that go on for 200 sols on Mars. It follows the same concept from Figures 7 and 9, but the distance doubled since the expected mission date doubled.

In NE Sytris, the rover is supposed to land at approximately 17°56'22.93"N, 77° 7'53.52" E. The end coordinates for the 100 sol journey are 17°55'41.62"N, 77° 7'26.28" E. The end coordinates for the 200 sol journey are 17°55'33.01"N, 77° 6'41.91" E. Within Jezero crater, the rover attempted to land near 18°28'24.12"N, 77°34'44.75"E. The end coordinates for the 100 sol mission are around 18°28'38.64"N, 77°33'35.06"E. The end coordinates for the 200 sol mission are around 18°28'50.82"N, 77°32'54.19"E.

Table 3. In the decision matrix adapted from the Reusable space tug concept and mission, shown in table 3, many things are vague or unexplained. The MMRTG is a Multi-mission radioisotope thermoelectric generator. The warm electronics box refers to the insulation box surrounding the electronics to prevent the electronics from freezing. The turret refers to the moving arm that collects samples.(NASA, 2020) The cameras refers to devices recording all types of light across the spectrum. The chemical makeup instruments are instruments such as ChemCam, DAN (Dynamic Albedo of Neutrons), RAD (Radiation Assessment Detector), APXS (Alpha Particle X-Ray Spectrometer), and CheMin (Chemistry & mineralogy X-ray diffraction). Whenever the word "elaborate" is used within the matrix, it means to be processed within the main computer of the rover (Cresto et al., 2016).

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

Table 3. The functionality Matrix shows all subsystems of the rover and what instruments and technology are needed to make that subsystem work.

SYSTEM LEVEL	SUB-SYSTEM LEVEL	MMRTG	Batteries	Power distributor	Power converting unit	Shielding panels around computer	warm electronics box	Airless tires	Motors	Rocker-bogie suspension system	Rover body	Rover arm	Turret	Drill	Microphones	Cameras (all spectrums of light,	Chemical make up instruments	Environmental Analyzer	Command & Acquisition unit	On-board computer	Memory unit	Receiver	transmitter	auto-tracking equipment	GPS receiver	Gyroscope/internal measurement unit	body attached sensors	Main Computer	telecommunications unit
Electrical power	To perform power production	X																											
	To perform power storage		X																										
	To perform power conversion				X																								
	To perform power distribution			X																									
Environmental Protection	To protect from debris					X					X																		
Temperature control	To prevent frozen electronics						X				X																		
Mechanical (structure, mobility, etc.)	To Move along the surface							X	X																				
	To allow traversing any terrain									X																			
	To collect rock/soil/debris samples											X	X	X															
	To house electronics										X																		
	to store samples																												
Data collection	To collect visual data															X													
	To collect auditory data														X														
	to perform analysis on rocks/soil/debris																X												
	To collect environmental data (temperature, wind velocity, humidity, and dust)																	X											
	to collect data on atmospheric makeup																X	X											
	to collect crust elemental makeup data																X												
Data handling	To receive monitoring Data																		X										
	To validate monitoring data																		X										
	to elaborate monitoring data																			X									
	to store elaborated monitoring data																				X								
	to receive control data																		X										
	to validate control data																		X										
	to elaborate control data																			X									
	to store elaborated Control data																				X								
	to receive command data																		X										
	to validate command data																		X										
	to elaborate command data																			X									
	to store elaborated command data																				X								
Tele-communications	To transmit mission data																						X						
	To elaborate data to be transmitted																							X					
	to obtain data to be transmitted																					X							
	to transmit telemetries																						X						
	to elaborate telemetries																							X					
	to obtain telemetries																					X							
	to receive commands from earth																					X							
	to elaborate receive commands from earth																							X					
Navigation	to receive GPS coordinates																								X				
	to elaborate coordinates																											X	
	to transmit coordinates																												X
	to receive gyroscopic data																									X			
	to elaborate gyroscopic data																											X	
	to transmit gyroscopic data																												X
	to receive positioning data																										X		
	to elaborate positioning data																											X	
	to transmit positioning data																												X

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

3.4 Rover CAD Models and Stress tests

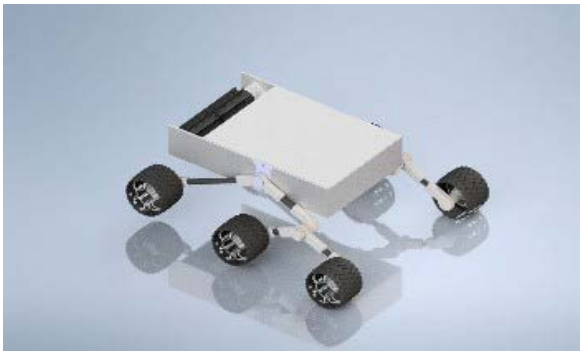


Figure 11. CAD render of v1 of rover

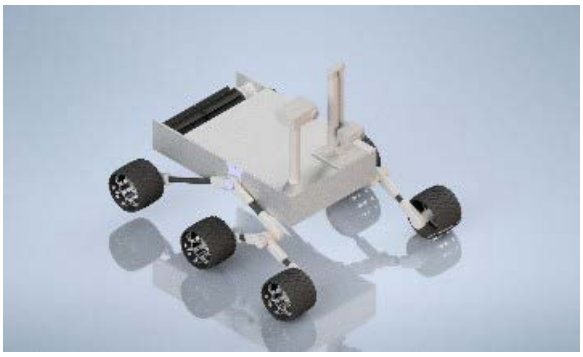


Figure 12. CAD render of finalized rover

The Computational designs, shown in Figure 11, show the first version of the CAD model of the rover. It missed components on top of the rover but displayed a relatively accurate look of the rover. Figure 12 shows when the camera and arm are attached to the rover. Both Figures 11 and 12 show the rover having a rocker-bogie chassis design. The MMRTG generator is at the back of the rover, and the cameras are at the front. The tires on these rovers are airless. The entire frame is theoretically made of aluminum. The computer would be located within the central box of the rover. Both of these designs are designed using Autodesk Inventor 2023.

$$F = mg \tag{2}$$

Equation 2. Force of gravity on Mars where F is force m is mass and g is the gravitational constant of acceleration on mars.

The force of gravity was the only factor taken into account for the stress tests for the rover models because that would be the bare minimum for traversing another planet. In this case, we are letting the computer handle the forces, and this equation is input into most of the stress tests taking place because we are only inputting gravity.

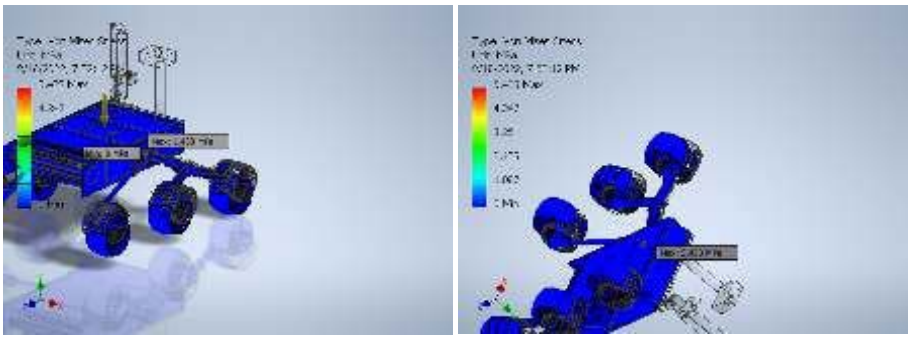


Figure 13. Von Mises Stress of Final Version of Rover.

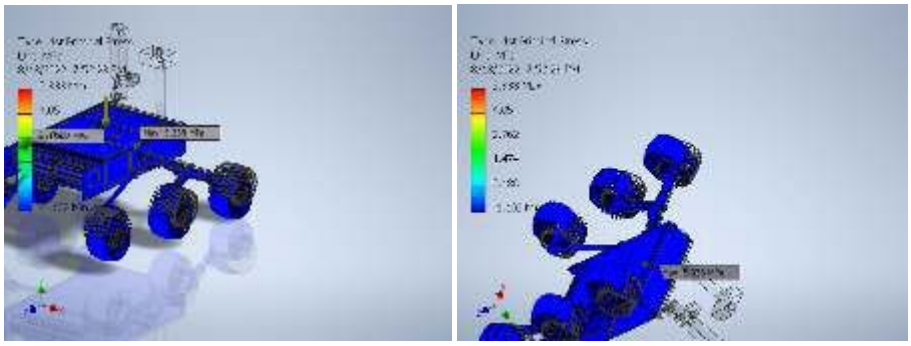


Figure 14. First Principle Stress of Final Version of Rover.

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

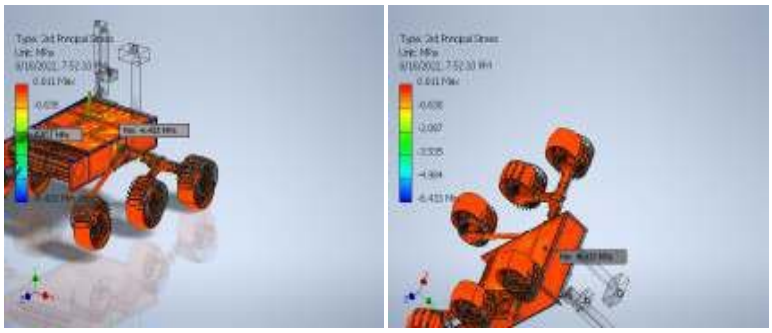


Figure 15. Third principle Stress of Final Version of Rover.

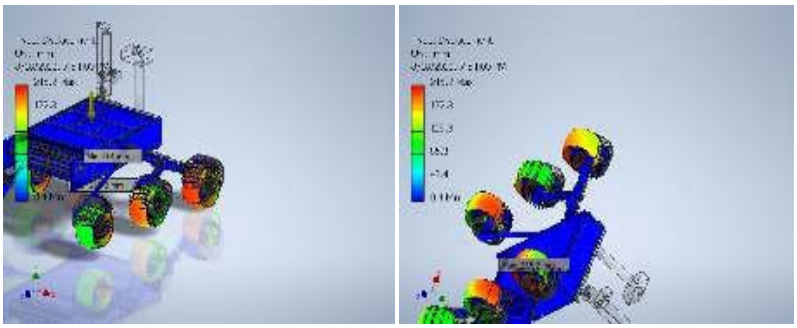


Figure 16. Displacement of Final Version or Rover.

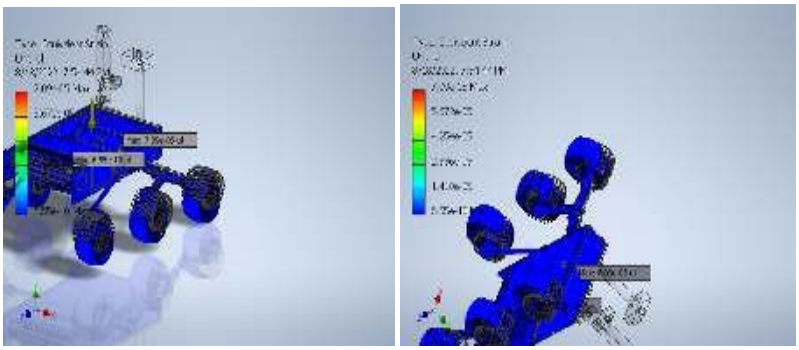


Figure 17. Equivalent Strain of Final Version or Rover.

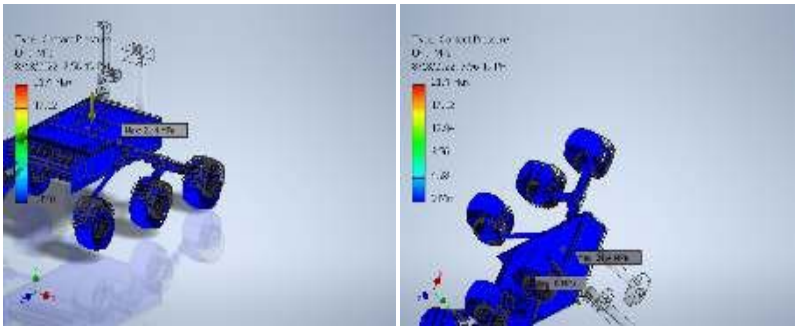


Figure 18. The Contact Pressure of the Final version of the Rover.

After adding more parts on top of the rover, 6 different stress tests were simulated using Autodesk Inventor. The Von Mises stress test results, shown in Figure 13, show a minimum value of 0 MPa, almost completely covering the outside of the rover. The places where the maximum value, 5.433 MPa, is shown are on the edge of the side plates and the inside. The Von Mises stress results, shown in Figure 24, have a minimum value of -1.102 MPa which also covers the rover except for a few spots on the top and bottom plate where it reaches a maximum of 5.338 MPa. From

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

the third principal stress test results, shown in Figure 15, we can see that the maximum is 0.811 MPa, which also covers the rover and lessens on the top and bottom plates. The minimum for these stress test results is -6.433 MPa and not most likely on the inside surfaces of the rover since it is not visible on the outside. The displacement stress test results, shown in Figure 16, show a minimum displacement of 0.4 mm and a maximum of 215.2 mm. The Equivalent strain test results, shown in Figure 17, show a maximum value of 7.09e-05 ul and a minimum of 6.55e-10 ul. The minimum value covers the rover's exterior, so the maximum must be on the inside. The Contact pressure test results, shown in Figure 18, show a maximum of 21.4 MPa and a minimum of 0 MPa. Like the equivalent strain tests, the minimum covers the outside while the inside shows the maximum.

4. Discussion

4.1 Rover Paths

Many factors could affect the rovers driving on Mars. To decrease the number of factors, we plan a course with the lowest change in elevation so the rover can explore without worrying about falling on its back. In Figures 9 through 12, the paths are chosen to get to a new area of soil in as straight a line as possible to analyze different soil types quickly. In the case of NE Syrtis, on Google Earth's view of Mars, there is a difference in color and number of craters west of the landing point. The rover landing point was constructed to cross different soils on the longitudinal lines and reach a completely different soil sample almost simultaneously. For the Jezero crater, The reasoning within this path is to reach the Jezero crater delta, the place most likely holds the answer to our mission questions, but not run into any craters while getting there. That is why the rover heads west first.

4.2 Decision Matrix Results

For the first question of the Decision Matrix, NES(Ne Syrtis) received a 5, and JC(Jezero crater) received a 7 because JC has more theoretical evidence of life on mars due to its geographical formations. The JC being a crater also shows that it could hold water, supported by the fact that there is a delta-like feature on the northern lip. For the second question, NES received a 7, and JC received a 3 because JC shows fewer rock types versus NES due to its volcanic history. For the third question, NES received a 7, and JC received a 5 because JC had only theorized water on it while NES had both water and volcanic activity. Due to the rich abundance of water theorized in each location, both locations received a 7 for the fourth question. For the fifth question, both locations received a 5 because there aren't plans to retrieve the rover samples from Mars, but they would both get samples once there. Question six also relates to the water content theorized in both places, so both received a 7. Question seven is about the maneuverability of the rover. Both places received a 5 because of many challenges in NES due to its geography, but launching a rover in JC locks the operable area inside the crater. The decision matrix ended with NE Syrtis only 4 points ahead of the Jezero crater, an insignificant difference. The best choice would be NE Syrtis because it is not a finite area unlike the Jezero crater. In the end, NE Syrtis would most likely have the best chances of finding evidence of life because of different soils, past volcanic activity, vast amounts of water, and clay deposits.

4.3 Rover CAD design

The rover design is the most crucial aspect of a rover's mission. The computation model has to reduce as many variables as possible and be able to travel the longest distance possible. To accomplish this, remove springs in place of a Rocker-bogie suspension system because it can traverse obstacles without having a single spring in it. Another variable that can be removed is the air inside the wheels because it risks popping and leaving the rover stagnant. (Rajan, 2022) Using airless wheels eliminates the risk of getting a flat tire. One part not developed fully is the design of the arm. It needs to have almost complete freedom to collect samples. The frame is aluminum, allowing the rover to have the most strength for the least density. By cutting the weight, the rover can hold more instruments. In the case

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

of this design, the MMRTG at the back is horizontal as opposed to on an angle on previous rovers, Curiosity, and Perseverance because it was the only generator file available to me on short notice.(Reed, 2018) The Rocker bogie chassis here is also not the best because where each axle is, there are supposed to be two motors that will move the wheel and any direction to achieve omnidirectionally. The wheels are from the Curiosity rover, which has passed through the test of time. By using the template of previous rovers, which has been tried throughout the years and shown long life, we can send rovers to Mars that would not become trash in a couple of years and be able to complete their missions within their lifetime.

4.4 Stress Test Results

The structure of the rover has to be strong enough to withstand the pressures on Mars. The simplest one to integrate is gravity.

In simplicity, Von Mises stress determines if a material will yield or fracture. In the Von Mises stress test results, shown in figure 13, the overall value throughout the rover is zero, telling us that the material chosen will not fracture under this pressure. Wherever the maximum value ends up, there will need to be more support in that area, especially on the inside.

First principal stress gives you the value of stress normal to the plane in which the shear stress is zero. Again in this test, shown in figure 14, the values throughout the rovers are zero indicating that the material will not break. The only places that are causes for concern within this are on the top where there is currently minimal support. This could be fixed by adding support structures throughout the central housing.

The third principal stress normally acts to the plane where shear stress is zero. It helps understand the maximum compressive stress induced in specific parts due to the loading conditions. The results of this test, shown in figure 15, are also similar to the first principal stress test results since the values hover around zero throughout the rover. Differing results are shown on the top paneling, indicating that more support is needed.

Displacement in a stress test is how far each specific part will move in relation to others and the forces acting on it. The displacement, shown in figure 16, of the rover, is the one part of these tests that are worrying because it seems that the wheels have fallen off. There could be a lot of circumstances that made the wheels not fully attach to the rover, but it is negligible because they are not fastened to the rover yet. This means that with proper assembly, the wheels would not fall off.

Equivalent strain is the limit for the values up to which the object will rebound and return to the original shape upon removal of the load. The equivalent strain, shown in figure 17, is almost negligible because it shows values of approximately 6.55e-10 ul. Wherever the maximum value ends up, there will need to be more support in that area, especially on the inside.

Contact pressure is the ratio of the normal load to the true contact area. In Figure 18, the contact pressure is also negligible since it's close to zero. Although these tests only use one factor to determine if they are concerns, it's useful to determine the base structural integrity of the rover, so adding more parts, tools, or scientific instruments would be easier.

4.5 Functionality Matrix Results

Table 3 contains information about the functionality matrix. The electrical power subsystem is to supply the entire rover with electricity to function. The environmental protection subsystem is vital to protect the robot from the harsh conditions of Mars. This subsystem works in tandem with the temperature control subsystem, ensuring the electronics function throughout the temperature fluctuations. The mechanical subsystem deals with all the raw robotics parts like structure and mobility. The data collection subsystem is arguably the most important because it's the one subsystem that gathers data on Mars. With the data collection, there needs to be a way to handle the data on board the rover. To send data to earth, the rover needs a telecommunications subsystem. To traverse Mars, a navigation subsystem is utilized so the rover and humans know where the rover is going. Without even one of these categories of electronics or tools, the mission would fail because it would be less than the bare minimum.

Vol. 2023 (2) 13 – 27
ISSN 2688-3651 [online]

5. Conclusion

In the end, Mars is just another planet in our solar system. Although widely unexplored, from the few times humans were able to explore Mars, there were still many discoveries that could help humans understand its history. Since humans can't get to Mars, the only way is to send a vehicle (rover) there to gather information for us. By sending more rovers onto the red planet, the more we can understand it at ground level. With more conceptual outlines of missions for these rovers, the less time it takes to understand the planet. There needs to be years of planning and more time building the rover itself. Ensuring everything will work beforehand is the key to making this mission succeed because once the engineers load the rover into the rocket, there is no going back to work on it. There are many years of plotting the path, designing the systems, perfecting the technology, building the equipment, and testing on earth, but there is only one chance to get to Mars. In conclusion, by preparing plans to send rovers to mars, we can reduce the time it takes to understand the planet and further advance our technology.

I would have liked to start from scratch and make the entire thing myself, but because of time constraints, it was improbable. I relied on past 3d models that would not have been what I was looking for. In the future, I recommend researchers find the most optimal path for the rover, hypothesize the robot landing the rover, make the rover drivetrain more mobile, and create a CAD design with all the necessary tools depicted on the outside.

References

Beck, K. (2021, August 14). Every rover, ranked by distance traveled on the moon and Mars. Mashable. https://mashable.com/article/moon-Mars-rover-distance-driven

Cresto Aleina, S., Viola, N., Antonietta Viscio, M., Stesina, F., & Ferris, S. (2016, July). Reusable space tug concept and mission. Acta Astronautica. https://www.researchgate.net/publication/304914910_Reusable_space_tug_concept_and_mission

Croston, T. (2014, February 25). Mars Curiosity wheel. Grabcad.com. Croston, T. (2014, February 25). Mars Curiosity wheel. Grabcad.com.

Dubov, D. (1997, July 8). About the Rover. Mars.nasa.gov. https://Mars.nasa.gov/MPF/rover/about.html

Lakdawalla, E., & McGovern, P. (2020, June 24). Every Mars landing attempt as of June 2020. The Planetary Society. https://www.planetary.org/space-images/every-Mars-landing-attempt-ever

Liu, J., et al. (2021, December 6). Geomorphic contexts and science focus of the Zhurong landing site on Mars. Nature Astronomy. https://doi.org/10.1038/s41550-021-01519-5

Journal of Research High School

Vol. 2023 (2) 28 – 35
ISSN 2688-3651 [online]

A Study on the Effects of Physical Appearance on Psychological Status and Cosmetic Surgery Preference: Differences by Generation

Gueyin Kim[1*]

[1]Hankuk Academy of Foreign Studies, Yongin-Si, Republic of Korea
*Corresponding Author: 17018@hafs.hs.kr

Advisor: Joohyun Lee, juue54@hanmail.net

Received February 25, 2023; Revised June 4, 2023; Accepted, June 28, 2023

Abstract

Physical appearance is not only a factor that influences an individual's attractiveness, but it can also have an impact on one's psychological status. In order to determine the correlation between one's physical appearance and self-esteem, a survey on participants of different genders and various generations was conducted. The questionnaire consisted of a Rosenberg self-esteem scale (RSE) and 14 additional items related to respondents' self-perception, social perceptions towards appearance, personal experiences of cosmetic surgery, and viewpoints on cosmetic surgery. Correlation analysis showed that the responses of each section of the questionnaire indeed had a significant correlation, meaning that the respondent's physical attractiveness, self-esteem, and thoughts on plastic surgery were associated. Through gender-specific and age-specific analysis, the results of different participant groups were compared. Results indicated that generally, the older age group exhibited a weaker correlation between their appearance and self-esteem compared to the younger age groups. This phenomenon was discussed in detail and possible explanations for this phenomenon were presented. First, a tendency of older generations to prioritize values other than superficial standards was explained. When forming a self-perception, they were more likely to have standards such as social success and wealth rather than only physical attractiveness. Secondly, a possibility of the decline of cognition and perception caused by brain aging was suggested as a reason for the elderly to be less sensitive about their appearance compared to younger individuals.

Keywords: Physical attractiveness, Self-esteem, Self-perception, Cosmetic surgery

1. Introduction

Appearance is generally understood as a factor that influences the attractiveness of an individual. Physical appearance is closely related to attractiveness and therefore is associated with one's life and perceptions. Humans naturally tend to pursue physical attractiveness and admire those that coincide with their standards of beauty. Studies show that behavioral preference for beauty is inclined from our innate tendency towards beauty (Mo, et al., 2016). Therefore, the preference towards aesthetic objects should not be viewed as a phenomenon caused by environmental aspects; it can be more accurately explained as a natural consequence of human psychology and behavior. When one evaluates their own attractiveness, they tend to form a self-perception. Self-perception refers to an individual's view of one's own physical, mental attributes. Depending on the values one prioritizes, the standards in assessing their own value vary as well.

As there were aspects of the association between psychological factors and physical attractiveness, it seemed meaningful to analyze the relationship between one's psychological status and physical appearance. In addition, diverse viewpoints on cosmetic surgery and undergoing surgical procedures to change or improve one's appearance also seemed to be worth examining. It was hypothesized that respondents from older age groups would have higher self-esteem and be less sensitive toward physical appearance, resulting in less willingness to go through plastic surgery.

It is largely understood that physical attractiveness and self-esteem are closely correlated. It is observed that improving physical appearance improves someone's attitude and self-esteem, and therefore physical attractiveness can be viewed as a major factor that has a substantial influence on self-esteem (Patzer, 1997). Those who establish a more

Vol. 2023 (2) 28 – 35
ISSN 2688-3651 [online]

positive self-image, are more likely to have a stable psychological state and therefore achieve higher confidence. Having a positive image of oneself leads to higher self-esteem and allows people to have mental clarity and form a more realistic viewpoint on their strengths and weaknesses. These factors all contribute to the success of an individual, meaning that having a positive self-image and high self-esteem is essential for success and achievement as well (Barclay, 2023). Studies on the self-perception of people from different genders and various age groups show that there is a difference between individuals in terms of how one perceives their own physical appearance. Age-specific studies determine how participants form a different self-image depending on their age. A cross-sectional study designed to observe the characteristics of adolescents' self-perceptions (Abdo, et al., 2023) concludes that the vast majority of participants from younger generations exhibited a negative self-perception of their physical appearance. Studies indicate that approximately half of adolescents experience low self-esteem (Hirsch and Dubois, 1991). Forming a negative self-image and having low self-esteem, especially at a young age, may lead to further psychological issues. Conversely, findings indicate that older generations tend to have higher self-esteem and self-efficacy compared to those who are younger (Dietz, 1996). This indicates that those who are older tend to form a more stable, confident psychological state.

The objective of this study is to investigate the differences between age groups in terms of the correlation between physical appearance and psychological status. Differences between generations are frequently discussed, and people thought that there might be a difference in the standards and viewpoints of people depending on age as well. We specifically focused on the aspect of one's viewpoints toward appearance, for discussion about the standards of beauty is always a controversial topic. It was thought that the beauty standards of generations would differ, and therefore, their assessments of their own appearance would differ as well.

In this report, we conducted a survey with participants of different genders and various age groups. The questionnaire was formed of 4 sections. Through the analysis of the responses, the correlation between the sections was observed. By making comparisons between the respondent groups, the correlation between one's appearance and self-esteem was determined. The age-specific analysis was especially studied in detail, and an explanation for the results was presented. We hypothesized that those in the older age group would exhibit a weaker correlation between their self-esteem and physical appearance compared to younger individuals.

2. Materials and Methods

This survey consisted of 24 questions regarding self-esteem, self-image and self-perception, and cosmetic surgery. The questionnaire was designed to enable an estimation of the participants' self-esteem and self-image to be made. In addition, personal opinions about plastic surgery were collected. Responses to each of the sections were collected in order to make a correlation analysis. Respondents were of different gender and various age groups.

2.1 Participants

The sample was formed by volunteer sampling. The questionnaire was sent by email to participants, and those who volunteered to participate in the survey were selected as the sample group. The nationality of the participants is South Korean, and email recipients who voluntarily completed the survey were selected as the sample group.

Table 1. Demographic characteristics

Age	Women	Men	Total
10~19	83	47	130
20~29	47	27	74
30~39	24	5	29
40~49	15	11	26
Over 50	20	15	35
Total	189	105	294

2.2 Questionnaire

The questionnaire was intended to evaluate self-esteem, perception of one's own appearance, and opinions on going through surgical procedures to modify one's appearance. Question 1~10 (section 1) is based on the Rosenberg self-esteem scale (RSE) (Rosenberg, 1965), a measure consisting of 10 items that were intended to diagnose the level of self-esteem of

Vol. 2023 (2) 28 – 35
ISSN 2688-3651 [online]

respondents. Question 11~15 (section 2) is related to the participant's self-perception of their own appearance. Question 16~20 (section 3) is related to the participant's thoughts on social perceptions towards appearance, and question 21~24 (section 4) asks for personal opinions about cosmetic surgery. Answer choices are numbers 1 to 5, 1 meaning "strongly disagree," 2 meaning "disagree," 3 meaning "neutral," 4 meaning "agree," and 5 meaning "strongly agree." Among the 24 questions, question 2, question 5, question 6, question 8, question 9, and question 15 were negative questions to check the consistency of respondents. In order to analyze the data more effectively, the responses to these negative questions were converted as 1 to 5, 2 to 4, 4 to 2, and 5 to 1. Responses to these questions were classified into each section for the efficiency of analysis. Questions on the survey are shown in Table 2.

Table 2. Questionnaire

1	Overall, I am satisfied with myself.
2	Sometimes, I think that I am good at nothing.
3	I think of myself as someone who has many strengths.
4	I can complete tasks as well as other people do.
5	I think that I have nothing to be proud of.
6	Sometimes, I view myself as useless.
7	I think that I am a valuable person.
8	I think that I do not have enough respect towards myself.
9	Overall, I think of myself as a failure.
10	I have an optimistic viewpoint toward myself.
11	I think that I have a decent appearance.
12	People around me often compliment my appearance.
13	I am satisfied with my appearance.
14	People say that my appearance is likeable.
15	I feel dissatisfied about my physical appearance.
16	The most influential thing among appearance, wealth, and education level is appearance.
17	I think that it is easier for someone who is physically attractive to achieve social success compared to someone who isn't.
18	I think that an attractive appearance is helpful in everyday life.
19	I think that including a photo of the applicant in a resume is better than not including one.
20	I think that it is unfair that viewpoints towards someone differ depending on his or her appearance.
21	I have gone through plastic surgery before.
22	I am willing to undergo plastic surgery in order to change my appearance.
23	I think that plastic surgery can help improve one's quality of life.
24	I have a positive viewpoint toward people who changed their physical appearance by plastic surgery.

2.3 Procedure and analysis methods

An email invitation with the survey link was sent to the participants on November 8, 2022. A total of 294 answers were collected by November 14, 2022. Statistical analysis, including estimating the mean, standard deviation, standard error, and correlation coefficient of the data, was performed. By evaluating the data on these questions, it was possible to estimate the correlation between each of the factors. The answers of the respondents were analyzed for a general analysis. In addition, gender-specific and age-specific comparisons were also made as well. The analysis of the results was confirmed by the Korea Policy Research Group.

IBM SPSS Statistics 23.0 was used for statistical analysis. The Pearson formula was used to calculate the correlation coefficients. For the age-specific analysis, the null hypothesis was that there is no difference in the mean values between different age groups. For the gender-specific analysis, the null hypothesis was that there is no difference in the mean values between different gender groups. A 95% confidence interval was used, and therefore, the p-value for the analysis was 0.05.

3. Results

The responses of all the participants were analyzed by each section of the questionnaire. Results are shown in Figure 1. Based on the analysis of section 1, it is observed that the questions with the highest average are "All in all, I am not inclined to feel that I am a failure." (4.27), "I feel I do have much to be proud of" (4.10), "I am able to do things as well as most other people" (4.09), "I feel that I'm a person of worth, at least on an equal plane with others" (4.06) had the highest average among the ten questions. In contrast, "I do not think I am no good at all at times" (3.44), "I wish I have enough respect for myself" (3.55), and "On the whole, I am satisfied with myself" (3.78) had the lowest average. The results of section 2, which are related to the self-image of

Vol. 2023 (2) 28 – 35
ISSN 2688-3651 [online]

the participants, show that "I am satisfied with my appearance" (3.45), "People often compliment my appearance" (3.28), "I think of my appearance as attractive" (3.17) exhibited the highest average among the five questions. The results in section 3, which concern social perceptions about appearance, show that "Having an attractive appearance makes life easier" (4.28), "People who are good-looking are more likely to succeed than those who aren't" (3.60) comparatively had the highest average among the five questions. The answers to the question of whether participants experienced plastic surgery before show that 58 participants (19.7%) had an experience undergoing cosmetic surgery whereas 236 participants (80.3%) never went through plastic surgery. Answers to section 4, exhibiting the participants' opinions on cosmetic surgery, show that "I agree that plastic surgery can improve one's life" (3.55) has the highest average among the three questions related to personal opinions on cosmetic surgery. For the question asking whether the respondent has experienced cosmetic surgery before (question 21), 58 people (19.7%) answered that they have gone through cosmetic surgery whereas 236 people (80.3%) answered that they haven't.

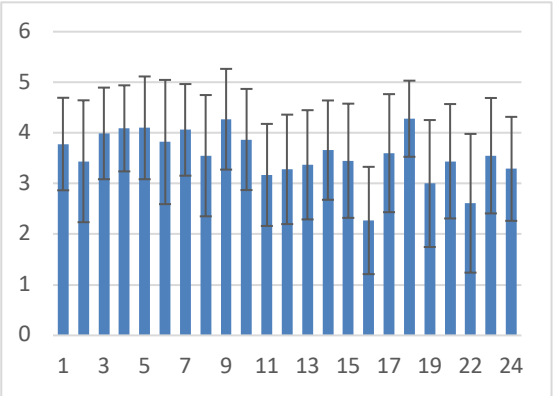


Figure 1. Analysis of answers to the questionnaire

Correlation analysis shows that there is a clear correlation between one's self-esteem and self-image. In other words, participants who had a more positive self-image of themselves tended to exhibit higher self-esteem. A partial correlation between self-esteem and opinions toward cosmetic surgery is also confirmed. Those who had lower self-esteem showed a slight tendency to choose plastic surgery as a solution. Also, participants who answered that they perceived themselves as someone of less worth compared to others were inclined to feel that surgical procedures could help them strengthen their self-esteem.

4. Discussion

4.1 Age-specific analysis

Age-specific comparison in each section shows that generally, those who are older tend to have higher self-esteem, form a more positive self-image, and think of the importance of appearance less than those who are younger. In section 1, the question with the most drastic change in responses across different age groups was "Sometimes I feel useless." In the following sections, the answers to the questions such as "People often compliment my appearance," "Including a photo of the applicant in a resume is better than not including one," and "I have a positive viewpoint towards people who have gone through plastic surgery," showed the most rapid change according to age group, respectively. Conclusively, the results show that the answers of the participants in each section exhibit a significant difference between different age groups.

4.2 Gender-specific analysis

Gender-specific analysis shows that generally, men tend to form a more positive self-image. In section 1, it is shown that men exhibit higher self-esteem compared to women. The average value of the responses of men was 3.92, whereas the average of the responses of women was 3.70. Also, as shown in section 2, male respondents showed a more positive perception of their own appearances. For the question that showed the largest difference between gender, "I think that I have a decent appearance," the average response of male participants was 3.35, which is larger than 3.07, the average response of female participants. Through the comparison made between responses to section 4 of both genders, it can be concluded that female participants have a more positive viewpoint towards cosmetic surgery compared to male participants. Also, female participants were more inclined to go through surgical procedures than male participants and generally displayed a more affirmative attitude toward plastic surgery. To the question "I am

Vol. 2023 (2) 28 – 35
ISSN 2688-3651 [online]

considering undergoing plastic surgery in the future," the average response of women was 2.88, which is larger than 2.11, the average response of men. Also, for the question "I have a positive viewpoint towards someone who has gone through surgical procedures to change or improve one's appearance," the mean value of the responses of female participants (3.43) was higher than that of male participants (3.05). Conclusively, results show that men generally form higher self-esteem and have a more positive self-perception about their appearance.

4.3 General Analysis

The hypothesis concerning differences between generations is generally supported by the results above. In the figures below, a larger response number indicates a more positive view of oneself, whereas a smaller response number signifies a comparatively negative view of oneself. However, results show an unexpected trend (Figure 2). The responses to section 2, which is related to the self-perception of the participants towards their own appearance, exhibit an increase from age group 10~19 to age group 20~29, and decline sharply in age group 30~39. Then, it shows an incline again through age groups 40~49 and those over 50. The age group of those between 20~29 years old exhibited a significantly positive self-image, whereas the age group 30~39 showed the lowest response average. This phenomenon suggests that those in the age group 20~29 tend to have a positive self-image and evaluate their appearance with a more positive attitude compared to age groups 10~19 and 30~39. This may be explained that because people in age 20-29 are generally more physically attractive due to their youth, those in their twenties view their appearances positively. However, those in the age group 10~19 tend to show lower self-esteem and a more negative self-perception compared to those in other age groups. This suggests that teenagers are not yet psychologically mature and therefore are more sensitive about their appearance and tend to form a negative self-image of their physical appearance generally. The sharp decline in the age group 30~39 suggests that people in their thirties recognize that the aging process has begun and perceive themselves as not as beautiful compared to younger age groups. However, for respondents in the age groups 40~49 and 50 and older, their self-perception appeared to get more positive again. This incline can be explained that as people age and become emotionally mature, they prioritize values such as social success, wealth, and respect over superficial standards such as physical attractiveness. Also, they tend to be less sensitive when viewing their own appearance, which leads to a more positive self-evaluation of their physical appearance. Within the same age group, female participants were more likely to exhibit a smaller average response, or a more negative self-perception.

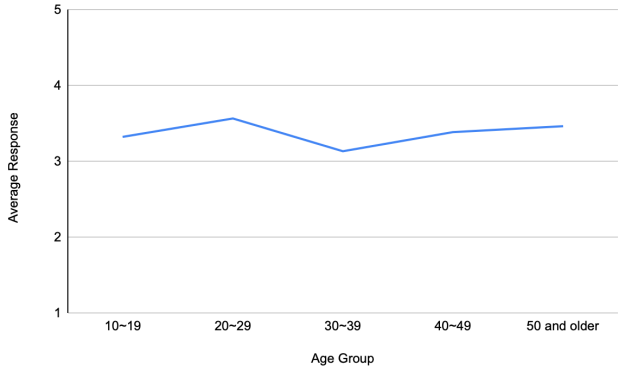


Figure 2. Trend of average responses to questions related to self-perception on appearance (section 2)

Self-perception and self-image responses showed a strong correlation to the responses to section 4, the section asking respondents' viewpoints toward plastic surgery. The trend of the average responses of each age group exhibits an ascent through age groups 10~19 and 20~29 to age group 30~39 and declines through age groups 40~49 and those over 50 (Figure 3). It is observed that those in their thirties were most likely to undergo cosmetic surgery in the future. Those in the age group 30~39 responded with the most positive attitude towards cosmetic surgery and were also willing to experience surgical procedures. In other words, the age group that showed the most negative

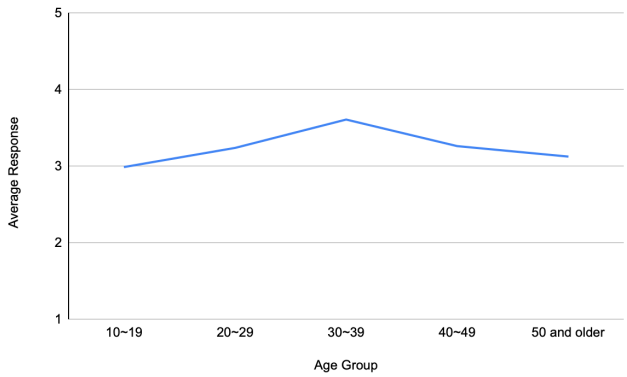


Figure 3. Trend of average responses to questions related to viewpoints on cosmetic surgery (section 4)

Vol. 2023 (2) 28 – 35
ISSN 2688-3651 [online]

self-perception of their physical appearance was most inclined to go through plastic surgery. Those in the age groups 40~49 and 50 and older were less eager to change their appearance compared to the younger age groups. This can be explained that although those over 40 are generally less attractive than those in the age group 30~39 due to aging, the older generations are not as keen about their appearance and therefore are less enthusiastic about using plastic surgery to change or improve their physical attractiveness. Within the same age group, female participants were more likely to have a positive viewpoint on plastic surgery compared to male participants.

As reported, it can be concluded that the older age groups generally had a more positive perception of their appearance and found that the relation between appearance and psychological stability such as self-esteem was not as strong as the younger participants. This phenomenon can be interpreted in two ways. First, it could be explained that the older generations tend to prioritize values such as social status and wealth, other than appearance. Another explanation could be that people with older age tend to be loose in perception as the brain ages, leading to being less sensitive towards their physical appearance.

Values and standards that relate to one's self-esteem vary between generations. A study (Antonucci, et al., 1979) reports that those of various ages showed a significant difference in which values they prioritized the most, especially those that are work-related. Those from the oldest generation rated the values related to their social success and work higher than the youngest generation. This suggests that when evaluating one's self-image and self-esteem, the younger generations are more likely to think of appearance as an essential value. The older population regards other values such as success, ambition, and financial stability as well as physical attractiveness. However, younger generations who have a comparatively narrow perspective of life may depend solely on their appearance when estimating their own values. The differences between generations can be viewed as a result of psychological maturity. A study (Icenogle, et al., 2019) states that there is a difference between legal maturity and psychological maturity. This "maturity gap" is formed because whereas the adult level of cognitive capacity is reached around the age of 16, the adult level of psychosocial maturity can be reached after the age of 18. This demonstrates why those who are legally "adults" can still experience psychological problems related to maturity and confidence. Therefore, it is possible for those over the age of 18 to vary in levels of psychological and emotional maturity. Psychological maturity is a term used to describe the level of emotional and psychological development of an individual. It is a broad concept that encompasses a wide range of traits, including self-awareness, self-regulation, empathy, and emotional intelligence. Maturity and self-esteem are both complex and multi-faceted concepts that are closely related. There is a strong correlation between maturity and self-esteem, as maturity can greatly influence self-esteem. As individuals mature and develop, they develop a more positive self-image and self-concept and tend to have a more positive sense of self-worth. Additionally, as individuals mature, they are better able to handle challenges and setbacks. It becomes easier to be resilient and cope with difficult situations as they develop a better understanding of their own strengths and weaknesses and can use that understanding to make positive changes in their lives. All the factors contribute to higher self-esteem, indicating that individuals who are more mature are likely to have better psychological stability. Similarly, immaturity can lead to lower levels of self-esteem. Immature individuals may have a negative self-image and self-concept, and may doubt themselves and their abilities, leading to lower levels of self-esteem. Although maturity levels may vary even in the same generation, it is widely accepted that those of an older age are more likely to develop higher levels of psychological and emotional maturity. A contributing factor to the lower psychological stability of young generations can be viewed as a result of social media usage. Those of the younger generation tend to be exposed to social media more actively than the older generation. A study assessing the association between social media usage and self-esteem (Valkenburg, et al., 2021) reported a significant negative relationship between social media usage and self-esteem of adolescents was examined, meaning that participants who spent more time on social media were inclined to exhibit lower levels of self-esteem. This provides an explanation for younger participants to show lower levels of self-esteem and be more sensitive to how self-images are perceived by others than the older participants.

Another contributing factor to this phenomenon is the impact that the aging of the brain has on cognition and perception. It is shown that there is a significant interrelation between perception and cognition, meaning that poor cognition can cause perception skills to deteriorate as well (Roberts and Allen, 2016). Perceptual decline may not only cause difficulties in mastering perceptual tasks, but it may also result in impoverished sensitivity in the process of perception. In addition, as well as cognitive and perceptual decline, the sensitivity to objects also is shown to worsen. Sensory impairment includes a decline in visual and auditory acuity. A deficit in cognition, perception, and sensory

Vol. 2023 (2) 28 – 35
ISSN 2688-3651 [online]

acuity may affect how one perceives their own physical appearance and form a mental image of their looks. Those who are younger tend to be keener about the flaws they spot in their physical selves, causing them to establish comparatively less confident self-images. According to a study on adolescents' self-esteem (Harter and Whitesell, 2001), adolescents are more self-observant and self-reflective, meaning that they pay more attention to themselves and their physical image than older people do. They tend to apply more strict and idealistic standards on themselves. Conversely, those from older generations become less sensitive in terms of viewing and analyzing their physical appearance. This may lead to a more positive perception of oneself since one becomes more insensitive about the subtle aspects, including minor flaws, of their physical body.

5. Conclusion

An individual's physical appearance has been found to impact both self-esteem and one's willingness to go through cosmetic surgery. Generally, those with lower self-esteem tend to form a more negative self-image and are more likely to take the risk of doing plastic surgery. Results show that this degree of correlation differs between various participant groups. Our age-specific analysis shows that respondents of the older age groups exhibited a weaker degree compared to those of the younger age groups and explanations for this phenomenon were suggested. Deteriorated cognition as a result of brain aging and emotional maturity were presented as possible explanations. Conclusively, it can be suggested that those of younger age should make an effort to form a positive self-perception for their mental health.

References

Abdo, Z. A., Seid, S. A., & Woldekiros, A. N. (2023). Self-perception of physical appearance of adolescents and associated factors in Addis Ababa, Ethiopia. *PLOS ONE*, *18*(1). https://doi.org/10.1371/journal.pone.0281202

Antonucci, T., Gillett, N., & Hoyer, F. W. (1979). Values and self-esteem in three generations of men and women. *Journal of Gerontology*, *34*(3), 415–422. https://doi.org/10.1093/geronj/34.3.415

Dietz, B. E. (1996). The relationship of aging to self-esteem: The relative effects of maturation and role accumulation. *The International Journal of Aging and Human Development*, *43*(3), 249–266. https://doi.org/10.2190/bhxn-xqm2-rmab-hd4p

Harter, S., & Whitesell, N.R. (2001). On the importance of importance ratings in understanding adolescents' self-esteem: Beyond statistical parsimony

Hirsch, B. J., & Dubois, D. L. (1991). Self-esteem in early adolescence: The identification and prediction of contrasting longitudinal trajectories. *Journal of youth and adolescence*, *20*(1), 53–72. https://doi.org/10.1007/BF01537351

Icenogle, et al., 2019. Adolescents' cognitive capacity reaches adult levels prior to their psychosocial maturity: Evidence for a "maturity gap" in a multinational, cross-sectional sample. *Law and Human Behavior*, *43*(1), 69–85. https://doi.org/10.1037/lhb0000315

Mo, et al., 2016. Natural tendency towards beauty in humans: Evidence from binocular rivalry. PLOS ONE, 11(3). https://doi.org/10.1371/journal.pone.0150147

Patzer G. L. (1997). Improving self-esteem by improving physical attractiveness. *Journal of esthetic dentistry*, *9*(1), 44–46. https://doi.org/10.1111/j.1708-8240.1997.tb00915.x

Roberts, K. L., & Allen, H. A. (2016). Perception and cognition in the ageing brain: A brief review of the short- and long-term links between perceptual and cognitive decline. *Frontiers in Aging Neuroscience*, *8*. https://doi.org/10.3389/fnagi.2016.00039

Rosenberg, M. (1965). Rosenberg self-Esteem Scale. *PsycTESTS Dataset*. https://doi.org/10.1037/t01038-000

Tim Barclay, P. D. (2023, January 17). *How does physical appearance affect self-esteem?* Innerbody.

Vol. 2023 (2) 28 – 35
ISSN 2688-3651 [online]

https://www.innerbody.com/relationship-between-body-image-and-self-esteem

Valkenburg, et al., 2021. Social Media use and adolescents' self-esteem: Heading for a person-specific media effects paradigm. *Journal of Communication*, *71*(1), 56–78. https://doi.org/10.1093/joc/jqaa039

Journal of Research High School

Vol. 2023 (1) 36 – 39
ISSN 2688-3651 [online]

Mechanics of HTLV - 1 Driven ATL

Shivam Mohanty[1] *

[1]Round Rock High School, Round Rock, TX USA
*Corresponding Author: shivam.mohanty08@gmail.com

Advisor: Natalia Khalatyan, natalia.khalatyan@gmail.com

Received November 21, 2022; Revised May 15, 2023; Accepted, June 8, 2023

Abstract

HTLV-1 (Human T-lymphotropic virus 1) is a retrovirus that can lead to the development of Adult T-cell leukemia/lymphoma (ATL), a non-Hodgkin's lymphoma, after a long latency period. Despite its low incidence rate, ATL is difficult to treat and has a poor prognosis with a median survival rate of 8 months and a 4-year survival rate of 12%. Recent research has suggested that ATL is host-driven, with low genetic diversity across endemic regions, challenging the previous assumption that ATL is solely caused by viral proteins. Blood sample studies from infected patients have revealed low genetic diversity, indicating the involvement of host genetic factors in ATL development. Host cancer genetic markers have also been found near the integration sites of the viral genome. This review highlights the correlation between patient genetics and HTLV-1 driven ATL, and how understanding the role of genetic markers and genes can lead to the development of new treatments or vaccines to improve patient outcomes.

Keywords: HTLV-1, Leukemia, ATL, Cancer, Viral Genome

1. Introduction

HTLV-1(Human T-Lymphotropic virus) is a non-Hodgkins retrovirus. Discovered in 1980 by Robert C. Gallo, HTLV-1 was the first retrovirus to have been discovered as described by Vahlne. ATL (Adult T-Cell Leukemia) is cancer that develops from HTLV-1's infection of T-Cells.

One of the most well-known retroviruses is HIV-1 (Human Immunodeficiency Virus), which was first identified in 1983 by LUC Montagnier's team at the Pasteur Institute of Paris. Similar to HTLV-1, HIV is responsible for causing AIDS (Acquired Immunodefiency Syndrome) an autoimmune disease. HIV is most transmitted through sexual contact between individuals. As of right now, there is no vaccine available for either HTLV-1, or HIV. This means that individuals infected with either virus will remain infected for life. As with other retroviruses, HIV uses reverse transcription to insert its RNA genome.

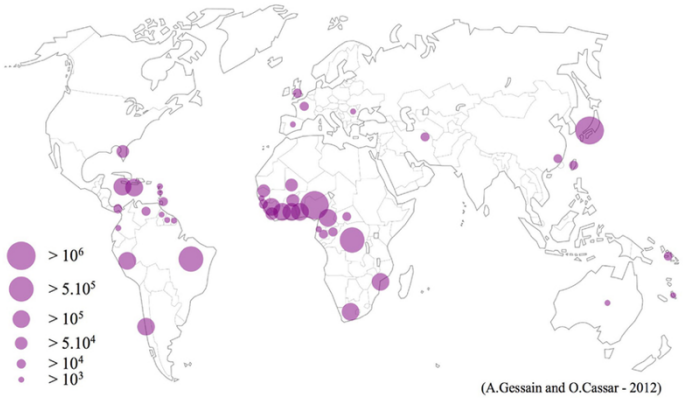


Figure 1: Geographical Distribution of HTLV - 1 cases (Gessain and Casser, 2012)

The global distribution of HTLV-1 subtypes can provide insight into the spread of this retrovirus across distinct regions of the world. As illustrated in Figure 1, the Japanese subgroup of HTLV-1 has spread to eastern China and Korea, while the African subgroup has been observed to mutate and diversify within specific regions of Africa. The higher prevalence of HTLV-1 in Africa, especially among men with multiple wives, suggests that cultural practices may contribute to the spread of the virus. This is important considering that the male-to-female transmission rate is 60.8%, while the female-to-male transmission rate is merely 0.4% shown in recent studies by Nunes et al. (2019).

Similarly, the South American subgroup has been observed to travel to North America, showcasing the influence

Vol. 2023 (1) 36 – 39
ISSN 2688-3651 [online]

of physical geography on HTLV-1.

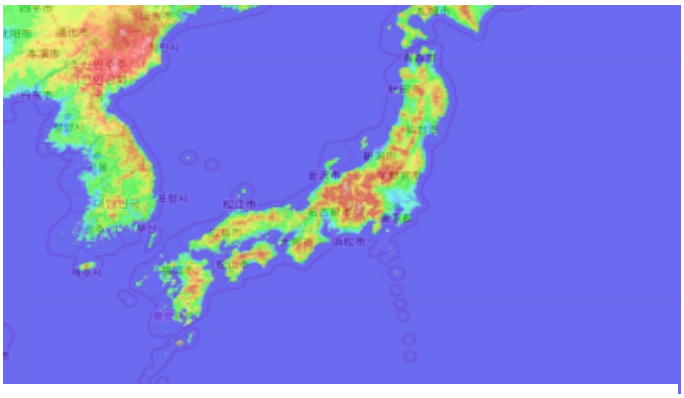


Figure 2: Elevation Heat Map of Japan; Japanese macaques are known to inhabit the mountainous regions of Japan (Yoshikawa et al., 2014).

Furthermore, the discovery of a primate counterpart to HTLV-1, known as STLV-1 (Simian T-Cell Leukemia Virus), has led researchers to hypothesize that the virus may have been transmitted to humans through contact with infected primates according to Jègado et al. (2019). This theory is supported by evidence that the southern region of Japan, where HTLV-1 is prevalent, has a higher elevation level as shown in Figure 2 below. This makes it a potential habitat for primates.

Similar to African hunters who hunt red colobuses in East Africa, which are known to carry STLV-1 (Leendertz et al., 2004), it is common for farmers or hunters in Japan to come into contact with these primates. Monkeys were known to terrorize villagers, leading to retaliation and the killing of the monkeys (Ohkuma & Tsuji, 2009). The African subtype of HTLV-1 is prevalent in central and East Asia, where red colobuses are found (Gessain et al., 2013). It is possible that the virus was spread through food preparation, like SARS-CoV-2, which is believed to have originated from horseshoe bats (Mallapaty, 2020).

Alongside STLV-1, a parasite known as *Strongyloids steracoralis* can contribute to a higher risk of developing ATL compared to only having HTLV-1 (Carvalho et al., 2015; Gotuzzo et al 2015). The mechanism by which *Strongyloids steracoralis* accelerates the development of ATL is not fully understood, but it is thought to involve the parasite's ability to suppress the host's immune system and increase HTLV-1 proviral load (Ito et al., 2014; Koga et al., 2005). Therefore, it can be hypothesized that the parasite plays an important role and should not be overlooked when considering the prevention and management of ATL in endemic areas.

In this review, we will explore the significance of HTLV-1 and ATL, which have a considerable impact on public health worldwide. Even though less than 5% of individuals infected with HTLV-1 develop ATL, the disease is highly aggressive and has a low survival rate. Additionally, there is currently no established treatment or vaccine available, making the development of effective interventions a pressing concern. By highlighting the correlation between patient genetics and HTLV-1 driven ATL, this review aims to emphasize the importance of understanding the role of genetic markers and genes in the development of ATL. By doing so, we can identify potential therapeutic targets and develop new treatments or vaccines to improve patient outcomes.

2. Discussion

In central Africa, there are many diseases such as Malaria, Yellow fever, Dengue, and Ebola. This high prevalence indicates that genetics must play an important role in developing ATL. A lot of these other diseases are rampant in this region due to many factors, but genetics plays a big role. It could be that the genetics of many people in Central Africa and Japan could be suffering from a malignancy in genetics. This would make sense because HTLV-1 originates from both countries, A gene could be passed on from a previous generation that may cause vulnerability in the genes of their descendants. Recently, a study has been found that states that integration sites of HTLV-1 are near genetic cancer markers.

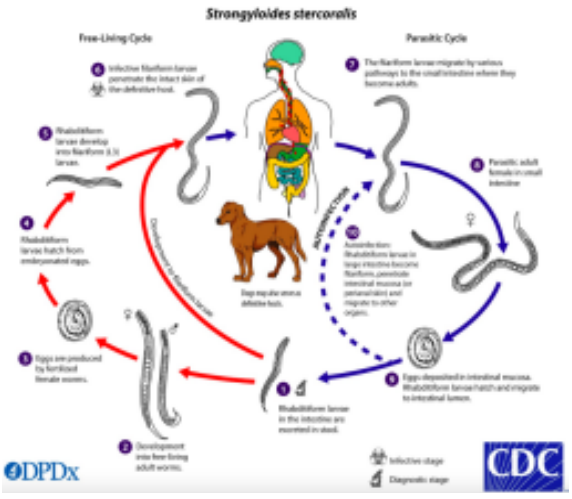


Figure 3: Strongyloids Stercoralis life cycle (https://www.cdc.gov/dpdx/strongyloidiasis/modules/Strongyloides_LifeCycle_lg.jpg)

Vol. 2023 (1) 36 – 39
ISSN 2688-3651 [online]

(Rosewick, N., Durkin, K., Artesi, M. et al., 2017). The people of these endemic regions could also have these genetic markers from which they could be getting infected. In Southern Japan, Ebola, and Dengue are also prevalent. These diseases may be associated with HTLV-1 and also support the possibility that genetics influences disease rates. It is known that parasites like *Strongyloids Stercoralis* can infect the host that may also have HTLV-1. So, it is reasonable to infer that other viruses or diseases can help increase the chances of developing ATL. Testing the samples of other diseases in these endemic regions can majorly impact the knowledge of HTLV-1 and ATL.

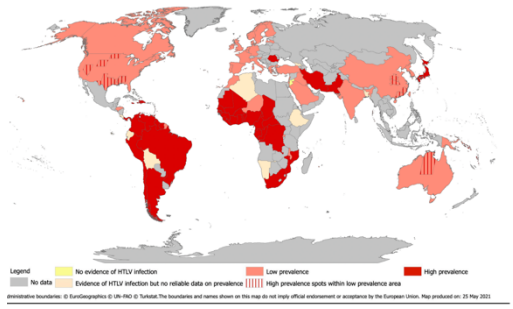


Figure 4: HTLV - 1 Distribution Map. Josh King-Robson,1 Timothy Hampton,2 Carolina Rosadas,3 Graham P Taylor, 3 Biba Stanton (2021)

In the central region of Africa alongside HTLV-1, some carriers have a parasite called *Strongyloids steracoralis*. This parasite would enter the body through soldiers' wounds during wartime. The life cycle of the parasite can be viewed in the figure below. This parasite can be found in areas where agricultural activities are performed. *Strongyloids steracoralis* can be transmitted through birds, reptiles, amphibians, primates, dogs, and cats.

When Co-infected with HTLV- 1 the growth rate of Adult T-Cell Leukemia is increased. This is due to the parasite's attack mechanism on the host's immune system which can increase the speed of leukemia. In the maps below, the prevalence of *strongyloids* and HTLV-1 is presented in the regions shown in Figures 4 and 5.

It can be seen in figures 4 and 5 Africa the regions of the high prevalence of *Strongyloids* and HTLV-1 are similar. It can be considered that in these regions HTLV-1 and *Strongyloids* can coexist in a person's body and cause ATL.

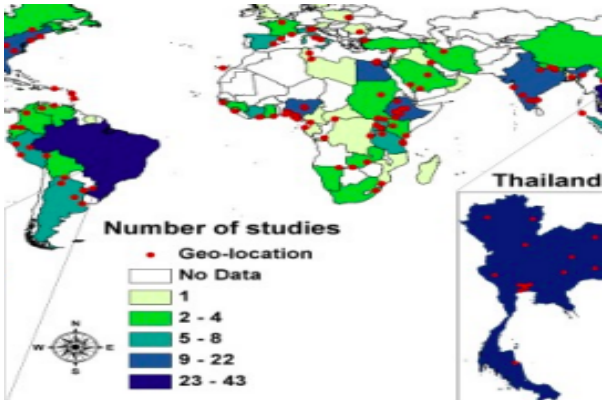


Figure 5: Strongyloids Stercoralis distribution map (Schär, et al., 2013).

3. Conclusion

HTLV-1 and ATL are not as widely recognized as many other viruses and diseases, and increasing awareness about them is crucial. Without adequate awareness, funding for research and the development of vaccines will be insufficient, hindering the ability to combat the spread of the virus. It is essential to understand the various factors contributing to ATL, such as *Strongyloids steracoralis*, genetics, and cancer markers. Genetics plays a significant role in determining the prevalence of ATL in certain regions, such as Southern Japan and Africa, and ancestry can be a predictor of one's likelihood of developing the disease.

Given the seriousness of this disease and its prevalence in endemic areas, it is critical to develop a vaccine for HTLV-1 and distribute it to affected zones. With more research, we can better understand the virus and develop effective prevention and treatment strategies. The fact that hundreds of thousands of people worldwide are affected by HTLV-1 and ATL cannot be ignored. Thus, we must prioritize finding ways to address the problem and work towards a solution.

Acknowledgments

I appreciate my parents for supporting me throughout my research. I also thank my research advisor Natalia for helping me throughout my research essay.

Vol. 2023 (1) 36 – 39
ISSN 2688-3651 [online]

References

Alvarez, C. et al. (2016). Family Aggregation of Human T-Lymphotropic Virus 1-Associated Diseases: A Systematic Review. Frontiers in Microbiology, 7, 1674.

Bangham, C. R., & Ratner, L. (2015). How does HTLV-1 cause adult T-cell leukemia/lymphoma (ATL)? Current Opinion in Virology, 14, 93–100. https://doi.org/10.1016/j.coviro.2015.09.004

Center for Disease Control and Prevention (CDC).

Jégado, B. et al. (2019). STLV-1 as a model for studying HTLV-1 infection. Retrovirology, 16, 41. https://doi.org/10.1186/s12977-019-0503-0

Nobre, A. et al. (2018). Low genetic diversity of the Human T-cell Lymphotropic Virus (HTLV-1) in an endemic area of the Brazilian Amazon basin. PLoS One, 13(3), e0194184. https://doi.org/10.1371/journal.pone.0194184

Rosewick, N. et al. (2017). Cis-perturbation of cancer drivers by the HTLV-1/BLV proviruses is an early determinant of leukemogenesis. Nature Communications, 8, 15264. https://doi.org/10.1038/ncomms15264

Schär, F., et al. (2013). Strongyloides stercoralis: Global Distribution and Risk Factors. PLoS Neglected Tropical Diseases, 7(7), e2288. https://doi.org/10.137

Shimoyama, M. (1991). Diagnostic criteria and classification of clinical subtypes of adult T-cell leukemia-lymphoma. A report from the Lymphoma Study Group (1984-87). British journal of haematology, 79(3), 428–437. https://doi.org/10.1111/j.1365-2141.1991.tb08051.x

Vahlne, A. (2009). A historical reflection on the discovery of human retroviruses. Retrovirology, 6, 40. https://doi.org/10.1186/1742-4690-6-40

World Health Organization (WHO).

Sato, Y., and Shiroma, Y. (1989). Concurrent infections with Strongyloides and T-cell leukemia virus and their possible effect on immune responses of host. Clinical Immunology and Immunopathology, 52(2). https://doi.org/10.1016/0090-1229(89)90173-6

Journal of Research High School

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

JRHS Outstanding Research Paper Award

An AI-Powered Assistive Device for the Visually Impaired

Katherine Hua[1] *

[1]Woodbridge High School, Irvine, CA, USA
*Corresponding Author: shivam.mohanty08@gmail.com

Advisor: Andrew Gibas, AndrewGibas@iusd.org

Received April 13, 2022; Revised June 8, 2023; Accepted, June 28, 2023

Abstract

Global studies report that 253 million people suffer from visual impairment. Most rely on traditional aids including Braille, white canes and guide dogs which lack versatility and adaptability. This study intended to utilize A.I. text, object recognition models and ultrasonic technology to create an effective vision aid. The prototype was constructed using a Raspberry Pi board with a Pi camera, ultrasonic sensor, earbuds, and other peripherals. Text and object recognition algorithms were implemented to convert printed text and real objects captured by the camera's live video feed to text output. Then, the text-to-speech code programmed in the device helped convert its text output to speech that could be heard through earbuds. Additionally, the ultrasonic sensors were programmed to determine the distance to the objects by measuring the time between emitting and receiving the reflected ultrasonic waves. During testing, the system recognized all sample words in 2.5 seconds on average and sample sentences in 6.4 seconds with 100% accuracy. Additionally, the device took 2.9 seconds on average to detect 80% of tested objects accurately and could detect large objects such as cars as far as 584cm away. The functional testing indicated that the prototype could inform users of recognized words and sentences, the types of detected objects, and their distance through audio. Thus, the results support the hypothesis that an AI-powered electronic device has the potential to provide reliable visual assistance to the visually impaired in their daily life.

Keywords: Visually impaired, Raspberry Pi, Vision aid, OCR, Text recognition, Object recognition, Ultrasonic sensor

1. Introduction

Vision plays a significant role in our daily lives. Our eyes allow us to explore and understand the physical world, learn new things, avoid danger and hazardous objects, and interact with other people. According to the Vision Loss Expert Group 2015 study, thirty-six million people suffer from blindness and an additional 217 million have medium to severe visual impairment (Bourne, et al., 2017). A report conducted by the National Center for Health Research found that one in five visually impaired people was unable to perform personal care activities while 46% had limited ability (National Center for Health Research, 2004). Their vision impairment causes them physical suffering, loss of productivity, lack of self-confidence, and an inferior quality of life. The economic burden of vision loss in the United States was estimated to be $134.2 billion in 2017, which included $98.7 billion in direct costs such as medical expenses for an eye examination, diagnosis, and corrections, and indirect costs such as informal care and social assistance programs (Rein et al., 2022). There is both a significant humanitarian demand and financial motive for developing a visual assistive aid using current-gen technology to grant much greater in-depth navigational and interpretive capability for the visually impaired. The question is, can it be achieved?

The visually impaired currently are provided with traditional aids such as Braille, white canes, and guide dogs as assistance in their daily life. However, these traditional mechanisms have major disadvantages and flaws. Since Braille is designed to function solely through touch and tactile recognition, it can take some time to develop the touch sensitivity, and practice using braille. This may become a challenge for seniors with slower cognitive and mobile abilities**.** While white canes help users locate obstacles to avoid, they cannot detect mobile objects or out-of-reach

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

hazards. Guide dogs are more versatile and responsive to situations, but they cost $32,000 for training and annual maintenance and can only work for 6-8 years (*The cost of a service dog*, 2022). Thus, a smart tool that has longevity and employs modern assistive technology to help the visually impaired is urgently needed.

Computer vision is one of the most widely used Artificial Intelligence applications and allows computers to gain a high-level understanding of contents from digital images and videos. The two core pre-existing technologies behind computer vision are deep learning and convolutional neural networks, known as CNN. CNNs are the core of computer vision as they are specifically modeled like the human brains to be able to recognize patterns and interpret visual data. Meanwhile, deep learning is a type of machine learning with algorithms structured in a hierarchy of increasing complexity and abstraction to develop high-level comprehension. It attempts to mimic the human mind's capabilities and thought processes through these trained algorithms (Voulodimos, et al., 2018).

Text and object recognitions are the two most prominent applications of AI in assistive aid for the visually challenged. While humans can quickly identify letters and words based on their ingrained knowledge and their trained visual recognition, machines do not have these natural abilities. Instead, computer vision allows machines to utilize image processing algorithms to understand the text within images through OCR, or Optical Character recognition. OCR technology trains and enables machines to be able to extract visual data from the pictures of printed text and convert the images of text into information that can be understood by a computer program. There are two steps involved in the OCR process. In the first stage, pre-trained models are used to detect text within images. Then, Deep Neural Network models process those images and implement text recognition through the OCR module (Onkar, et al., 2019).

Object recognition is a computer vision technique capable of detecting, locating, and identifying objects of a certain class (humans, automobiles, etc.) within an image or camera feed (Mwiti, 2019). It uses image classification to assign a class label to an image and classify it into a category. Meanwhile, it applies image localization to locate the object in the video feed and draw a bounded box around it (Brownlee, 2019). Many applications on cell phones, cars, computers, and cameras have used object recognition to achieve their desired functions such as video surveillance, facial detection, object tracking, pedestrian detection, etc. through identifying objects and their locations within video frames. The Single Shot Detector (SSD) is one of the best object recognition models and uses a convolutional neural network's pyramidal feature for the efficient detection of objects of various sizes. The tasks of object localization and classification are done in a single forward pass of the network. Additionally, the MultiBox technique is used for the bounding box regression function (Remanan, 2019).

During the past years, researchers have attempted to develop all kinds of assistive aids for the visually impaired by utilizing technologies such as A.I. object and text recognition algorithms, sonar technology, global positioning systems (GPS), etc. in their designs. The Smart Reader overcame the limitation of traditional Braille by integrating an AI text-reading system with a Raspberry Pi controller and reading out the text through a speaker (Ravil, et al., 2020). Another unique device based on Raspberry Pi utilized computer vision for identifying real-time objects through the image captured by the camera module and then provided audio narration about the detected objects through an audio receiver (Abedalrahim, et al., 2022). Researchers from the Center for Research and Advanced Studies in Mexico developed AI glasses that use artificial intelligence to recognize locations, read signs, and identify objects to help the blind navigate through their environment (Borghino, 2014). Yet another type of vision aid such as a smart walking stick was embedded with Raspberry Pi, an app, a GPS system, and an alarm component. The device can emit a sound alarm regarding the obstacles detected by the sensor, use the GPS to track the position of the visually impaired and provide help via an app in an emergency (Sahoo, et al., 2019). Another aid, a walking stick used a microcontroller that emitted and received ultrasonic pulses to capture and process environmental data. By integrating the SOS navigation system into the smart stick, the device could assist the blind by providing warnings about the approaching obstacles in the path and helping them avoid emergencies (Mohapatra, et al., 2018). These revolutionary designs in past research demonstrated the capability and potential of using advanced technology to help the visually impaired regain independence in their daily life.

The purpose of this study was to build a smart visual device that could effectively help the visually impaired read printed text, identify objects in the surroundings and keep them informed of approaching obstacles through an alarm. The prototype was built by using a Raspberry Pi microprocessor and incorporating OCR text recognition, object

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

recognition, and ultrasonic technology. By utilizing both artificial intelligence algorithms and ultrasonic sensors for this project, this single pocket-sized device maximized its efficiency by providing the visually impaired three types of critical information: printed text in the environment, the type of objects detected, and the distance to objects via sound and vibration. My hypothesis is that this electronic device has the potential to help the visually impaired deal with the challenges they face each day.

2. Materials and Methods

The Raspberry Pi 4B was chosen to be the microcontroller of the device and was connected to the following peripherals: Raspberry Pi Camera V2, HC-SR04 ultrasonic distance sensor, mini vibration motor disc, and earbuds (Figure 1). The OS used was the Raspbian Buster OS which was flashed onto a 16GB micro-SD card using the Raspberry Pi Imager. After the device was powered on and booted up, the initial setup and configuration were completed to connect the Pi to Wi-Fi, enable the camera, and install any software updates. To run the object recognition algorithm within a Python script, the OpenCV library, TensorFlow Lite dependencies, and Google's sample quantized SSD Lite Mobile Net V2 object detection model (trained on the COCO dataset) were installed within a Python virtual environment. The base Python script meant to run the object recognition model over the camera's live stream was downloaded into the virtual environment by cloning a public GitHub repository (Bradway, 2022). Within the same virtual environment, the Tesseract OCR library, and packages such as Pytesseract, OpenCV-python, and NumPy were also installed (*Build live text recognition with the Raspberry Pi (OCR),* 2022).

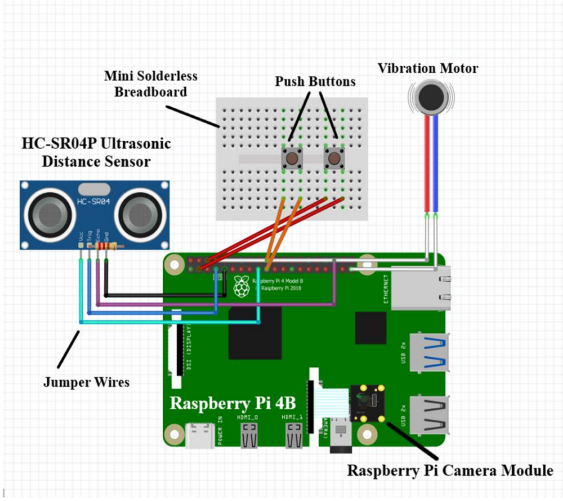


Figure 1. Image of prototype design: Pi camera, ultrasonic sensor, vibration motor, and wires were connected to the Raspberry Pi board's ports.

There were three separate Python scripts for this device. One script controlled the ultrasonic sensor which was used to calculate the distances between the device and surrounding objects. The transducer on the sensor emitted high-frequency sound waves at the speed of sound and then received them reflected from objects. The time between the emission and reception of the sound waves was multiplied by the speed of sound and then divided by two to calculate the distance between the sensors and the objects (Burnett, 2018). Depending on the magnitude of the calculated distance, the script made the mini-vibration motor discs emit vibrations at different intensities as a proximity warning. The intensity of the vibration was inversely related to the distance between the sensor and the object as the intensity increased whenever an object or obstacle was closer to the sensor.

Another Python script within the cloned GitHub repository took individual frames from the camera's live video view and processed them using TensorFlow's object detection model to recognize objects within the image. The script was modified to calculate the objects' center coordinates, which were then used to determine the object's relative position within the image frame: left, right, or forward. The code was also modified to include a text-to-speech synthesizer function which used an installed TTS engine to vocalize outputs about the recognized object type through audio output that can be heard from earbuds.

The third Python script ran text recognition using the Tesseract OCR library and Pytesseract. There were four steps in the text recognition process. In the first step, the paper with printed text messages was placed in front of the Pi camera so that the camera could capture the images of the text. The second stage of pre-processing checked whether the captured images got skewed towards the left or right, and then used the appropriate skew correction for the images. During the third step, the OCR algorithm code processed the images containing characters and identified the words.

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

Lastly, the Text-to-Speech (TTS) code translated the text output into corresponding speech output through the earbuds attached to the Raspberry Pi board.

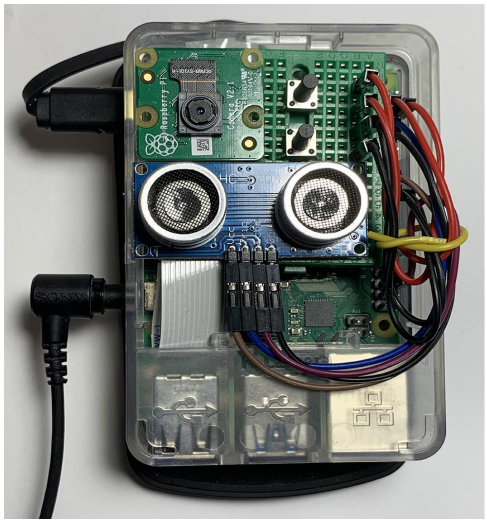


Figure 2. Image of actual prototype design: Pi camera, ultrasonic sensor, and wires were connected to the Raspberry Pi 4B board.

3. Results

The prototype design consisted of Raspberry Pi 4B board, Raspberry Pi camera, ultrasonic sensor, vibration motor, earbuds, and battery as shown in Figure 2. The Pi 4B microprocessor took the images of printed text or objects caught by the Pi camera as inputs and then produced the speech feedback of the texts or objects through earbuds as output. Additionally, the ultrasonic sensor implemented the object detection function and then triggered the motor to vibrate when the object was approaching within a certain distance.

When the Raspberry Pi was turned on, the booting screens appeared in the terminal. There are two buttons on the device. When button one was pressed, the device performed the text recognition function. While the Pi camera caught the text images and bound the images in the box as shown in Figure 3, the system programmed with the OCR text recognition algorithm processed the images, recognized the texts, and then generated speech output of the text messages that can be heard through the earbuds.

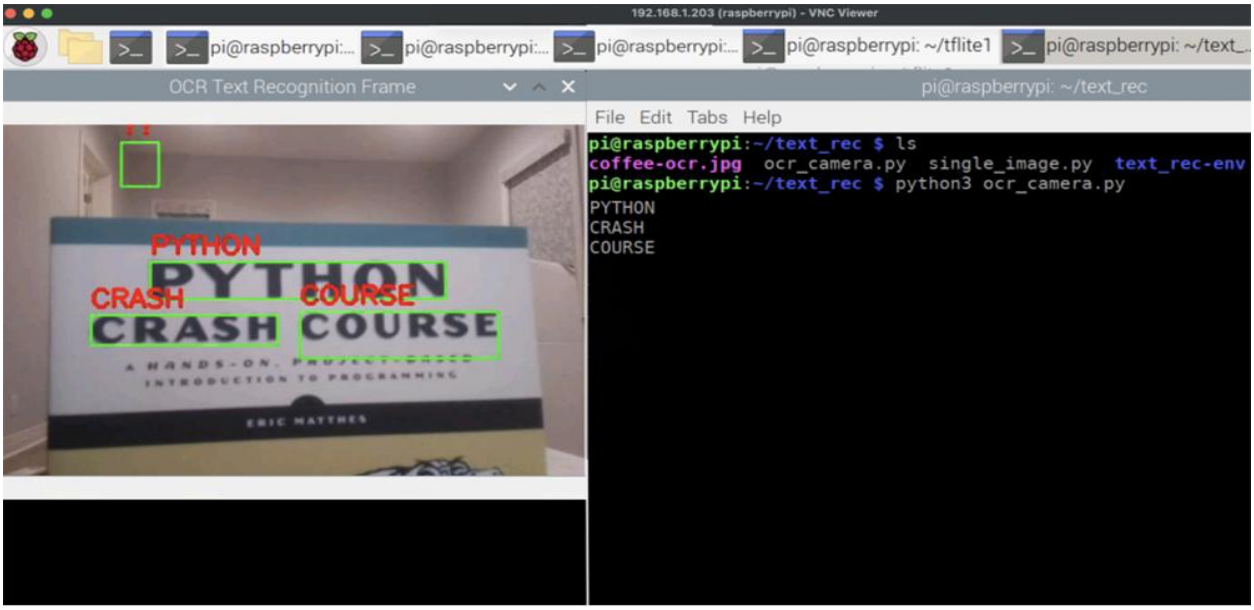


Figure 3: Image of text recognition testing: OCR algorithms recognized the words in the camera frame and indicated the text output in the terminal.

Testing was conducted by placing the paper with printed words and sentences in front of the Pi camera. The data collected from testing OCR text recognition is shown in Table 1. The collected data indicated that the device recognized all single words ranging from five-letter words to thirteen-letter words and all short sentences with 100% accuracy (Table 1). The average time taken to recognize single words was 2.5 seconds, while the average time taken to recognize short sentences was 6.4 seconds.

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

Table 1: Subset of Testing Data for Texas Recognition

Text (specific words)	Audio Output	Trial 1 time (s) to successfully recognize text	Trial 2 time (s) to successfully recognize text	Trial 3 time (s) to successfully recognize text	Average Time (s)
egg	egg	2.9	3.2	3.2	3.1
cash	cash	2.8	3.2	3.1	3.03
orange	orange	2.7	2.6	3	2.77
problem	problem	2.5	2.4	2.7	2.53
assistive	assistive	2.1	2.3	2.5	2.3
technology	technology	2	1.9	2.3	2.07
competitor	competitor	1.8	2.2	2.1	2.03
environment	environment	2.4	2.3	1.9	2.2
functionality	functionality	3.1	2.4	2.4	2.63
I ate an apple.	I ate an apple.	5.1	5.3	5.6	5.33
I love to watch movies.	I love to watch movies.	6.5	6.7	7.9	7.03
We went to the beach.	We went to the beach.	6.7	6.3	7.2	6.73

When button two was pressed, the device performed the object recognition function. The window was the live feed from the Pi Camera with the object detection API. It detected the book and orange in the camera's live feed by bounding the object within a box that contained the object label and generated the text output "book" and "orange" in the window as shown in Figure 4. Meanwhile, the object location identification code produced the output of the object's relative location such as forward, left side, and right side (Figure 5). Then, the device used Text-to-Speech synthesis code to generate speech output about the object's class label and location relative to the users through the earbuds' audio output. Additionally, the ultrasonic sensors on the device calculated the distance between the objects and the device (Figure 5) and then generated vibration from motors as an alarm when the object was closer to the user within a certain distance.

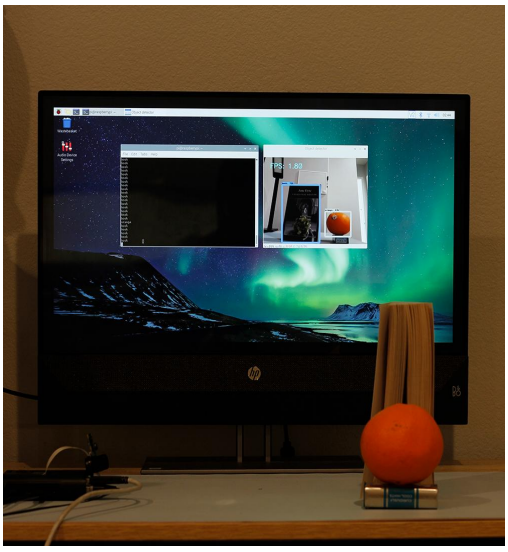


Figure 4. Image of object recognition testing: object detection model recognized the book and orange in the camera frame and printed the class name/label in the terminal.

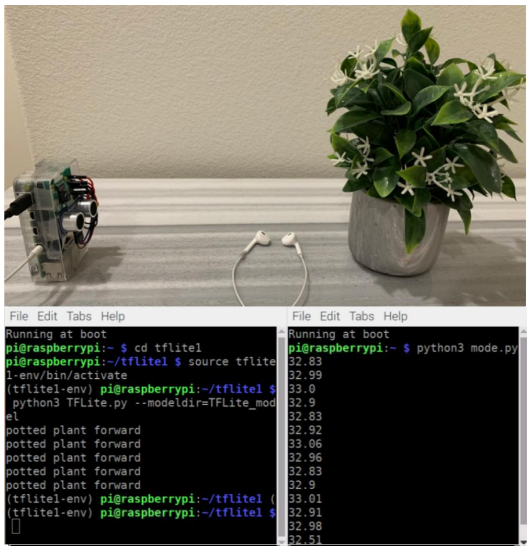


Figure 5. Images of object recognition testing: object detection algorithms identified the relative position of the objects (left, right, or forward) while ultrasonic sensors calculated the distance between the object and the sensor.

The testing data from experiments on object recognition and ultrasonic distance measurement is shown in Tables 2 & 3. The device could recognize 80% of real objects in testing correctly. It took 2.86 seconds on average to recognize

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

the objects and produce speech output through earbuds. Additionally, the average maximum distance the ultrasonic sensor was able to detect the objects was 165 cm. The device could detect large objects such as cars as far as 548cm away.

Table 2: Subset of Testing Data for Object Recognition

Tested Objects	Audio Output	Actual location	Detected location	Processing time	Max. Distance Detected
Water Bottle	Bottle	left	left	4.3 s	110 cm
Laptop	Laptop	forward	forward	2.0 s	134 cm
Backpack	Backpack	forward	forward	2.4 s	123 cm
TV	TV	forward	forward	1.4 s	147 cm
Cellphone	Cellphone	left	left	1.8 s	82cm
Person	Person	left	left	1.3 s	177 cm
Fire Hydrant	Fire Hydrant	left	left	5.3 s	230 cm
Trash Bin	Trash Bin	forward	forward	3.8 s	503 cm
TV	TV	right	right	1.4 s	147cm
Person	Person	left	left	1.3 s	177cm
Streetlamp	Traffic Light	right	right	9.0 s	191 cm
Chair	Chair	right	right	2.2 s	247 cm
Plant	Potted plant	right	right	2.8 s	87cm
Car	Car	forward	forward	2.5 s	548 cm
Tree	Not Recognized	N/A	N/A	N/A	139cm

4. Discussion

The results from my experiments show that the device can quickly provide feedback to visually impaired users without sacrificing accuracy. It can read out all words in the sentences with 100% accuracy. While nearly all the objects were detected by the ultrasonic sensor, only 80% of objects were correctly identified by the device. The object recognition model had its limitations when recognizing real-life objects: it either misclassified the objects or could not recognize the objects at all. Future improvements can be made through further training of the object recognition model with more images of real-world objects. Additionally, the device could not identify the traffic light because it was not able to recognize the color patterns. The model's recognition of color and symbols such as? #, and numbers will need further research and training in order to perform these functions reliably enough for incorporation into the device.

Potential future experiments could include the redesign of the device with advanced hardware which could improve the processing speed, accuracy of identification results, and addition of more functions. The current design was based on the Raspberry Pi 4B microcontroller. Replacing the Raspberry Pi board with Navida's Jetson Nano will speed up image processing and allow several computer vision packages to run simultaneously. While the ultrasonic sensor is a cost-effective solution for detecting the distance between obstacles and the user, it cannot detect the objects which are beside or beneath the sensor. On the contrary, the Lidar sensor or 3D depth sensor is compatible with the Jetson Nano microcontroller board and allows devices to provide a high-resolution three-dimensional view of their surroundings. The second version of the visual device which consists of Jetson Nano, a USB camera, and a Lidar or 3D depth sensor will enable the device to run object, text, and color recognition simultaneously without sacrificing its processing speed. To address the challenge that earbuds pose by potentially preventing the user from hearing environmental noises and sounds, a new design could incorporate both earbuds and speakers. The users can switch between earbuds and speakers according to real-life situations and the

Table 3: Statistics of Average Processing Time and Maximum Distance Detected by Ultrasonic Senso

Descriptive Statistics			
Average Processing Time (s)		Average Max. Distance Detected(cm)	
Mean	2.86	Mean	164.75
Standard Error	0.262818729	Standard Error	26.44712047
Median	2.6	Median	128.5
Mode	2.7	Mode	98
Standard Deviation	1.175361087	Standard Deviation	129.5639006
Range	4.2	Range	512
Minimum	1.3	Minimum	36
Maximum	5.5	Maximum	548

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

surrounding environment. Additionally, the user can choose to wear only one earbud, allowing their free ear to pick up the background and surrounding noise easier.

Currently, there are several vision aid devices on the market to help the visually impaired. The research on this existing alternative indicates that the AI-powered device offers an innovative solution for vision impairment. Its concept and technology-oriented design make it more effective, convenient, and accessible to visually impaired individuals than any other existing devices.

Sunu Band is a sonar wristband used in combination with a white cane to detect objects no more than six feet away. By using ultrasonic sensors to calculate the distance to objects, Sunu Band can inform users about approaching obstacles and provide prompt warnings. According to their website, the Sunu Band also pairs with an app to provide GPS assistance as real-time navigation support. However, Sunu Band cannot identify the type of obstacles in the visually impaired user's path. By utilizing TensorFlow's object recognition model that processes live video feed from the camera, the developed device not only alerts users of oncoming objects through vibration but also identifies and declares their object class via audio feedback.

Another product in the market is IrisVision's headset which utilizes mobile virtual reality by pairing Samsung Galaxy's smartphone with the Gear VR headset to help low-vision people observe objects, read and shop. The headset uses a magnification bubble to achieve magnification power so that people with low vision can see the enlarged images through goggles. However, this product is designed to help low-vision people only and is not recommended for outdoor walking because of its limited peripheral vision and depth perception. The developed device uses A.I. text and object recognition algorithms to process real-time video images from the camera and then immediately generates audio output by using a built-in text-to-speech function. It was specifically designed to target the blind and severely visually impaired, but it can also help low-vision individuals.

The Brainport Vision Pro is another alternative in the assistive aid market. It is a unique intraoral headset for the visually impaired and consists of a video camera and tongue array working with user control. It transforms the images caught on wearable cameras into electrical stimulation patterns which can be felt by the tongue. The users need to be trained to learn how to interpret the moving bubble-like pattern on the tongue to 'see' the shape and size of the objects. It is neither convenient nor easy to work with since its users must undergo complicated training programs. On the contrary, the developed device utilizes the most advanced A.I. technology thus far and is much more intuitive to operate without prior training.

Through improvements and further development, this vision aid can become a major player in the assistive aid industry and potentially assist a large number of visually impaired individuals in managing their daily challenges and regaining their independence. This study has important implications for the design of assistive aids for the visually impaired by demonstrating possible functions and indicating areas of future improvement for scientists and engineers to develop a truly reliable aid in the near future.

5. Conclusion

The development of AI text and object recognition models have enabled the construction of electronic vision aids for the visually impaired. The test results supported my hypothesis that this AI-powered device can provide a viable solution to the problem I intended to solve for this study. The device took 2.5 seconds on average to recognize single words and 6.4 seconds to recognize short sentences with 100% accuracy. This proved its potential as a text reader for the visually impaired. Meanwhile, the prototype provided quick and accurate real-time aid by verbalizing objects' names and their distance to the device with an 80% accuracy rate within 2.86 seconds on average. This demonstrated its advanced functionality and projected performance in identifying real-life objects in the surroundings. Since the device was programmed with a text-to-speech function, the text outputs from text and object recognition were automatically transformed into audio output for users to hear. The device's accuracy, effectiveness as a vision aid, and real-time speech feedback show that the prototype has fulfilled the requirement of its design criteria. Through improvements and further research, this vision aid can become a viable solution that is more effective than traditional aids such as Braille, walking canes, and guide dogs, and more affordable than existing assistive aids in the market.

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

The device's innovative integration of computer vision algorithms into one pocket-sized microcontroller will enable it to address a long-overlooked problem.

References

Abedalrahim J., et al. (2022). Intelligence interfaces for assisting blind people using object recognition methods. *International Journal of Computer Science and Applications*. Vol 13 (5). https://doi.org/10.14569/IJACSA.2022.0130584

Borghino, D. (2014). *Hi-tech glasses aim to assist the blind with directions and obstacle detection.* Newatlas. https:// newatlas.com/stereoscopic-ultrasound-gps-ai-glasses-blind-assistance/32166.

Bourne, R., et al. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *Lancet.* Vol. 5 (9). https://doi.org/ 10.1016/S2214-109X (17)30293-0

Bradway, D*. (2022) Tutorial to set up TensorFlow Object Detection API on the Raspberry Pi.* Github. https:// github.com/EdjeElectronics/TensorFlow-Object-Detection-on-the-Raspberry-Pi

Brownlee, J. (2019). *A gentle introduction to object recognition with deep learning*. Machine Learning Mastery. https:// machinelearningmastery.com/object-recognition-with-deep-learning.

Build live text recognition with the Raspberry Pi (OCR) (2022). Tutorials for Raspberry Pi. https://tutorials-raspberrypi.com/raspberry-pi-text-recognition-ocr.

Burnett, R. (2018). *Understanding how ultrasonic sensors work.* MaxBotix. https: //maxbotix.com/articles/how-ultrasonic-sensors-work.htm

Mohapatra S., et al. (2018). *Proceedings of 2018 2nd international conference on trends in electronics and informatics*. IEEE Xplore. https:// ieeexplore.ieee.org/document/8553935.

Mwiti, D. (2019). *A 2019 guide to object detection.* Heartbeat. https:// heartbeat.fritz.ai/a-2019-guide-to-object-detection-9509987954c3

National Center for Health Research (2004). *Blind adults in America: their lives and challenges.* https:// center4research.org/blind-adults-america-lives-challenges.

Onkar, O.S., et al. (2019). Optical character recognition for the blind using Raspberry Pi. *International Research Journal of Engineering and Technology*. Vol. 6(3). https://www.irjet.net/archives/V6/i3/IRJET-V6I3520.pdf

Ravil, A., et al. (2020). Raspberry pi based smart reader for blind people. *Journal of Analysis and Computation (JAC)*. Vol. XIII (III). https://ijaconline.com/wp-content/uploads/2020/06/1012-Paper-Dr.-A.-Ravi.pdf.

Rein, D.B., et al. (2022). The economic burden of vision loss and blindness in the United States. *Ophthalmology.* Vol. 149 (4). 369-378. https://doi.org//10.1016/j.ophtha.2021.09.010

Remanan, S. (2019). *Beginner's guide to object detection algorithm.* Medium. https://towardsdatascience.com/beginners-guide-to-object-detection-algorithms-6620fb31c375

Sahoo N., et al. (2019) Design and implementation of a walking stick aid for visually challenged people. *Sensors*. Vol. 19 (1). https:// doi.org/10.3390/s19010130 mdpi.com/1424-8220/19/1/130.

The cost of a service dog. (2022). Guide Dogs of America. https://www.guidedogsofamerica.org/cost-of-a-service-dog/

Voulodimos, A., et al. (2018). Deep learning for computer vision: a brief review. *Computational Intelligence and Neuroscience*. Vol 2018. https://doi.org//10.1155/2018/7068349.

Journal of Research High School

Vol. 2023 (2) 48 – 56
ISSN 2688-3651 [online]

Effect of Coating Seeds with Micronutrients and Bacterial Consortia on Stomatal Conductance and Yield of Cluster Bean

Sachin Vashisht[1*], Dhruva Shankara[1], Punya Shree[1], Suvratha Herur[1], Shreeadithya Kashyap[1], Venkata Krishna Bayineni[2]

[1]Samvida school, Bengaluru, Karnataka, India
[2]Prayoga Institute of Education Research, Bengaluru, Karnataka, India
*Corresponding Author: sachinjv28@gmail.com

Advisor: Dr. Venkata Krishna Bayineni, krishna.bayineni@prayoga.org.in

Received October 22, 2022; Revised June 20, 2023; Accepted, July 11, 2023

Abstract

The study aims to evaluate the effect of coating seeds with micronutrients in combination with bacterial consortia on stomatal conductance and yields of cluster bean. Greenhouse experiments were carried out during August – December 2021 at Prayoga Institute of Education Research, Bangalore, India. The experiment consisted of 8 different treatments laid out in a randomized block design with five replications. Plant growth-promoting rhizobacteria (PGPR) are known to influence plant growth by various direct or indirect mechanisms. In search of efficient PGPR strains with multiple activities, a total of 10 bacterial strains were isolated from different rhizosphere soil and plant root nodules in the vicinity of Prayoga campus. These test isolates were screened in vitro for their plant growth-promoting traits like the production of *Indoleacetic acid,* production of *Ammonia* and *Phosphate solubilization* and three isolates were selected to develop consortia in different combinations. The micronutrients and bacterial consortia were applied to the seeds in combination as per the studied treatments and drought stress was created during growth stage. The study result revealed improved yields with a decrease in stomatal conductance due to plant adaptation to the stress induced by drought. There was a significant difference in the seed yield of treated treatments with that of the untreated control. Stomatal conductance (mmol m−2 s−1) of five randomly tagged plants was recorded by using a leaf porometer after 40 days and seed yield was recorded. Hence, coating seeds with bacterial consortia and micronutrients had a significant influence on the stomatal conductance and seed yield of cluster bean and this approach can serve as an effective strategy to enhance cluster bean crop performance.

Keywords: Cluster bean, Micronutrients, Rhizobacteria, Seed coating, Stomatal conductance, Yield

1. Introduction

Agriculture provides the principal means of livelihood for the majority of the Indian population. Modern agriculture is not sustainable in long run; Modern agriculture, characterized by intensive practices and heavy reliance on synthetic inputs, has been instrumental in meeting the global demand for food. However, mounting evidence suggests that this system is not sustainable in the long run. Hence the concept of sustainable agriculture has emerged in recent years with an emphasis more on the conservation of natural resources and the environment. Cluster bean (*Cyamopsis tetragonoloba L. Taub*.) or Guar is a major vegetable and industrial crop grown for its tender pods and endospermic gum (30-35 %). Tender pods are rich in nutrients: they contain 16 Kcal of energy, 3.2 grams of protein, 1.4 grams of fat, 10.8 grams of carbohydrate, 65.3 IU of Vitamin A, 49 mg of Vitamin C, 57 mg of calcium and 4.5 mg of iron for every100g of an edible portion (Ashraf & Iram, 2005). Due to their hardiness and tolerance for poor soil and moisture stress conditions, cluster bean is mostly grown in rainfed conditions in arid and semi-arid parts of tropical India. India produces over 80% of the world's cluster bean production (Punia, *et al*., 2009).

Vol. 2023 (2) 48 – 56
ISSN 2688-3651 [online]

Cluster bean seeds are primarily used for extracting endospermic gum, which has excellent binding characteristics and is in high demand in the culinary industry as an ingredient in sauces and ice creams, among other things. Guar gum is also utilized as a sizing agent, thickener, stabilizer, protective colloid, absorbent, flocculating agent in various industries (Sabahelkheir, *et al*., 2012; Srivastava, *et al*., 2011). Cluster bean pods are utilized as a treatment for diabetes patients, in addition to having a high caloric and nutritional content. For monogastric animals, cluster bean meal (high protein content) produced from the seed coat and germ cell is a good feed. For export, cultivars with a high gum content (>32 %) and viscosity (4000-5000 cps) are desirable (Kumar & Rodge, 2012; Meena & Asrey, 2018). In addition, cluster bean is grown as a green manuring crop in different parts of the world. The husk of cluster bean is used for cattle feed because it contains high protein content (Goudar, *et al*., 2016).

Despite their nutritional importance, cluster bean production has remained low even during the green revolution era. One of the many factors contributing to the cluster bean's low production is a lack of micronutrients in addition to macronutrients. Exogenous sources of these micronutrients are badly needed in different cropping systems due to excessive removal by high-yielding varieties. Micronutrients are extremely effective in little doses to generate optimal results. Micronutrients are important for increasing agricultural productivity. Zn insufficiency is a common micronutrient deficiency in Indian soils. Molybdenum deficiency, by and large, is associated with acid soils. Micronutrient insufficiency was caused by intensification of agriculture with high-yielding crop types, continual application of high-analysis chemical fertilisers, limited supply of organic manures, and little crop residue return to the soil. Micronutrients are required in lesser amounts than macronutrients, yet they are just as vital. Growth suppression or even complete inhibition can occur if any of these components are deficient in the soil or are not properly balanced with other nutrients. In addition to serving as cofactors in enzyme systems and participating in redox reactions, micronutrients play a number of other important roles in plants. Most notably, micronutrients are involved in the fundamental physiological processes of photosynthesis and respiration, and their lack can obstruct these activities thus limiting yield gain in many crops (Mengel, *et al*., 2001). In most crops, seed treatment with micronutrients improves stand establishment, accelerates phenological events, boosts yield, and increases micronutrient content in grain. In some cases, seed treatment with micronutrients outperformed comparable to soil application and foliar spray techniques (Singh and Gandhi, 2015; Quddus, *et al*., 2020). Seed treatment with polymer coating is a simple and cost-effective strategy for resource-poor farmers because of its pronounced influence during the early stages of seedling establishment (Johnson, *et al*., 2005). These treatments can also alter the stomatal response which can play a key role in plant adaptation.

Stomata consist of pores scattered over the relatively waterproof and CO2-tight cuticle covering the leaf surface. Stomata play a key role in plant adaptation to changing environmental conditions as they control both water losses and CO_2 uptake. Stomatal conductance (gl) is a measure of the degree of stomatal opening and can be used as an indicator of plant water status. Stomatal conductance (mmol m−2 s−1) measured by a porometer is the rate of CO2 entering, or water vapour exiting through stomata (Damour, *et al*., 2010), which is dictated by the degree of stomatal opening. In irrigated trials, the handheld porometer allows for quick measurement of leaf stomatal conductance and is suggested for use under water stress (unless extremely slight) because the stomata are normally closed (Shinde, *et al*., 2016).The rhizosphere is a complex ecology in which plant impacts on soil microbes and microorganism effects on plants interact and are interdependent (Mukerji, *et al*., 2006).

Plant-growth promoting rhizobacteria (PGPR) are a type of bacteria that colonizes the roots of plants and promotes their growth (Ghadamgahi, *et al.,* 2022). Sitepu (2008) proved that a variety of bacteria from the genera *Azospirillum, Alcaligenes, Acinetobacter, Bacillus, Burkholderia, Enterobacter, Erwinia, Flavobacterium, Pseudomonas, Rhizobium,* and *Serratia* are associated with the plant rhizosphere and can benefit plant growth. The key processes through which PGPR contributes directly to plant phytohormone production, such as auxins, cytokinin's, and gibberellins are through mineral solubilization, siderophores, and enzyme production (Bhattacharyya, *et al*., 2012).

Undoubtedly, one of the most significant environmental factors affecting crop plant productivity worldwide is drought. And this was reported to greatly affect the yield of cluster bean cultivation. Drought stress decreases the rate of photosynthesis (Kawamitsu, *et al*., 2000). To conserve water, plants cultivated in drought conditions have a decreased stomatal conductance. Keeping in mind the above-mentioned beneficial properties and importance of PGPR

Vol. 2023 (2) 48 – 56
ISSN 2688-3651 [online]

and micronutrients in promoting plant growth and their adaptation, we hypothesized that coating the seeds with them would increase the PGPR in the rhizosphere and hence regulate the stomatal conductance of cluster bean. Also, we hypothesized that these results would contribute to a better understanding of the responses of cluster bean plants to drought stress.

2. Materials and Methods

2.1 Study Area and Sample Collection

The rhizosphere soils of plants were collected from Prayoga Campus (Ravugodlu, Bengaluru 560082, Karnataka, India) from Prayoga campus (Fig. 1 and 2). The collected soil samples were transported to the laboratory and refrigerated (4 ºC) for further processing. Bacteria were isolated using the serial dilution technique and spread plate technique on nutrient agar (NA). 1gm of rhizosphere soil sample was suspended in 100ml autoclaved 0.85% saline. After sedimentation of solid particles, dilution was made upto10^6. 0.1ml of each dilution was spread by L- shaped rod on NA. The selected isolates were tested for their plant growth-promoting properties, such as *IAA* production, *P* solubilization, and *ammonia* production, were evaluated for some isolates.

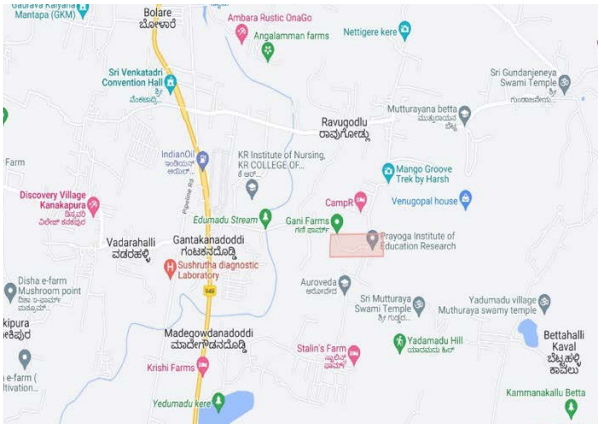


Figure 1. Location of Sampling Site in the Main Campus of Prayoga, Ravugodlu where the Rhizosphere Soil Samples were Collected.

Figure 2. Collection of Rhizosphere Samples from Prayoga Campus, Ravugodlu.

2.2 Test for Plant Growth Promoting Properties

Indole Acetic Acid (IAA) Production

Indole acetic acid production was detected as described by Bric, *et al.,* (1991). Bacterial isolates were inoculated in NA amended with L-Tryptophan and incubated at 37 ºC for 48 h. Fully grown cultures were centrifuged at 3000 rpm for 30 m, the supernatant (2ml) was mixed with two drops of orthophosphoric acid and 4ml of the Salkowaski reagent (50 ml 35% of perchloric acid, 1ml of 0.5mFeCl3 solution). The development of the pink color was indicative of IAA production.

Phosphates Solubilization

Phosphate solubilizing ability of the isolate was checked on Pikovskaya (PVK) medium, incorporated with tricalcium phosphate ($Ca_3 (PO_4)_2$). The isolates were spot inoculated on PVK medium. The formation of a transparent halo zone around the developing colonies indicated *phosphate* solubilizing ability.

Materials required to make Pikovskaya (PVK) medium:

PVK medium contained per litre: glucose,10 g; $Ca_3(PO_4)_2$, 5 g; $(NH_4)_2SO_4$, 0.5 g; NaCl, 0.2g; $MgSO_4W7H_2O$, 0.1 g; KCl, 0.2 g; Yeast extract, 0.5g; $MnSO_4WH_2O$, 0.002 g; and $FeSO_4W7H_2O$, 0.002 g (Surange, *et al.,* 1997).

Vol. 2023 (2) 48 – 56
ISSN 2688-3651 [online]

Assay for Ammonia Production

The rhizobacterial isolates were tested for the production of *ammonia* in peptone water after incubating at 30 ºC For 48 h. Nessler's reagent (0.5 ml) was added to each tube and observed for the development of a brown to yellow color (Cappucino, *et al*., 1992).

2.3 Gram Staining

The Gram stain technique was carried out using a Gram stain kit. The kit includes 4 bottles of 1X solutions of safranin, acetone-alcohol, Gram's iodine, and crystal violet. The rhizosphere bacterial isolates were maintained and cultivated for 24 hours at 37°C on NA to prepare the microscope slides for use in this study. Each bacterium was smeared separately onto a slide and heat-fixed just before staining. The Gram stain method was carried out in accordance with the instructions provided in a typical microbiology laboratory manual (Michael & Burton, 2002).

2.4 Bacterial Consortia

The isolated bacterial strains were grown on NA for routine use and maintained in nutrient broth with 20% glycerol at - 80 °C for long-term storage. A single colony of bacteria was transferred to 500 ml flasks containing 250 ml of NB and were grown on a rotary shaker (150 rpm) for 48 h at 30 °C. An equal volume of the bacterial suspension was used to make the consortia.

2.5 Coating the Seeds

The micronutrients (like potassium molybdate, $ZnSO_4$, boron etc.,) are applied to the seeds of cluster bean either individually or in combination and grown in a greenhouse facility (Fig. 3 and 4). Micronutrients (like potassium molybdate, $ZnSO_4$, boron etc.,) and/or bacterial consortia are applied to the seeds of cluster bean either individually or in combination (Table 1). The biometric observations on stomatal conductance and growth and yields are recorded. For recording such observations, five plants at random from net plot area were selected from each plot.

Drought stress was created during the vegetative stage by withholding re-watering at and after flowering. Then, around once every week, plants were irrigated with tap water. Hand weeding was used to keep the plots free of weeds. The effects of the drought treatment and covering the seeds with bacterial consortia and micronutrients on seed yield were evaluated at the conclusion of the crop cycle.

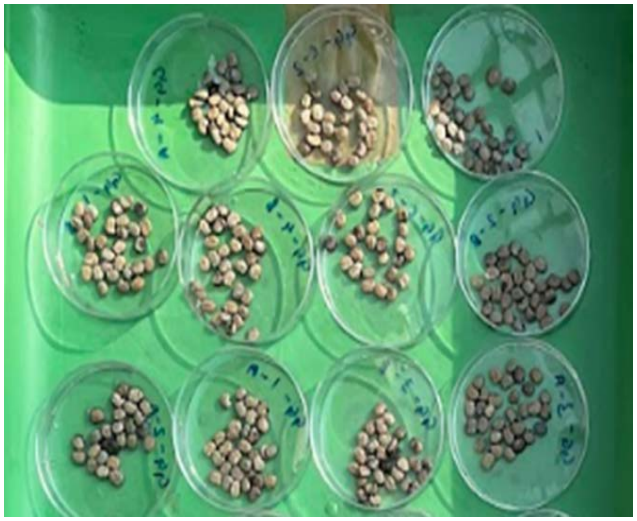


Figure 3. Cluster Bean Seeds Coated with Bacterial Consortia and Micronutrients

Table 1. List of treatments showing respective experimental design

Set Code	Treatment
1a	1+2+MN
1b	1+2
2a	1+2+3+MN
2b	1+2+3
3a	1+3+MN
3b	1+3
4a	2+3+MN
4b	2+3
Control 1	No coating
Control 2	Only starch
Control 3	Only micronutrients

MN - Micronutrients
1, 2, 3 – Isolates

Vol. 2023 (2) 48 – 56
ISSN 2688-3651 [online]

3. Results and Discussion

Plant growth-promoting rhizobacteria are usually applied to a wide range of agricultural crops for the purpose of growth enhancement, including increased seed germination, plant weight, and harvest yields. PGPR colonization triggers plant growth by bacterial synthesis of plant hormones including *indole-3-acetic acid*, *cytokinin*, and *gibberellins* as well as by increased mineral and nitrogen availability in the soil. Some of them were also known to protect their host plant from pathogenic microorganisms (Handiganoor, *et al*., 2018). In our investigation, four rhizobacteria were isolated from four healthy selected plants from Prayoga.

Figure 4. Planting Seeds in a Small Green House

3.1 Isolation of PGPR

The morphological characteristics of isolates were examined to evaluate colony diversity. It was thought that the isolates' physiological traits varied considerably between the various colonies (Fig. 5). Therefore, 10 bacteria were chosen and used in the following experiments based on the morphological characteristics of isolates, such as the form (circular, filamentous, and irregular), color (white, whitish, yellow, yellowish, creamy, and transplant), elevation (convex, flat, raised, crateriform, and umbonate), and margins (entire, filiform, and undulate) of colonies (Fig. 5 and 6). The isolates were sub-cultured and preserved on NA slants and used for further experiments.

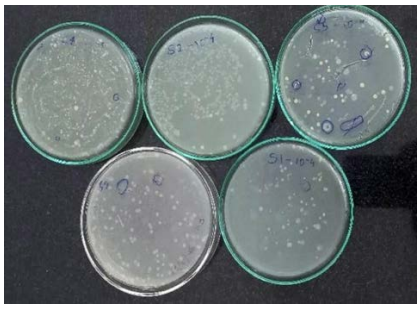


Figure 5. Isolation of Plant Growth Promoting Rhizobacteria by Spread Plate Technique

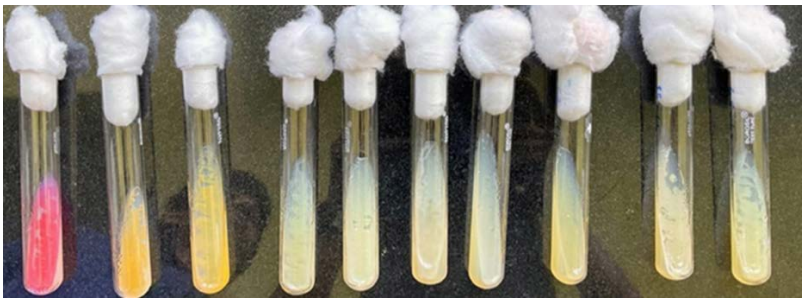


Figure 6. Selected Isolates Maintained on NA Slants

3.2 Physiological properties of isolates

Indole Acetic Acid (IAA) production

A total of 10 selected isolates of PGPB were tested for the production of IAA in the presence of tryptophan. With no addition of tryptophan, production of IAA was not observed. Out of 10 isolates tested, 9 showed positive. The production of IAA was highest in isolates of S4-C3 and S3-C and no IAA was observed in S5-C2 (Fig. 7).

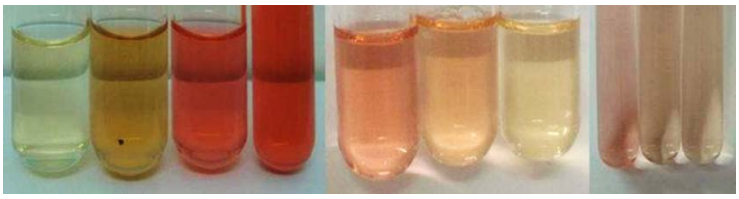


Figure 7. Result of *Indole Acetic Acid* (IAA) production test

Vol. 2023 (2) 48 – 56
ISSN 2688-3651 [online]

Figure 8. Result of *Phosphate* Solubilization Test

Phosphates Solubilization

A total of 10 selected isolates of PGPB were tested for the production of *phosphate* solubilization. Out of 10 isolates tested, 6 showed positive results. The solubilization of *phosphate* was highest in isolates of S4-C1, S5-C2, S5-C4 and S3-C2 (Fig. 8).

Assay for Ammonia production

A total of 10 selected isolates of PGPB were tested for the production of *ammonia*. Out of 10 isolates tested, 9 showed positive. The production of IAA was highest in isolates of S4-C1 and S5-C2 (Fig. 9). The growth promoting activity of rhizobacterial isolates is summarized in Table 2.

Table 2. Growth promoting activity of Rhizobacterial isolates

Rhizobacterial Isolates	IAA production	PO_4 solubilization	NH_3 production
S4-C3	++	++	++
S4-C1	+	++	+++
S5-C2	-	++	-
S5-C4	+	+	++
S3-C2	++	++	+
S5-C5	+	+	+
S3-C3	+	-	++
S3-C1	+	-	++
S5-C3	+	-	+
S5-C6	+	-	+

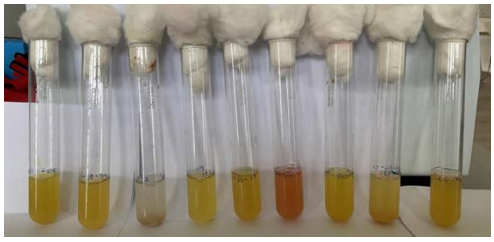


Figure 9. Result of *Ammonia* Production Test

3.3 Gram staining

The results obtained from the assay of the physiological properties of isolates were used to select some isolates for further study. These isolates were tested for gram staining to understand the gram staining result and their morphology (Fig. 10, 11 and 12). Three isolates were selected (S4-C3 – Gram-negative bacilli; S5-C4 – Gram-positive cocci; S3-C2 – Gram-positive bacilli) after gram staining and these isolates were used to make the bacterial consortia using all possible combinations.

Figure 10. Selected Isolates Streaked on NA

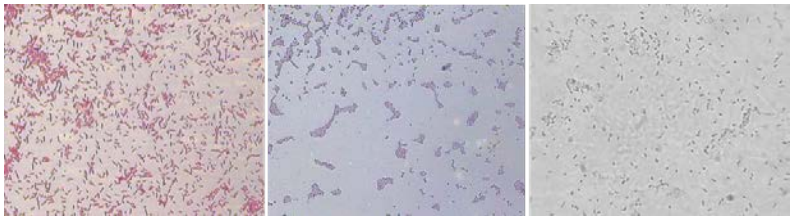


Figure 11. Gram's Staining Result of the Selected PGPR (From left to right: S4-C3 – Gram-negative bacilli; S5-C4 – Gram-positive cocci; S3-C2 – Gram-positive bacilli)

Figure 12. Selected PGPR isolates maintained on NA slants

3.4 Measurement of Stomatal Conductance and Growth

The rhizosphere is a complex ecology in which plant impacts on soil microbes and microorganism effects on plants interact and are interdependent. The key processes through which PGPR contributes directly to plant phytohormone production, such as auxins, cytokinins, and gibberellins, as well as boosting plant nutrition through mineral solubilization, siderophores, and enzyme production (Taiz *et al*., 1991). The turgor pressure and osmotic potential of guard cells are directly related to stomatal conductance (Buckley *et al*., 2013). Stomatal conductance is a function of stomatal

Vol. 2023 (2) 48 – 56
ISSN 2688-3651 [online]

density, stomatal aperture, and stomatal size (Ziegler *et al.*, 1987). Reductions in stomatal conductance prevent further decreases in water potential by reducing transpiration; also, reductions in water potential can induce stomatal closure, resulting in lowered stomatal conductance. Coating the seeds with PGPR consortia and micronutrients showed beneficial effects on plants adaptation to drought stress and increased seed yield (Fig. 13) (Table 3).

Figure 13. Leaf Porometer (AP4 leaf porometer, Delta-T Devices, UK) used during the experiment (measurement of stomatal conductance)

Stomatal conductance decreased in some of the treatments when they were imposed to drought stress. One of the first responses of plants to drought is stomatal closure, restricting gas exchange between the atmosphere and the inside of the leaf. 'Treatment 2a' showed the lowest stomatal conductance and a better seed yield under drought conditions.

Table 3. Influence of Coating Seeds with Micronutrients and Bacterial Consortia on Stomatal Conductance (mmol m−2 s−1) of Cluster Bean: List of Treatments Showing Respective Experimental Design. The values are the mean and standard deviation of three replications.

Set Code	Treatment	Length (cm)	Stomatal Conductance (mmol m−2 s−1)	Cluster bean yield in g (per plant)
1a	1+2+MN	36 ± 1.14	107 ± 2.61	82.5 ± 2.56
1b	1+2	29 ± 0.74	106 ± 3.44	62.6 ±2.27
2a	1+2+3+MN	49 ± 1.87	77 ± 1.87	118.5 ± 3.65
2b	1+2+3	40 ± 1.47	144 ± 4.81	102.6 ± 2.07
3a	1+3+MN	32 ± 0.98	128 ± 3.65	71.2 ± 1.42
3b	1+3	34 ± 0.87	160 ± 4.03	59.4 ± 0.98
4a	2+3+MN	39 ± 1.04	166 ± 3.75	48.3 ± 1.36
4b	2+3	28 ± 1.01	176 ± 2.93	39.6 ± 1.87
Controls				
C1	No coating	32 ± 0.73	180 ± 4.97	43.6 ± 0.57
C2	Starch	31 ± 0.81	175 ± 4.71	32.4 ± 0.71
C3	Micronutrients	29 ± 0.42	178 ± 3.49	51.4 ± 1.33

MN - Micronutrients
1, 2, 3 – Bacterial isolates

4. Conclusion

Micronutrients viz., zinc, boron and potassium molybdate in combination with standardized PGPR consortia significantly influenced the stomatal conductance of leaf, helping in the better establishment of seedlings and higher yield. Seed coating of cluster bean with the combination of micronutrients with PGPR consortia was shown to be effective in increasing the yields. Stomatal conductance decreased in several of the treatments when they were applied to drought stress. Under drought conditions, "Treatment 2a" displayed the lowest stomatal conductance and the best seed yield. These micronutrients along with the PGPR consortia may be supplied to the plants through seed treatment to improve the stand establishment, better seed yield under drought conditions, micronutrient contents in grain in the

Vol. 2023 (2) 48 – 56
ISSN 2688-3651 [online]

cluster bean crop. Being an easy and cost-effective method, seed treatment by polymer coating offers an attractive option for resource-poor farmers through its pronounced effect during the early stage of seedling establishment. This positive impact on stomatal conductance was attributed to the synergistic effects of micronutrients and the beneficial bacterial consortia, which likely enhanced nutrient availability, facilitated nutrient uptake, and promoted plant growth. This approach can serve as an effective strategy to enhance crop performance, optimize resource utilization, and contribute to sustainable agriculture. Further studies are warranted to elucidate the underlying mechanisms involved in the observed improvements and to evaluate the long-term effects of seed coating on soil health and crop resilience.

Acknowledgments

We would like to thank Prayoga Institute of Education Research for funding and material support and for giving us the opportunity to work in a research laboratory as high school student. Many thanks to our supervisor, Dr. Venkata Krishna B for guiding us throughout the duration of the research study. This report and the research behind it would not have been possible without his exceptional support. Thanks to Dr. Nagabhushana K S, Research Director, Prayoga for his constant cooperation, encouragement, and facilities. We also thank Mr. Raju and Mr. Narayan, who helped us procure all the milk samples. We would like to extend our gratitude towards the Samvida management for introducing us to this opportunity.

References

Ashraf, M., & Iram, A. (2005). Drought stress induced changes in some organic substances in nodules and other plant parts of two potential legumes differing in salt tolerance. *Flora*, 200(6), 535–546.

Bhattacharyya, P. N., & Jha, D. K. (2011). Plant growth-promoting rhizobacteria (PGPR): emergence in agriculture. *World Journal of Microbiology and Biotechnology*, *28*(4), 1327–1350.

Bric, J. M., Bostock, R. M., & Silverstone, S. E. (1991). Rapid In Situ Assay for Indoleacetic Acid Production by Bacteria Immobilized on a Nitrocellulose Membrane. *Applied and Environmental Microbiology*, *57*(2), 535–538.

Buckley, T. N., & Mott, K. A. (2013). Modelling stomatal conductance in response to environmental factors. *Plant, Cell & Environment*, *36*(9), 1691–1699.

Cappucino, J. C., & Sherman, N. *"Microbiology A Laboratory Manual," 3rd Edition, Benjamin/Cumming Pub. Co., New York, 1992. - References - Scientific Research Publishing*. (n.d.). Www.scirp.org. Retrieved October 21, 2022, from

Damour, G., Simonneau, T., Cochard, H., & Urban, L. (2010). An overview of models of stomatal conductance at the leaf level. *Plant, Cell & Environment*, 33, 1419–1438.

Ghadamgahi, F., Tarighi, S., Taheri, P., Saripella, G. V., *et al.* (2022). Plant Growth-Promoting Activity of Pseudomonas aeruginosa FG106 and Its Ability to Act as a Biocontrol Agent against Potato, Tomato and Taro Pathogens. *Biology*, *11*(1), 140.

Goudar, R., Srinivasa, V., & Lakshmana, D. (2016). Genetic variability and divergence studies in cluster bean (Cyamopsis tetragonoloba L.) under hill zone of Karnataka. *Legume Research*, *40*(2), 237–240.

Handiganoor, M. G., Patil, S. B., & Vasudevan, S. N. (2018). Effect of Seed Coating Polymer and Micronutrients on Stomatal Conductance and Resistance at Different Growth Stages of Pigeonpea. *Advances in Research, 17*(4), 1-8.

Johnson, S. E., Lauren, J. G., Welch, R. M., & Duxbury, J. M. (2005). comparison of the effects of micronutrient seed priming and soil fertilization on the mineral nutrition of chickpea (Cicer arietinum), lentil (Lens culinaris), rice (Oryza sativa) and wheat (Triticum aestivum) in Nepal. *Experimental Agriculture, 4,* 427-448.

Vol. 2023 (2) 48 – 56
ISSN 2688-3651 [online]

Kawamitsu, Y., Driscoll, T., & Boyer, J. S. (2000). Photosynthesis during Desiccation in an Intertidal Alga and a Land Plant. *Plant and Cell Physiology*, *41*(3), 344–353.

Kumar, D., & Rodge, A. (2012). Status, scope and strategies of arid legumes research in India-A review. *Undefined*.

Meena, N. K., & Asrey, R. (2018). Tree Age Affects Postharvest Attributes and Mineral Content in Amrapali Mango (Mangifera indica) Fruits. *Horticultural Plant Journal*, *4*(2), 55–61.

Mengel, K., Kirkby, E. A., Kosegarten, H., & Appel, T. (2001). Nutrient Uptake and Assimilation. *Principles of Plant Nutrition*, 111–179.

Michael, J. L., & Burton, E. P. (2002). *Microbiology: Laboratory Theory and Application - PDF Drive*. (n.d.). Www.pdfdrive.com. Retrieved October 21, 2022.

Microbial Activity in the Rhizoshere. (2006). In K. G. Mukerji, C. Manoharachary, & J. Singh (Eds.), *Soil Biology*. Springer Berlin Heidelberg, 349.

Punia, A., Yadav, R., Arora, P., & Chaudhury, A. (2009). Molecular and morphophysiological characterization of superior cluster bean (Cymopsis tetragonoloba) varieties. *Journal of Crop Science and Biotechnology*, *12*(3), 143–148.

Quddus, M. A., Anwar, M. B., Naser, H. M., Siddiky., *et al.* (2020). Impact of Zinc, Boron and Molybdenum Addition in Soil on Mungbean Productivity, Nutrient Uptake and Economics. *Journal of Agricultural Science*, *12*(9), 115.

Sabahelkheir, M.K., Abdalla, A.H. and Nouri, S.H. (2012). Quality Assessment of Guar Gum (Endosperm) of Guar (Cyamopsis tetragonoloba). *ISCA Journal of Biological Sciences, 1, 67-70. - References - Scientific Research Publishing*. (n.d.). Www.scirp.org. Retrieved October 21, 2022, from

Shinde, P., & Vasudevan, S. R. D. S. N. (2016). Influence of seed polymer coating with micronutrients and foliar spray on seed yield of chickpea (cicer arietinum l.). *Legume Research*, *40*(4), 704–709.

Singh, B., Natesan, S., Singh, B., & Usha, K. (2005). Improving zinc efficiency of cereals under zinc deficiency. *Current sci, 88,* 36-14.

Sitepu, I. R., Hashidoko, Y., Aryanto, A., Turjaman., *et al.* (2008). Studies on functional bacteria of indonesian tropical forest plants for biorehabilitation of degraded lands. *Indonesian Journal of Forestry Research*, *5*(1), 21–36.

Srivastava, S., Anees, K., & Ramani, R. (2011). Promise of Guar meal. *Undefined*, 38–39.

Stomatal function. (1987). In E. Zeiger, G. D. Farquhar, & I. R. Cowan (Eds.), *Library Catalog (Blacklight)*. Stanford University Press.

Surange, S., Wollum II, A. G., Kumar, N., & Nautiyal, C. S. (1997). Characterization of *Rhizobium* from root nodules of leguminous trees growing in alkaline soils. *Canadian Journal of Microbiology*, *43*(9), 891–894.

Taiz, L. and Zeiger, E. (1991). *Plant physiology Mineral nutrition. The Benjamin Cummings Publishing Co., Inc. Redwood City, 100-119. - References - Scientific Research Publishing*. (n.d.). Www.scirp.org. Retrieved October 21, 2022.

Journal of Research High School

Vol. 2023 (2) 57 – 61
ISSN 2688-3651 [online]

How Can We Accelerate the Early Identification of Biomarker Glycoprotein NMB for The Early Detection of Breast Cancer

Harshita Jinaga[1*]

[1]American School of Bombay, Mumbai, India
*Corresponding Author: harshitajinaga@gmail.com

Advisor: Bryan Nolasco, bnolasco.jr6@gmail.com

Received December 21, 2022; Revised June 16, 2023; Accepted, July 7, 2023

Abstract

The current methods for screening for any type of breast cancer involve the use of radiation levels or some form of waves. Widely used methods are CT scans, mammograms, MRIs, and ultrasounds, majority requiring the use of harmful radiation. A mammogram uses low energy x-rays to examine the human breast, a CT scan involves the usage of ionizing radiation, and an ultrasound transmits sound waves into the body. By identifying specific biomarkers like Glycoprotein NMB which are overexpressed in conditions like TNBC, we can find a more efficient and less harmful method of diagnosis. Usage and identification of these biomarkers to diagnose TNBC would also help with quicker and earlier detection compared to existing screening measures. The results of clinical trials experimenting with Glycoprotein NMB and various cancers point in a positive direction.

Keywords: Glycoprotein NMB, Triple Negative Breast Cancer, Radiation

1. Introduction

Triple negative breast cancer (TNBC) is regarded as one of the most aggressive and rarest cancers. The lack of hormone receptors, including estrogen and progesterone receptors, on breast cells renders it particularly difficult to treat with hormone or HER2 therapy. TNBC affects approximately 13 in 100,000 women each year in the USA and has an overall mortality rate of 77% (Chen. L, et. al, 2016). TNBC accounts for approximately 15% of all breast cancers diagnosed all over the world (Sheng. J., 2023). Given its rapid course of progression, there also tends to be a short window for timely intervention. Currently screening and detection methods rely heavily on mammography's, ultrasounds, and MRIs. These forms of testing sometimes result in false negatives or fail to detect a cancer that's present, making them a suboptimal form of testing. The sensitivity of the current forms of testing ranges from 80%-95% and the specificity ranges from 90%-98%. A new protein of interest in recent studies is Glycoprotein NMB (GPNMB) due to its expression in specific types of cancers. GPNMB is a protein in humans encoded by the GPNMB gene (McCarthy, A. M., et al., 2021). This glycoprotein has been found to be overexpressed in many cancers, including triple negative breast cancer. Given the aforementioned limitations in current screening and detections measures, I believe, it could be beneficial to investigate targeting GPNMB as a potential biomarker for TNBC. Blood tests can be used to identify the presence of the biomarker GPNMB in the patient's blood. GPNMB is a form of testing for various types of cancer, not just TNBC. Currently there has been limited research performed that analyzes the relationship between GPNMB and TNBC in patients. My aim is to identify whether identification of GPNMB is a potential method for detecting early-stage TNBC. Through blood tests testing for GPNMB we can find a more accurate and efficient way of detecting TNBC.

Triple negative breast cancer is a subtype of breast cancer known for its rarity and difficulty to treat. The reason the term triple negative breast cancer was coined for this type of cancer was due to the lack of estrogenic or progesterone receptors (ER or PR) as well as limited production of the protein known as HER2. It does not have crucial receptors which are commonly found in other types of breast cancer (Scott et al., 2019). This makes it particularly

Vol. 2023 (2) 57 – 61
ISSN 2688-3651 [online]

difficult to identify and treat. The route of metastasis in TNBC correlates with the survival of TNBC patients which have brain metastases (the poorest survival indicator), followed by liver metastases, pleura, bone, and lastly lung (Prakash O., et al., 2020). The cause of TNBC is yet to be identified, but researchers seem to correlate the BRCA 1 genetic mutation as an indicator. As a tumour suppressor gene BRCA1 plays an important role in DNA repair mechanisms and preventing aberrant growth, but an oncogenic hit could reverse the course and cause cells to be more vulnerable to cancer (Stewart, 2019). Along with this specific reproductive and lifestyle-related factors link to an increased risk of TNBC (Cho, B., et al., 2021).

A glycoprotein is a protein consisting of oligosaccharide chains which are covalently attached to an amino acid side chain. Glycans can attach to either lipids or amino acids, through a bond or process known as glycosylation (Raas, 2023). Glycosylation is a process where secreted extracellular proteins are glycosylated. Glycoproteins play an important role in cell-cell interactions as well as in integral membrane proteins. Another essential role played by glycoproteins is their role in primary and secondary immune responses allowing for white blood cell migration around the body (Dent R., et al., 2007). The synthesis of a glycoprotein occurs between two organelles: the endoplasmic reticulum and the Golgi apparatus. Glycoprotein NMB is a transmembrane glycoprotein which is encoded in humans by the GPNMB gene. GPNMB is characterized by two transcript variants which encode 560 and 572 amino acids. GPNMB is usually expressed in different cell types like melanocytes, osteoclasts, osteoblasts, dendritic cells, along with being overexpressed in different types of cancer. Previously GPNMB was a gene which did not express or was expressed poorly in highly metastatic cell lines and in metastatic human melanoma cell lines. But recent studies have identified high GPNMB expressions in aggressive melanoma, glioma, as well as breast cancer specimens.

Triple negative breast cancer is a rare cancer which affects approximately 13 in 100,000 women each year (Sussman D., et al., 2014). Around 91% of all women with TNBC are still alive after 5 years of diagnosis. Although, if the cancer spreads to the lymph nodes near the breast the 5 year survival rate reduces to 65%. Lastly, if the cancer spreads to distant areas in the body, the 5 year survival rate further reduces to (Birdia A., et al., 2019). TNBC being more aggressive than other cancers, causes it to be harder to treat and also increases the chances of recurrence. TNBC is usually treated with surgery, radiation, chemotherapy, or a combination of all three. Chemotherapy, a medicine which kills cancer, is usually the first treatment therapy used for TNBC. This is often followed by radiation therapy.

The current method to identify TNBC is through ultrasounds, MRIs, and mammograms. These methods are used to first detect breast cancer in a patient. Once an ultrasound, MRI, or mammogram indicates that the patient has breast cancer, a biopsy is conducted. The biopsy is used to assess cells of a patient and further check for estrogen, progesterone, and HER2 receptors to determine the person's breast cancer subtype. This process is lengthy, and also includes the patient being exposed to a lot of radiation, which is unhealthy. TNBC tends to double in size every 180 days, or every 6 months, so it is crucial for its identification to be quick. By using GPNMB as an indicator for TNBC, the diagnosis of this subtype of cancer can be much faster, and avoid using unhealthy radiation.

As mentioned previously, GPNMB is highly expressed in cancers or tumour's like melanoma, glioma, breast cancer, and cholangiocarcinoma. Currently researchers are exploring the potential of targeting GPNMB in osteosarcoma. Using human osteosarcoma samples, researchers are identifying the expression of GPNMB (Weterman, 2020). A similar research is being conducted for glioblastoma tissues and the mediation of glioma progression (Zhang, 2017). In both of these studies there has been high expression of GPNMB and clinical trials are being conducted to assess the efficiency of targeting GPNMB in patients.

My aim is to identify a method of detecting triple negative breast cancer which does not include radiation or biopsies. TNBC is one of the most difficult breast cancers to detect, so finding a way to detect it fast and accurately is crucial. Using GPNMB is a relatively new method as clinical trials and research regarding it are still being pursued. I would like to research more regarding the identification of TNBC through GPNMB, a new and upcoming method.

Based on preliminary studies showing that GPNMB has been effective in diagnosing other conditions like Alzheimer's disease, Parkinson's disease , Non-Alcoholic Fatty Liver Disease, etc (Budge, et al., 2017). Various clinical studies and trials have been conducted, experimenting with GPNMB and neurological diseases, but very few experimenting with cancer, especially TNBC.

I hypothesize that using glycoprotein NMB to target or identify patients with triple negative breast cancer will be a possible, effective, and efficient method of diagnosis.

Vol. 2023 (2) 57 – 61
ISSN 2688-3651 [online]

2. Methods

Based on my interest in TNBC I analysed recent studies that have conducted experiments on GPNMB. I looked through various trials that had been conducted so far about this topic. Since the use of GPNMB is still upcoming and fairly new, research on it is limited. Additionally, research with the use of GPNMB as a biomarker of TNBC is further specialized.

There are multiple ways experiments with GPNMB are carried out. Each experiment uses a tumor sample and tests it to find expression of GPNMB. Most studies or clinical trials use immunohistochemistry which is a laboratory method that uses antibodies to identify whether certain antigens are present (Horwitz S, et al., 2019).

Some of the criteria I used while picking the clinical trials to use for my research were ensuring that the data was focused specifically on TNBC. I also tried to use sources which were acclaimed or reliable like Nature and PubMed by the National Library of Medicine. I also focused on ensuring the experimental design used either immunohistochemistry or gene expression data. This is so that it would give me results which were specific to what I was looking for.

The data for the first clinical trial was collected through testing with immunochemistry, in-silico survival analysis, cell culture reagents and transfection, invasion assay as well as western blot analysis. 759 patients who tested positive for TNBC and underwent primary surgery were enrolled into the clinical trial. Various clinico- pathological characteristics like TNM staging, histological grade, as well as tumor type were all matched through the WHO classification system. The clinical parameters collected by oncologists and past medical records. The representative areas of each tumor were carefully selected and constructed into tissue microarrays (TMA). The histology score through a semi-quantitative assessment of the percentage of positive-stained carcinoma cells as well as the staining intensity was assessed. The range of H scores, which ranged from 0 to 300 were evaluated by pathologists and categorized into high and low subgroups.

The immunostainings were performed on 4 micrometer paraffin-embedded tissue sections. The slides were then placed in a 10mM citrate buffer for 20 minutes in a pressurized heating chamber. The tissues were incubated with the antibodies against GPNMB like E-cadherin and vimentin. After the slides were taken out and washed with phosphate-buffer saline, bound antibodies were detected. Tissues were identified as ER or PR positive breast cancer based on the amount of nuclei stained. If the percentage was higher than 10 it was classified as ER or PR positive breast cancer, and as HER-2 positive breast cancer if 3+HER2 expression was identified. After identifying this the Kaplan-Meier analysis was done and western blot analysis was done to quantify protein levels. Invasion assay was conducted to identify the number of invaded cells, and research was analyzed and concluded through statistical analysis.

The second clinical trial was conducted with a lower number of patients. The significance of GPNMB expression was addressed by analyzing GPNMB levels in various gene expression data sets. The clinical trial involved using two independent tissue arrays from human breast tumors. To identify the GPNMB expression IHC stainings were used. The IHC stainings were analyzed to identify the intensity of GPNMB expression. The sample was analyzed and the percentage of positively stained carcinoma cells were analyzed. The GPNMB expressing breast cancer cells that were identified were further analyzed to understand the significance of using GPNMB as a prognostic biomarker.

Both the clinical trials used women of ages between 20-70. The women were from different ethnicities and different backgrounds, but age was a variable that remained constant. The patients enrolled in both trials consisted of women ranging from different severities of TNBC. In the first trial majority of the women had undergone primary surgery, whereas in the second trial many women hadn’t.

3. Results

There have been primarily positive results with identifying GPNMB as a marker for TNBC. Various clinical trials have been done which test the expression of GPNMB in different cancers.

In one clinical trial, 759 specimens were collected, and in immunohistochemistry it was found that GPNMB was expressed in different subtypes of cancer but was significantly higher in TNBC. The data was collected from 759

Vol. 2023 (2) 57 – 61
ISSN 2688-3651 [online]

patients with primary breast cancer with a median follow-up of 74 months. Among the patients with TNBC, the median age was 55 years. The Kaplan-Meier analysis revealed that the overexpression of GPNMB in TNBC was associated with an advanced or worse prognosis, in specific distant metastasis, especially visceral metastasis which includes the lungs, liver, and brain. The trial results showed that in TNBC the mean and medium of GPNMB expression were 102.9 and 87.5 with the interquartile ranges of 45.0 to 165.0 (Huang, et al, 2021). The average H-scores of GPNMB in TNBC subtypes were also significantly higher compared to those of non-TNBC subtypes. These results demonstrated that GPNMB might be overexpressed in TNBC preferentially. In silico analysis there was high expression of the mRNA of GPNMB and it can be correlated with distant metastasis. Additionally, GPNMB was overexpressed in TNBC in the silico analysis. The protein levels of Twist and MMP2 were also upregulated by GPNMB overexpression in TNBC cells. These graphs and diagrams from the study in the Figure 1 indicate that GPNMB expression does correlate with breast cancer. The graph on the bottom right indicates that low GPNMB expression results in a long survival time compared to high GPNMB expression. The pictures help us identify the GPNMB expression intensity in different samples.

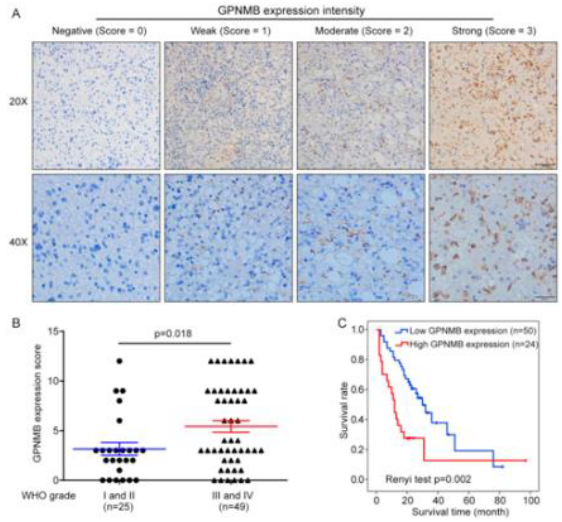


Figure 1. GPNMB expression correlated to triple negative breast cancer

In another study the expression of GPNMB was studied specifically in TNBC. Through immunohistochemical analysis, findings show that GPNMB is commonly expressed in breast tumors. These results were found in two studies specifically. In the first study, it was found that GPNMB was detected in approximately 71% of breast tumors. Whereas, in the second study GPNMB was detected in 64% of breast tumors, and additionally 10% of tumors expressed this gene in the tumor epithelium (Burris, et al., 2009). Through this study it was found that GPNMB expression in the tumor epithelium could be seen as an independent prognostic indicator of breast cancer recurrence. Epithelial GPNMB expression was highest in triple negative breast cancer and found to be a prognostic marker for this type of breast cancer subtype. In addition to this, GPNMB expression in breast cancer cells is also capable of promoting cell migration, invasion, as well as metastasis in both in-vitro and in-vivo.

Both clinical trials resulted in positive results proving that GPNMB is a potential and effective indicator of TNBC. From the first clinical trial the results indicated that GPNMB might be overexpressed in TNBC preferentially and that it may be an indication of worse prognosis. Even so, it does indicate the presence of TNBC. The second clinical trial the results also pointed in a positive direction. GPNMB was found in 71% of the breast tumors tested, which is a relatively high percentage. Along with this there was a finding of another independent prognostic indicator: GPNMB expression in the tumor epithelium.

4. Discussion

Overall I believe that GPNMB is a promising indicator for TNBC. A lot more research is yet to be conducted, but according to clinical trials which are currently being undergone, results seem to be pointing in a positive direction.

More research and clinical trials need to be conducted to ensure that GPNMB can be used to identify TNBC. Many factors like distant metastasis and recurrence of TNBC need to be looked into. Along with this more detailed trials need to be conducted which specifically look into GPNMB's effects at the protein and gene levels. Antibodies against GPNMB like glembatumumab vedotin need to also be considered.

Though results seem promising, there are many limitations regarding identifying GPNMB as a breast tumor. In both the clinical trials the follow up procedures with the patients were of different lengths and did not take into

Vol. 2023 (2) 57 – 61
ISSN 2688-3651 [online]

consideration each one's unique course of medication and treatment.

New information which has been found through the clinical trials above also could be potential indicators. GPNMB expression specifically in tumor epitheliums has not been looked at, but it could be another solution to the current problem we are facing.

5. Conclusion

TNBC is an extremely difficult form of breast cancer to detect, and overall using GPNMB to accelerate the identification seems to prove useful. According to the clinical trials currently being conducted, the results look promising and effective. Though much more research still needs to be conducted, usage of biomarker GPNMB has the potential to accelerate the identification of TNBC.

References

Biochemistry, Genetics and Molecular Biology | ScienceDirect Topics. (n.d.). Www.sciencedirect.com. https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/

Ferreira, A., & Weller, M. (2017). *Differential Association of the Lifestyle-Related Risk Factors Smoking and Obesity with Triple Negative Breast Cancer in a Brazilian Population*. *18*(6), 1585–1593. https://doi.org/10.22034/apjcp.2017.18.6.1585

Glycoprotein Synthesis - an overview | ScienceDirect Topics. (n.d.). Www.sciencedirect.com. https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular- biology/glycoprotein-synthesis

GPNMB. (2023, March 3). Wikipedia. https://en.wikipedia.org/wiki/GPNMB

Huang, Y.-H., et al., (2021). Expression pattern and prognostic impact of glycoprotein non-metastatic B (GPNMB) in triple-negative breast cancer. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-91588-3

Medical Definition of Glycoprotein. (n.d.). RxList. https://www.rxlist.com/glycoprotein/definition.htm

National Cancer Institute. (2019). *NCI Dictionary of Cancer Terms*. National Cancer Institute; Cancer.gov. https://www.cancer.gov/publications/dictionaries/cancer- terms/def/immunohistochemistry

Roth, M., Barris, D. M., Sajida Piperdi, Kuo, V., Everts, S., Geller, D. A., Houghton, P. J., E. Anders Kolb, Hawthorne, T., Gill, J., & Gorlick, R. (2015). *Targeting Glycoprotein NMB With Antibody- Drug Conjugate, Glembatumumab Vedotin, for the Treatment of Osteosarcoma*. *63*(1), 32–38. https://doi.org/10.1002/pbc.25688

Shabir, O. (2020, June 15). *What is a Glycoprotein?* News-Medical.net. https://www.news- medical.net/health/What-is-a-Glycoprotein.aspx

Taya, M., & Hammes, S. R. (2018). Glycoprotein Non-Metastatic Melanoma Protein B (GPNMB) and Cancer: A Novel Potential Therapeutic Target. *Steroids*, *133*, 102–107. https://doi.org/10.1016/j.steroids.2017.10.013

Triple-Negative Breast Cancer. (2019). Centers for Disease Control and Prevention. https://www.cdc.gov/cancer/breast/triple-negative.htm

Triple Negative Breast Cancer: What Is It, Symptoms, Treatment & Survival Rate. (n.d.). Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/21756-triple-negative-breast-cancer-tnbc

Tseng, L. M., et., (2013). Distant metastasis in triple- negative breast cancer. *Neoplasma*, *60*(03), 290–294. https://doi.org/10.4149/neo_2013_038

Shabir, O. (2020, June 15). *What is a Glycoprotein?* News-Medical.net. https://www.news- medical.net/health/What-is-a-Glycoprotein.aspx

Journal of Research High School

Vol. 2023 (2) 62 – 68
ISSN 2688-3651 [online]

Mechanical Design in Modular Reconfigurable Robotics: A Literature Review

Owen Y.H. Lai[1*]

[1]Newport High School, Bellevue, WA, USA
*Corresponding Author: lai.owen@outlook.com

Advisor: Robert Doran, doranr@bsd405.org

Received December 15, 2022; Revised May 27, 2023; Accepted, June 26, 2023

Abstract

This paper reviews recent research on mechanical aspects of modular reconfigurable robots (MRRs). Many newe MRRs have developed unique designs and hardware updated from the older, basic designs. The author compile research of recent innovations in MRRs to include in the contents of this paper. The paper will first begin with reviewin the background of the MMRs and the influences it took on to create it. MMRs were inspired by the idea of one joinin together with others to achieve a task not possible for a single individual. Based on this idea, MMRs have modules tha form different structures by attaching to one another in different positions. With these unique differences to othe robotics, the paper then discusses the numerous benefits MRRs brings to the table compared to traditional robots. Th discussion of MRRs' benefits also goes into territory of applications in space as MRRs could be used for i extraterrestrial expansion. Further in this paper, reoccurring designs of MRRs are analyzed, comparing older design to recent ones. Here, the paper goes over recent MRRs and review their designs and mechanisms in depth, breakin down how they are built and how they work. The paper then examines the sensing in MRRs and recent innovations i making modules more aware of their surroundings. The differences between old and recent for both internal an external sensing used in modules will also be discussed and covered. Throughout this paper, the discussions of th similarity and lack of change/innovation from old MMRs to recent MRRs will be covered. Then finally, the paper wi conclude with key findings determined by the author and what further research needs to be done on the topic.

Keywords: Modular Reconfigurable Robots, Space Applications, Robotics, Modular Robots, Truss Robots, Bonding Robots, Lattice Robots, Modular Self-reconfigurable Robots, Chain Robots.

1. Introduction

Over the past four years, around 17,500 scientific articles covering modular reconfigurable robots (MRRs) have risen up at the time this was written. This style of robotics has been inspired by the idea of individuals joining together to achieve a task not feasible for a single individual (Liu, et al., 2021). MRRs are robots made up of individual modules that work together as a system (Seo, et al., 2019). Modules reconfigure themselves into different shapes and forms by connecting to and detaching from other modules of the system, in other words, reconfiguring themselves. With MRR systems being able to function properly no matter the number of modules there are, these robotic systems are able to be repaired easily and able to continue its tasks without certain parts (Zhao, et al., 2022). This makes MRRs better than non-modular robotics in certain applications where repairs might not be convenient or even possible. MRRs also bring a factor of versatility to the table as well. Because of their ability to reconfigure into all different shapes and sizes, MRRs can change their configuration to fit the task they need to accomplish next. However, there are some challenges to MRRs that traditional robotics often solves. Modular robots need to be connected with one or more modules, making a system, in order to work properly. A single module of a MRR by itself cannot function as intended. This makes the issue of needing to make perfect modules to have a steady fixed connection to communicate to one another (Zhao, et al., 2022). Imperfect manufacturing could result in a system failure because the system relies on a

Vol. 2023 (2) 62 – 68
ISSN 2688-3651 [online]

connection that could easily be broken with poor dock alignment as an example. However, with these disadvantages aside, MRRs show great potential to solve a multitude of problems that traditional single-body robots cannot as later presented in this paper.

One of the earliest pioneers of MRRs is Yim et al.'s PolyBot (2000). The PolyBot displayed the use of individual modules that could attach to and detach from other modules. The PolyBot has around one thousand citations at the time this was written and influenced many other current MRRs' designs. Yim et al. (2003) also remarks in another paper that the PolyBot and other articulated robots would be ideal in space applications as it met requirements in space manipulation.

As more MRRs start showing up in the robotics industry, this paper aims to cover the different examples of new MRRs and its applications. This paper is structured as follows: Section 2 will give an overview and explanation of the space applications of MRRs; Section 3 reviews common MRR designs by going in depth into each robot of each design, seeing how they functions with their design; Section 4 will cover the hardware inside the MRRs and how it works; and concluding the paper in Section 5.

2. Applications of MRRs

MRRs have and are being used in vastly different applications over the years it has been developed upon. One of the most prominent fields MRRs are being used for is space applications. With the commercialization of space by private corporations becoming more relevant in recent years, extraterrestrial living solutions often cite back to MRRs (Romanov, et al., 2021). Robots could be sent in swarms to other planets or moons to build shelter autonomously. This reduces the risks and resources required to have humans build shelter outside of Earth. MRRs are the most ideal type of robot for these missions as they carry their own advantages over a traditional robot too. MRRs are better at navigating through obstacles that might be abundant in other planets/moons and are easier to maintain. Because of their module design, they can adapt to the environment easily by changing their shape and size. This also help with maintenance as faulty or trapped modules could be easily detached to the system of modules and the robots would still function together fine with the remaining modules in the system.

3. Commonly Seen Designs in MRRs

There have been various different designs of MRRs, where all are designed to take on different challenges and problems. This paper will be discussing commonly seen designs in MRRs.

3.1 3D Lattice design

The 3D lattice design has its MRRs be connected in a web-like network where modules could be attached to another module in any angle that is 90 degrees away from another point of connection. One of the earliest pioneers of a 3D lattice reconfiguration design in MRRs were Romanishin et al. and the M-blocks (2013). Romanishin (2013) designed modules that were cubes that could be set next to or stacked on top of other modules and would magnetically attach to one another. These modules could move up and down vertically to a different plane by rotating 180 degrees in a certain direction and move horizontally in the same plane by rotating 90 degrees. However, this reconfiguration process is only able to be done locally. Modules are only able to move to a position where there is a neighboring module present already. In past years, there has been barely any innovation or change in the designs of the 3-D lattice MRRs. The use of the hybrid design in MRRs have caused the 3D lattice design to decrease in popularity due to the hybrid design utilizing features from 3D lattice MRRs and other designs. Nisser et al.'s ElectroVoxel (2022) and Leal-Naranjo et al. (2021) both use designs almost identical to the M-Block. Both researches proposed a cube shaped module that had the same movements as the M-Block. Also, 3D lattice MRRs all seem to use magnets to attach to other modules.

Vol. 2023 (2) 62 – 68
ISSN 2688-3651 [online]

3.2 Chain design

The chain design has its MRRs be attached to one another in a chain-like architecture where modules connect to each other from the back and front, forming a link. A great example of this design is Murata et al.'s M-TRAN (2002). The M-TRAN modules are made up of two boxes attached with a link that provides the boxes to rotate. These boxes are of a semicircle prism shape with the rounded side of each box facing each other. On the flat side of each box, there are magnets that attach to the flat side of boxes of other modules. Murata (2002) designed the M-TRAN this way so it's modules can metamorphose into different configuration that can generate a walking motion without external help. The chain design offers length for the system of the MRR as modules could easily connect on either side and build a long structure. Other than the M-TRAN and other older MRRs, the chain design is not seen in newer innovations because designs of newer MRRs are usually of a hybrid design, a combination of 3D lattice, chain, and other designs.

3.3 Hybrid design

The hybrid design combines elements of the 3D lattice and chaining to assemble the MRRs. The best example of a MRR of this design is the SMORES-EP module by Liu et al. (2021). These modules are cube-shaped with four connectors that could attach to other modules on the sides of the cube. Three of these connectors are shaped in a circle and could work as wheels and rotators that spin the modules attached to it. The fourth connector functions as the bottom of the module and is square-shaped without the ability to rotate. They designed the modules this way to add vertical movement to the robot without have the modules stack on top of one another like Romanishin et al.'s M-blocks (2013). Hauser et al.'s Roombots (2020) are another example of the hybrid design for MRRs. Hauser's (2020) hybrid design takes 3-D lattice designs and added a new element to make it do so much more. Roombots modules have a rounded cube shape and are diagonally split into hemispheres. Each module has six connectors, one on every side of the cube, that can attach to other modules. The hemispheres of each module have three connectors on it and can rotate around the other hemisphere. The split in each module sets it apart from other designs of MRRs. Another unique hybrid MRRs design is found in Zhao et al.'s SnailBot (2022). As the name tells, Zhao (2022) designed his SnailBot modules to be snail shaped (sphere shaped) with a base that holds six wheels. These wheels are split into two trains, one with two sets of wheels and other with one set of wheels. Splitting the wheels into two trains makes climbing onto other modules possible. The train with two sets of wheels lifts up and rolls onto another module and the train with one set of wheels stays onto the surface it was originally placed on, whether it be another module or a flat surface. This way the module can attach to different modules magnetically and maneuver its way to different spots while on other modules. Zhao (2022) with the SnailBot takes older designs of 3-D lattice MRRs and their actuations and makes them smooth and less jagged, like how a snail would move. There seems to be a lot more variation in hybrid designs of MRRs compared to the other designs already covered. Like the 3-D lattice MRRs, hybrid MRRs are seen to have a more cube like shaped module with the exception of the SnailBot. However, hybrid designs tend to be used more in MRRs because of its versatility, getting the most range in motion and having the most convenient bonding motion. Additionally, hybrid MRRs, like 3-D lattice MRRs, also have magnets to assist the bonding processes or be the bond between two modules.

3.4 Truss design

The truss design in MRRs has its modules connected to one another in at least two points of contact and more often than not are connected to several other modules. Connections between truss modules are also manipulatable, which is the main method for the MRR moving. This way, when the modules change their size and/or degree they are connected to the other modules, the whole system is moved like that. One example of this design is Qin et al.'s TrussBot (2022). The TrussBot modules are triangular pyramid shaped. Qin (2022) designed each corner of each module has a hook with a rubber band that is used to attach to other modules, so that the connection flexible between the two modules. There are two different types of modules depending on object on the inside: battery modules and actuated modules. Each battery module holds a battery that supplies power to the whole system of modules with the

Vol. 2023 (2) 62 – 68
ISSN 2688-3651 [online]

other battery modules. Actuated modules have a servomotor in the center of the module that pulls in and releases a fishing line that is connected to other modules. This is for moving the system of modules by contracting and stretching the fishing line like a tendon. Another innovation in truss designs in MRR comes from Stuart et al. (2021) and the Balloon Animal Robot. Modules of the Balloon Animal Robot are untraditional where they don't make up the MRR shape but instead are cinches that are wrapped around an inflated tube that makes up the MRR. Stuart (2021) designed the Balloon Animal Robot to be soft so that the robot would be able to reconfigure into varying landscapes in vastly different situations. Each module is straw-shaped with rollers on the inside side that rotate with a drive mechanism. This way modules can move up and down the inflated tube they are wrapped around on. Each module also has a winching mechanism and a cinch mechanism. The winching mechanism is a motor attached to a string that is attached to other modules' winching mechanism on the other end. This is for pulling and pushing parts of tube closer or farther together by pulling in and releasing the string. The cinch mechanism is a motor also attached to a string, but the string is wrapped around the inflated tube and both ends of the string are attached to the same module. This is for closing the tube to make joints for bending the tube by reeling in the string. The combination of these mechanism makes the inflated tube be to twist and form structures from a previously straight inflated cylinder. Truss design in MRR vary in structural build up, but seem to have string to actuate movements. This design tends to be applicable in more specific situations and is more difficult to work with.

4. Hardware

When it comes to MRRs, the hardware that is needed to make the design come to life is equally as important as the design. In this section, the paper covers how modules bond together, move from rest, and sense their orientation and environment.

4.1 Actuation

Every MRR has a method of action physically interact with other modules or the environment around it. This action is known as actuation. In this sub-section, the paper will talk about actuations and the actuators, the mechanism that made the action happen, that MRRs employ for bonding and movement.

Bonding Actuation

Bonding actuation is the action that physically connects or attaches one module to another module. MRRs, similar to design, have vastly different methods for bonding to other modules and different mechanisms to do so. One common method of bonding is by using magnets. Romanishin et al.'s M-Block (2013), Murata et al.'s M-TRAN (2002), and Zhao et al.'s SnailBot (2022) are some examples of this bonding actuation, both using permanent magnets to bond modules together. However, Liu et al.'s SMORES-EP (2021) also uses magnets to bond modules together, but SMORE-EP uses electro permanent magnets instead of the popular permanent magnets. Liu's use of electro permanent magnets allows SMORE-EP modules to easily attach and detach by just turning on and off the magnet instead of having a separate mechanism to break the bond between modules that is needed for modules that use permanent magnets to bond together. This trend of using electro permanent magnets in bonding actuations is rising as Nisser et al.'s ElectroVoxel (2022) noted that they would research into the use of electro permanent magnets in the future for the ElectroVoxel. Putting aside this small difference between the uses of permanent magnets and electro permanent magnets, the actuation process for both of them are similar. Magnets are put on the surface of a module that is meant for attaching to other modules and magnetism holds the modules together. Another common method of bonding is using a hooking mechanism. Hauser et al.'s Roombots (2020) and Qin et al's TrussBot (2022) are some examples of this bonding actuation, both having their modules use hooks to latch onto other modules. Hauser (2020) has his modules hook onto connection plates of other modules and Qin (2022) has his modules hook onto rubber bands in other modules to physically attach both modules together. The use of hooks compared to magnets has its arguments. Hooks provide a stronger link between modules and more degrees of freedom for movement between the two linked modules than magnets do. However, magnets give modules an easier and faster method for attaching and detaching

Vol. 2023 (2) 62 – 68
ISSN 2688-3651 [online]

to other modules. It is ultimately dependent on the situation and task the MRR is in or tasked to do that determine which actuation is favored.

Movement Actuation

Movement actuation is the action that moves modules of an MRR to other modules or to the environment. With varying designs, MRRs have many different ways of moving from one place to another. One common movement actuation is rotating a part of a module. Whether this be wheels on Liu et al.'s SMORES-EP (2021) or hemispheres of Hauser et al.'s Roombots (2020), many MRRs just use a motor to rotate parts of a module to alter the module's position. In the case of Liu and Zhao et al.'s SnailBot (2022), their MRRs have wheels to roll to other places, whether it be to another module or the environment. Another movement actuation in MRRs uses string. Qin et al.'s TrussBot (2022) and Stuart et al.'s Balloon Animal Robots (2021) both use string to move their modules closer together or farther apart. In both robots, two modules that are distanced apart are physically connected via string and have a winching mechanism to reel in or let go the string to move their respective system. Lastly, another movement actuation in MRR is internal actuation. This type of actuation has its MRR move to different places without interacting with other modules and the environment physically. This is the case for Romanishin et al.'s M-Blocks (2013) and Nisser et al.'s ElectroVoxel (2022). Romanishin (2013) designs his modules to move from its bonded modules by using a flywheel inside each module to create torque strong enough to break away from the force from the magnets and move to a different location. On the other hand, Nisser (2022) designs his movement actuation mechanism to use magnets on each module to repel and attract other magnets on other modules to move the module to a different spot. Movement actuations in MRRs, more so than bonding actuations, are heavily dependent on the design of the MRR.

4.2 Sensing

In order to have physical actuations function correctly, MRRs also have to have some method of detecting other modules and the environment. Without sensing, modules would not know where to traverse to and bond with other modules if they were out of their initial positions. In this section, the paper will address innovations in internal sensing and external sensing.

Internal Sensing

Internal Sensing in MRRs is seen to be needed to do two things, provide precision for docking and track the position of each module's degrees of freedom. Degrees of freedom (DOF) stands for the parts of a module that have actuated movement (Liu, et al, 2021). Precision docking is incredibly important to ensure self-reconfiguration of MRRs as modules in order to connect to other modules reliably have to dock accurately. Docking is often seen being assisted by infrared and linear hall-effect sensors, for example in Hauser et al.'s Roombots (2020). The infrared and linear hall-effect sensors on each module of the Roombots detect the existence of other modules in the area and the distance it is from other modules or the environment. With the sensors put under the connection plates of the Roombots modules, the combination of infrared and linear hall-effect sensors with magnets guides the connection plates of modules to the perfect position for the bonding actuation. The position of each module's degrees of freedom is needed in order to know the next actuation for reconfiguration or moving. To know the positions of these degrees of freedom on each module, sensors such as encoders and estimation algorithms are used to calculate the positions. For example, Liu's SMORES-EP (2021) uses customizable sensors and an algorithm to predict each module's wheels'/connection plates' orientation. SMORES-EP modules use this info to determine how much to rotate each wheel/connection plate to get to the system's desired configuration or location.

External Sensing

External sensing in MRRs helps modules detect the environment and other modules in the area. External sensing is less essential to the MRRs function compared to internal sensing, but this doesn't excuse its existence. A great example of external sensing in MRRs is Hauser et al's Roombots (2020). Hauser (2020) proposes incorporating a camera to modules of the Roombots to enable computer vision for the robot. This could be used for detecting the

Vol. 2023 (2) 62 – 68
ISSN 2688-3651 [online]

environment, for example human beings, that could be useful for the furniture applications, what Roombots is intended to be applied to. However, most MRRs use a bus system where modules can communicate to one another as long as they are directly connected to a system of other already interconnected modules.

5. Conclusion

This paper reviews recent research of modular reconfigurable robots (MRRs), focusing on the mechanical aspects. Many newer MRRs have developed new, unique designs and hardware. This paper reviews and highlights the vastly different designs used and hardware implemented in recent MRRs, updating the repeated designs and hardware of older MRRs. Key findings in recent research of MRRs prove that mechanisms ideal for assisting in bonding or being the bond between two modules are dependent on the task the MRR is aimed to complete. Also, cube shaped modules are found to be more popular in design because its shape helps with the MRR's simplicity. However, further research needs to be done on external sensing of the environment. With space applications possible for MRRs, research should be conducted on MRRs being able to sense and adapt to its surroundings. Additionally, further research could be done on chain MRR designs, as there was a lack of chain MRRs compared to other designs like the popular hybrid design. With the chain design offering the option to be lengthy that other designs can't do as easily, research should be conducted to use the length of chain design MRRs in unique applications.

Acknowledgments

I would like to acknowledge the support and help given from Robert Doran, Koshu Takatsuji, and the UCI X GATI program.

References

Hauser, S. et al. (2020). Roombots extended: Challenges in the next generation of self-reconfigurable modular robots and their application in adaptive and assistive furniture. Robotics and Autonomous Systems, 127, 103467. https://doi.org/10.1016/j.robot.2020.103467

Leal-Naranjo, J.-A., Fichera, S., & Paoletti, P. (2021). Towards a modular robotic platform for construction and manufacturing. 2021 29th Mediterranean Conference on Control and Automation (MED), 1197–1202. https://doi.org/10.1109/MED51440.2021.9480264

Liu, C. et al. (2021). SMORES-EP, a Modular Robot with Parallel Self-assembly (arXiv:2104.00800). arXiv. https://doi.org/10.48550/arXiv.2104.00800

Murata, S. et al. (2002). M-TRAN: Self-reconfigurable modular robotic system. IEEE/ASME Transactions on Mechatronics, 7(4), 431–441. https://doi.org/10.1109/TMECH.2002.806220

Nisser, M. et al. (2022). ElectroVoxel: Electromagnetically Actuated Pivoting for Scalable Modular Self-Reconfigurable Robots. 2022 International Conference on Robotics and Automation (ICRA), 4254–4260. https://doi.org/10.1109/ICRA46639.2022.9811746

Qin, Y. et al. (2022). TrussBot: Modeling, Design, and Control of a Compliant, Helical Truss of Tetrahedral Modules. 2022 International Conference on Robotics and Automation (ICRA), 4218–4224. https://doi.org/10.1109/ICRA46639.2022.9812295

Romanishin, J. W., Gilpin, K., & Rus, D. (2013). M-blocks: Momentum-driven, magnetic modular robots. 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, 4288–4295. https://doi.org/10.1109/IROS.2013.6696971

Romanov, A. M., Yashunskiy, V. D., & Chiu, W.-Y. (2021). A Modular Reconfigurable Robot for Future

Vol. 2023 (2) 62 – 68
ISSN 2688-3651 [online]

Autonomous Extraterrestrial Missions. IEEE Access, 9, 147809–147827. https://doi.org/10.1109/ACCESS.2021.3124951

Seo, J., Paik, J., & Yim, M. (2019). Modular Reconfigurable Robotics. Annual Review of Control, Robotics, and Autonomous Systems, 2(1), 63–88. https://doi.org/10.1146/annurev-control-053018-023834

Stuart, A. D., Hammond, Z. M., & Follmer, S. (2021). Balloon Animal Robots: Reconfigurable Isoperimetric Inflated Soft Robots. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 6941–6947. https://doi.org/10.1109/IROS51168.2021.9635842

Yim, M., Duff, D. G., & Roufas, K. D. (2000). PolyBot: A modular reconfigurable robot. Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065), 1, 514–520 vol.1. https://doi.org/10.1109/ROBOT.2000.844106

Yim, M. et al. (2003). Modular Reconfigurable Robots in Space Applications. Autonomous Robots, 14(2), 225–237. https://doi.org/10.1023/A:1022287820808

Zhao, D., & Lam, T. L. (2022). SnailBot: A Continuously Dockable Modular Self-reconfigurable Robot Using Rocker-bogie Suspension. 2022 International Conference on Robotics and Automation (ICRA), 4261–4267. https://doi.org/10.1109/ICRA46639.2022.9811779

Journal of Research High School

Vol. 2023 (2) 69 – 77
ISSN 2688-3651 [online]

Barriers to Eating Disorder Treatment: A Review of Socioeconomic Obstacles and Gaps in the Healthcare System

Deshna Shekar[1*]

[1]Evergreen Valley High School, San Jose, CA, USA
*Corresponding Author: deshna.cs@gmail.com

Advisor: Juliana Kotz, j.kotz@ucla.edu

Received January 4, 2023; Revised June 7, 2023; Accepted, June 29, 2023

Abstract

With an increasing number of cases every year, eating disorders have the highest mortality rate of all mental illnesses globally. However, more than half of the diagnosed patients do not receive suitable treatment. This paper identified the barriers to both mental and physical care in eating disorder treatment and their impacts on patients' course of recovery. The reasons for these obstacles are highlighted based on patients' personal experiences and information compiled from different online sources, such as research articles and surveys found through Pubmed and Google Scholar. Research, surveys, and patients' feedback highlighted that treatment barriers are primarily due to high financial costs for care, stigma surrounding recovery, bias against minorities, inconsistent treatment, and insufficient outpatient care. Furthermore, from a medical standpoint, primary care physicians only have surface level knowledge on how to provide treatment to patients but end up having to do so because there are a limited number of eating disorder specialists. If patients are diagnosed and given adequate care in the beginning stages and throughout their eating disorders, then not only will they be more likely to avoid falling into critical mental and physical conditions, but they will likely reach the point of recovery quicker and more efficiently.

Keywords: Eating Disorders, Treatment Barriers, Recovery, Primary Care Physicians, Eating Disorder Specialists.

1. Introduction

Eating disorders (EDs) are psychological disorders with the highest mortality rates among all psychiatric illnesses, with the number being 5.86 times more than the general mortality rate (Neale and Hudson, 2020). Worldwide there has been an 18-fold increase in reported cases, from 0.2% to 3.7%, affecting over 30 million people in the U.S. itself (Maguen et al., 2018; Qian et al., 2013). The average onset age is mainly women between ages 15-19, however research shows that the incidence rate is continuing to become more frequent in younger children, especially those from ages 8-17 (Neale and Hudson, 2020; van Eeden et al., 2021).

Per standard definition, eating disorders are characterized by disordered eating habits and an abnormal relationship with food. However, studies and surveys show that there is most likely an underlying cause for these behaviors whether that be cultural, experiential, familial and/or societal (Fisher, 2006). Additionally there are many types of eating disorders, some of the more common ones being Anorexia Nervosa (AN), Bulimia Nervosa (BN), and Binge-Eating Disorder (BED), each with varying traits.

Despite the chronicity of eating disorders, only a small portion of patients receive treatment, with the average waiting time to accessing care being 5.28 years from onset to treatment-seeking (Hamilton et al., 2021). The principal reasons for these delays in acquiring care are socioeconomic treatment barriers, disruptive transitions between treatment, workforce shortages, and missed diagnoses.

The process of initially reaching out for treatment can be difficult, especially regarding (1) social and weight stigmas amongst patients and caregivers, (2) an under-prioritization of treatment for those from gender, sexual and/or

Vol. 2023 (2) 69 – 77
ISSN 2688-3651 [online]

ethnic minorities, (3) financial troubles in beginning and continuing care, and (4) patients' reluctance to seek treatment. Collectively, these hurdles can lead to missed diagnoses, as well as patients dipping into chronic and severe mental and physical states before receiving adequate treatment. During treatment, abrupt changes in programs and between healthcare providers can upset treatment routines that had formed for patients and guardians, causing relapses and readmissions to the hospital or other care. Even after treatment and a transition to outpatient care, if parents/guardians are not well equipped with resources to support patients, inpatient care may be required once again. This shift between forms of care can result in states of diminished mental health and motivation to recover.

Another, more unspoken cause of unbalanced treatment is workforce shortages. There are few specialists in the eating disorder field and referrals to them usually require waiting for large amounts of time, in many cases over a year (Brown et al., 2016). Furthermore, patients are rarely directed to these specialists, mainly because primary care physicians cannot identify the eating disorder and only intervene once the patient's physical health is impaired, even if their eating disorder has psychologically manifested prior to that.

In relation to workforce shortages, missed diagnoses are a huge factor in a person's eating disorder journey. Without a 'sick' appearance or significant weight loss, many caregivers cannot identify eating disorders unless there is medical instability. This is a harmful mindset as many patients can have atypical eating disorders where they are not underweight but suffer from the same condition and display the same behaviors as those that are. What's more is that general doctors do not screen patients in normal checkups, which could potentially help them catch any intrusive thoughts a child could be having. Both of the above factors impact caregivers as well as patients. Individuals that do not feel 'sick' enough, often will not access care, paving the way for their state to become more chronic.

This research review will explore the above factors, as each are hindrances to eating disorder patients' course of treatment. We believe this paper has brought attention to the many issues in the eating disorder medical system through a collection of quantitative and qualitive data. Our analysis illustrates that more effective eating disorder care can be accomplished by breaking stereotypes about the illness, improved training to identify and treat patients from early onset, and better understanding of the psychopathologies of the condition.

2. Socioeconomic Treatment Barriers

2.1 Financial

A major factor that prevents patients from seeking treatment is the cost of accessing care. The National Eating Disorder Association (NEDA) has found that inpatient care can cost more than $30,000 a month, with an average of $500-$2000 a day (Thompson and Park, 2016). Moreover, only ten states require insurers to cover treatment for anorexia and bulimia, which are two of a large list of eating disorders. Even with this requirement, insurers generally choose not to cover patients that are not medically unwell, leading to families having to pay out of pocket for psychological, nutritional, and other treatment needs. The medical director of Austen Riggs Center, a non-profit residential treatment facility, states, "insurance companies are using the "wrong criteria" for what makes something medically necessary. They pay enough only to stabilize someone's condition, but not actually to improve it." (Dangor, 2018). Having financial costs deter patients from seeking care can cause them to require more intensive treatment, which can be costlier than the financial coverage of providers in outpatient settings.

2.2 Bias against Ethnic, Gender, and Sexual Minorities

Eating disorders affect those of all genders and backgrounds but are thought to mainly impact the White population. Due to the fact that eating disorder studies are predominantly done with White individuals and that those from ethnic minorities generally prefer concealing their eating disorders, a small percentage of non-Caucasian patients receive care. Demonstrating this, a study showed that in a group of Asians, African-Americans, Latinos, and Caucasians, Caucasians were more likely to receive a recommendation for treatment from a health professional (60%) than individuals from the other backgrounds (31%), despite having the same symptoms (Thompson and Park, 2016).

Vol. 2023 (2) 69 – 77
ISSN 2688-3651 [online]

People from all ethnicities and nationalities can develop eating disorders, making it imperative that they are all given equal opportunity to acquire treatment.

Eating disorders are typically perceived as only affecting the female population, despite there being an increasing number of them in males. Males are largely underrepresented in mental health research and significantly in eating disorders, creating a lot of stigma in their care. A population survey estimates that about 10 million males in the US have or will experience an eating disorder during their life. In fact, approximately one in four patients with eating disorders are male. Regardless of these rising numbers, a study found that 25.6% of males with EDs were not referred to specialized care (Sangha et al., 2019). This is prominently because of the stigma around males seeking care for their EDs, and many being given a wrong diagnosis or none at all. Furthermore, less than 1% of ED research is centered around males, so there are many unknowns about their eating disorder psychopathologies (Lavender et al., 2017). As more physicians and researchers understand eating disorders in male patients, patients can be given accurate diagnoses and adequate care, with less stigma and misunderstandings.

A commonly overlooked minority in regard to eating disorders and mental health in general is the LGBTQ community. However, research shows that those from the LGBTQ population are more likely to develop eating disorders than those that are cisgender, due to reasons such as body dissatisfaction, minority victimization, and other mental health disorders, amongst other factors. For instance, a survey of about 300,000 college students found that trangender students had over 4 times more risk of being diagnosed with anorexia or bulimia than cisgender female students (Gordon, n.d.). An additional research found “approximately 54% of LGBT adolescents have been diagnosed with a full-syndrome eating disorder during their lifetime, with an additional 21% suspecting that they had an eating disorder at some point during their life.” (Parker and Harriger, 2020). Nevertheless, it is harder for LGBTQ patients to access care, not only in the eating disorder field, but the medical field in general. For example, a study found that 1 in 4 transgender people were denied equal health care (Jaffee et al., 2016). The LGBTQ community already faces many mental and physical health disparities, which themselves contribute to their development of eating disorders. As caregivers begin to learn more about how to approach LGBTQ patients, patients can start receiving more appropriate and equal care.

2.3 Stigma

Although there are treatment options where patients with eating disorders can be cared for in residential treatment, where they are secluded from the public, they ultimately must continue to recover and re-enter society. Not only can this be challenging after diagnosis, but stigma in healthcare and from the public can lead to hesitancy in initially asking for help as well. In fact, a study recorded stigma causing longer illness durations and “that a substantial proportion of people with anorexia nervosa waited a long time before visiting a physician and starting treatment because they were afraid of being stigmatized. The average period between disease onset and the first medical consultation or treatment initiation was around 8–9 months.” (Brelet et al., 2021). Stereotypes and stigmatization of those with eating disorders can provoke negative feelings around seeking and being in treatment, as well as causing hesitancy to recover.

Even away from health care, the general population tends to have negative perceptions of individuals with eating disorders. For instance, a survey aimed at the public, found that “more than a third of the respondents blamed people with EDs for their situation, thought that people with EDs would be able to pull themselves together if they wanted to, and found communication with them challenging.” (Brelet et al., 2021). Stigma is not just communicated orally but is largely involved online with pro-ED websites. An article studying social support noted that when websites advertise “ED support” it is generally superficial and correlates to the maintenance of ED behaviors (Rouleau and von Ranson, 2011). As people learn more about eating disorders, they can become more cognizant of how to support and interact with anyone with the illness, encouraging a helpful environment.

Vol. 2023 (2) 69 – 77
ISSN 2688-3651 [online]

3. Disruptive Transitions Between Levels of Care and Treatment Programs

Transitions in eating disorder care are very common and are normally required when patients' support needs change. The most beneficial transitions are well thought out and thoroughly discussed with the patient, their family, and their current and future treatment teams. However, due to circumstantial urgencies and limited availability in programs, transitions can be uncalled for and disruptive for a patient, ultimately leading to negative outcomes in their care.

The transition from adolescent to adult care can be overwhelming, especially when a patient's entire treatment team suddenly changes. A study that found "out of 76 cases of transition between CAMHS [Children and Adolescent Mental Health Services] and general adult services (not confined to[but including] eating disorder services), only 4 met all their criteria for optimal transition." (Crockett et al., 2017). Furthermore, a carer shares their daughter's experience when transitioning to an adult:

"My daughter feels that she's not being taken seriously. Her anorexic thoughts and body dysmorphia are still strong. She sees an ED nurse every 2–3 weeks if she's lucky. That's it." (Crockett et al., 2017). Adults are automatically expected to be more independent in dealing with their EDs even when they need the same support as when they were adolescents, leading to relapses and instability in treatment.

Another notable transition can occur when patients go to a university/college, away from home and any treatment established there. A paper found that compared to the general eating disorder population, 14% of females and 4% of males that go to universities have clinical EDs. Moreover, 54% of females and 19% of males exhibit sub-clinical EDs that can easily develop into clinical ones. In relation, a study found that many individuals forget to consider treatment options before going to and at universities because of the notion of having a "fresh start." This on top of a newfound freedom and distance from caregivers can cause a decline in patients' health and stability (Webb and Schmidt, 2020).

An additional change is the transition between inpatient and outpatient care, affecting patients of all ages equally. A qualitative study recorded a patients' feedback on transition support:

"I feel it's a process that you need to go through with a key worker or a therapist or whatever, just to walk you through it and make sure you're actually ok with the stepping-down process and the discharge process … because it is such a big deal." (Bryan et al., 2022). Many factors need to be taken into account when transitioning from treatment, especially when moving to less intensive care, and this process should be as smooth as possible for both the patient and their family.

There is also distress going from outpatient care to more intensive treatment. Patients can feel a complete loss of control and loss of contact with family, friends, and their work/education. Parents may feel relief that their child is receiving adequate care, but may also feel a loss of control regarding their child's treatment (Crockett et al., 2017). The above transitions in treatment demonstrate that although smooth adjustments are required for patients to have positive outcomes in care, they are hard to come by for several reasons. If more time was taken to discuss and consider these changes, individuals could receive the support they need in a manner that they can navigate and manage.

4. Parents/Guardians Lack of Experience for Outpatient Care

One of the most efficient methods for treating eating disorders in outpatient care is Family-Based Therapy (FBT), a type of treatment where parents are empowered to help the patient manage their eating disorder behaviors with guidance from a trained professional. A 2009 study at the Child Guidance Clinic in Philadelphia found that there were successful outcomes in 86% of the patients (Webb and Schmidt, 2020). Moreover, a study conducted at Maudsley Hospital in London with a group of adolescents with AN, reported that 90% of patients who took part in family therapy fell into "good" or "intermediate" categories (regarding weight and presence of disordered symptoms) compared to the 18% in individual therapy (Bryan et al., 2022). Although FBT doesn't work for all families, it is proven to have the most successful outcomes compared to other forms of therapy.

However, few patients and families have access to these services, with there being only 52 certified FBT providers within 13 U.S. states, and with wait times from receiving an initial evaluation to starting therapy ranging from 5-10

Vol. 2023 (2) 69 – 77
ISSN 2688-3651 [online]

months (Bruett et al., 2022). With this treatment gap, parents must take charge and care for their child, but many are inexperienced and struggle without resources. In a qualitative study recording parent experience with dealing with a child with an eating disorder, one father states:

"She became weaker and I stood alone without much help from anyone. This (illness) is unknown to most of us when it occurs. You do not have anything to compare it with. You have no experience or references. You feel helpless, do not know what to do or whether you are pushing the right buttons or making the right moves." (Karlstad et al., 2021). Many parents are not equipped with resources to care for their child, especially if their child hasn't been part of specialized treatment.

This helplessness can have health impacts on parents/guardians as well. In a study measuring distress levels in individuals caring for someone with an eating disorder, about 36% had high mental difficulties and 17% faced high psychological stress (Whitey et al., 2007). If parents received proper resources and support when caring for their child with an eating disorder, more intensive treatment could be avoided and more patients could stay stable at home.

5. Workforce Shortage

5.1 Limited Number of Specialists

The majority of patients with eating disorders end up requiring or seeking treatment from eating disorder specialists. However, there is a shortage of specialists in this field, which leads to the majority of patients' referrals being disregarded. A 2020 study found, "Systemic factors such as clinicians' concerns about use of manualized evidence-based treatments and a shortage of clinicians trained in these can impede dissemination and implementation of evidence-based treatments so that even those who do seek care may not receive appropriate treatment." (Mclean et al., 2020). Additionally, psychological support is also crucial but can be limited to individuals as well. To resolve this problem, caregivers have started providing online treatment, which has proven advantageous, but still is restricted to the availability of therapists. A report states, "While these approaches [online therapy] clearly have advantages in making treatment more available, they still rely on trained eating disorder therapists who are in short supply." (Cooper and Bailey-Straebler, 2015). As more physicians and therapists receive credentials for eating disorder specialization, more patients will receive the care that they require.

5.2 Few Patients Receive Specialist Care

Primary physicians play a key part in diagnosing eating disorders in patients and guiding families to access more specialized/professional care. However, this can be a huge demand of these doctors, who often do not have experience with eating disorders. In a qualitative study, in which general physicians that dealt with eating disorders spoke about their experiences, themes of uncertainty regarding what to tell and how to treat patients/families were identified. In these instances, many physicians would seek guidance from more experienced eating disorder specialists or refer back to the treatment guidelines (Fogarty and Ramjan, 2016). Primary physicians' borderline knowledge of eating disorders can make it hard for them to guide patients and families in pursuing care.

Situations where general physicians care for eating disorder patients rather than specialists are common due to few specialist services being available and accepting patient referrals. A review from south London about adult eating disorder patients' treatment found, "only two-thirds of referrals resulted in an assessment. Of these, only three-quarters entered treatment. Finally, only half of those who entered treatment completed it." (Waller et al., 2009). The numbers presented display that only one-fourth of the patients referred to specialist services actually complete treatment. If there are more eating disorder services, not only in densely populated areas, but in more rustic areas as well, there will be a higher chance of more eating disorder patients reaching recovery.

Vol. 2023 (2) 69 – 77
ISSN 2688-3651 [online]

6. Missed Diagnoses

6.1 Ineffective Screening

Similar to screening of other mental illnesses, such as depression and anxiety, there are effective methods to screen individuals for eating disorders, such as the Eating Disorder Screen for Primary Care (EDS-PC) and the SCOFF questionnaire. However, many primary care settings do not routinely use these screening mechanisms to detect eating disorders in individuals. Screening can pose negative effects such as false-positive results that lead to added stigma and unnecessary financial expenditures (Davidson et al., 2022). On the other hand, without screening, 50% to 80% of eating disorders go undetected or incorrectly diagnosed (Baranauskas et al., 2022). As of right now, a universal screening system for eating disorders hasn't been implemented because there is not enough evidence to back up if its benefits outweigh its cons. Regardless, a survey found that in a group of medical providers, 54% of them supported applying a screening mmethod (inville et al., 2010), which professionals are continuing to perfect. As screening policies become more accurate and efficient, they can be useful assets to medical providers for diagnosing eating disorders.

6.2 Psychological Manifestation vs. Physical Manifestation

Despite eating disorders being psychological disorders, they are mainly diagnosed based on physical manifestation, such as weight loss or emaciated appearance. However, over half of eating disorder patients do not meet the full hospital admission criteria but have the same psychological issues and behaviors (Fisher, 2006). In fact a study recorded that the percentage of patients with atypical anorexia (who have all the symptoms of anorexia, but are not underweight) had risen from 8% to 47% in four years (Strand et al., 2020). Additionally, a participant (parent) of a Cambridge University study regarding patient and guardian treatment experience gave feedback about the treatment they received:

"One of the main challenges is getting help early enough. There is too much attention paid to weight loss and not enough to the behavior patterns and so they are dangerously low in weight before you get referred…" Individuals generally only receive care once reaching a medically dangerous place, which typically happens due to the little attention given to the psychological aspect of their illness.

Patients also mentally struggle at home despite reaching healthier weights, but are not given the required support, which can lead to relapses and readmissions. This is highlighted by a patient taking part in the above study:

"I spoke to my psychiatrist once I got to a healthy weight, and she said there is not much they can do if I'm not ill which really upset me…I wish she hadn't said there was nothing she could do because I wasn't ill physically because that makes you think you need to make yourself ill to get help which shouldn't be the case!" (Mitrofan et al., 2019). Lulls in proper care can deprive a patient of sufficient support in reaching "complete" recovery, and in some cases lead to severe relapses.

6.3 Few Patients Reach Out for Help

The majority of individuals with eating disorders are reluctant to seek treatment; those that do so are mainly adolescents brought to caregivers by their guardians (NHS England, 2015). The chief reasons for this are: patients' not acknowledging problematic behaviors, shame around asking for help, mainly due to perceived and experienced stigma, and few known resources to reach out to. A study shows that in a sample of individuals with eating disorders, only 30.7% felt the need to reach out for help (Sonneville and Lipson, 2018). Another study by a 'Health Body Image' Program communicated similar results, with 41.7% of the screened patients claiming they didn't need therapy, 19.9% being unsure of how serious their condition was, and 19.5% not having time for treatment (Bryant et al., 2022). This reluctance to receive care highly impacts the amount of time before being treated and both the physical and mental severity of the patient's illness when they are.

Vol. 2023 (2) 69 – 77
ISSN 2688-3651 [online]

6.4 Lack of Experience

The majority of patients with eating disorders are treated by their primary care physicians (PCPs) or general doctors, many of whom tend to have little experience surrounding eating disorders and provide inadequate support to patients. A survey in Queensland found that out of 136 primary care providers, 73% felt that they weren't experienced enough in treating eating disorders (Bryant et al., 2022). Furthermore, many providers also lack the understanding that eating disorders are illnesses not decisions. For example, a study found that out of a total of 82 caregivers (32 nurses and 50 residents), 58.2% of them felt the patient was responsible for their disease. Additionally, 68.8% of the residents and 45.2% of the nurses had often felt frustrated with patients (Raveneau et al., 2013). These statistics show that a large percentage of caregivers do not truly understand the nature of an eating disorder, which can prevent them from giving the emotional support needed for patients.

7. Conclusion

The results of this review illustrate the various challenges that present themselves in eating disorder treatment. Socioeconomic barriers, such as financial difficulties, stigma, bias, disruptive treatment transitions and inadequate outpatient care, all contribute to the discouragement of seeking care and the stunting of progress made in recovery. Furthermore, the process of receiving treatment is challenging itself due to shortages in the number of caregivers, namely specialists, and the lack of patient referrals from primary physicians that are actually treated in specialized care centers. Lastly, primary physicians generally are not experienced in diagnosing patients especially without a difference in physical appearance. Eating disorders are mental illnesses and many individuals go undiagnosed because they do not appear “sick.” What’s more is that there is not an effective screening practice in place as a baseline to help physicians diagnose patients. This can be detrimental as few patients truly reach out for help on their own or without a guardian doing so for them. In the future, there can be more positive outcomes in treatment if healthcare providers and the general population learn more about how to provide a safe and supportive environment for patients with eating disorders.

References

Baranauskas, M., Kupčiūnaitė, I., & Stukas, R. (2022). Potential triggers for risking the development of eating disorders in non-clinical higher-education students in emerging adulthood. Nutrients, 14(11), 2293. https://doi.org/10.3390/nu14112293

Brelet, L., et al. (2021). Stigmatization toward people with anorexia nervosa, bulimia nervosa, and binge eating disorder: A scoping review. Nutrients, 13(8), 2834. https://doi.org/10.3390/nu13082834

Brown, A., et al. (2016). The Freed Project (first episode and rapid early intervention in eating disorders): Service model, feasibility and acceptability. Early Intervention in Psychiatry, 12(2), 250–257. https://doi.org/10.1111/eip.12382

Bruett, L. D., et al. (2022). Development of evidence-informed bridge programming to support an increased need for eating disorder services during the COVID-19 pandemic. Journal of Eating Disorders, 10(1). https://doi.org/10.1186/s40337-022-00590-1

Bryant, E., et al. (2022). Screening, assessment and diagnosis in the eating disorders: Findings from a rapid review. Journal of Eating Disorders, 10(1). https://doi.org/10.1186/s40337-022-00597-8

Clark Bryan, D., et al. (2022). Transitions from intensive eating disorder treatment settings: Qualitative investigation of the experiences and needs of adults with anorexia nervosa and their carers. BJPsych Open, 8(4). https://doi.org/10.1192/bjo.2022.535

Vol. 2023 (2) 69 – 77
ISSN 2688-3651 [online]

Cooper, Z., & Bailey-Straebler, S. (2015). Disseminating evidence-based psychological treatments for eating disorders. Current Psychiatry Reports, 17(3). https://doi.org/10.1007/s11920-015-0551-7

Crockett, P., Morris, J., & Winston, T. (2017). Managing Transitions When the Patient has an Eating Disorder. College Report. Retrieved January 3, 2023, from https://www.rcpsych.ac.uk/docs/default-source/improving-care/better-mh-policy/college-reports/college-report-cr208.pdf?sfvrsn=ace4d5db_2

Dangor, G. (2018, December 16). Burden of proof: Insurance companies are refusing to pay for mental healthcare and regulators are letting them. patients are left to fend for themselves. CUNY Academic Works. Retrieved January 3, 2023, from https://academicworks.cuny.edu/gj_etds/309/

Fisher, M. (2006). Treatment of Eating Disorders in Children, Adolescents, and Young Adults. Pediatrics in Review. Retrieved January 4, 2023, from http://nutricion.sochipe.cl/subidos/catalogo3/Eating_disorders.pdf

Fogarty, S., & Ramjan, L. M. (2016). Factors impacting treatment and recovery in anorexia nervosa: Qualitative findings from an online questionnaire. Journal of Eating Disorders, 4(1). https://doi.org/10.1186/s40337-016-0107-1

Gordon, A. R. (2018, February 20). Shining a light on gender identity and eating disorders. National Eating Disorders Association. Retrieved January 3, 2023, from https://www.nationaleatingdisorders.org/blog/shining-a-light-on-gender-identity-and-eating-disorders

Hamilton, A., et al. (2021). Understanding treatment delay: Perceived barriers preventing treatment-seeking for eating disorders. Australian & New Zealand Journal of Psychiatry, 56(3), 248–259. https://doi.org/10.1177/00048674211020102

Jaffee, K. D., Shires, D. A., & Stroumsa, D. (2016). Discrimination and delayed health care among transgender women and men. Medical Care, 54(11), 1010–1016. https://doi.org/10.1097/mlr.0000000000000583

Karlstad, J., et al. (2021). The experiences and strategies of parents' of adults with anorexia nervosa or bulimia nervosa: A qualitative study. BMC Psychiatry, 21(1). https://doi.org/10.1186/s12888-021-03345-5

Lavender, J. M., Brown, T. A., & Murray, S. B. (2017). Men, muscles, and eating disorders: An overview of traditional and muscularity-oriented disordered eating. Current Psychiatry Reports, 19(6). https://doi.org/10.1007/s11920-017-0787-5

Linville, D., et al. (2010). Medical providers' screening, training and intervention practices for eating disorders. Eating Disorders, 18(2), 110–131. https://doi.org/10.1080/10640260903585532

Maguen, S., et al. (2018). Screen for disordered eating: Improving the accuracy of eating disorder screening in primary care. General Hospital Psychiatry, 50, 20–25. https://doi.org/10.1016/j.genhosppsych.2017.09.004

McLean, S. A., et al. (2020). Credentialing for eating disorder clinicians: A pathway for implementation of clinical practice standards. Journal of Eating Disorders, 8(1). https://doi.org/10.1186/s40337-020-00332-1

Mitrofan, O., et al. (2019). Care experiences of young people with eating disorders and their parents: Qualitative study. BJPsych Open, 5(1). https://doi.org/10.1192/bjo.2018.78

Neale, J., & Hudson, L. D. (2020). Anorexia nervosa in adolescents. British Journal of Hospital Medicine, 81(6), 1–8. https://doi.org/10.12968/hmed.2020.0099

NHS England. (2015, July). Access and Waiting Time Standard for Children and Young People with an Eating Disorder. Retrieved January 4, 2023, from https://www.england.nhs.uk/wp-content/uploads/2015/07/cyp-eating-disorders-access-waiting-time-standard-comm-guid.pdf

Vol. 2023 (2) 69 – 77
ISSN 2688-3651 [online]

Parker, L. L., & Harriger, J. A. (2020). Eating disorders and disordered eating behaviors in the LGBT population: A review of the literature. Journal of Eating Disorders, 8(1). https://doi.org/10.1186/s40337-020-00327-y

Qian, J., et al. (2013, August). Prevalence of eating disorders in the general population: A systematic review. Shanghai archives of psychiatry. Retrieved January 3, 2023, from https://pubmed.ncbi.nlm.nih.gov/24991159/

Raveneau, G., et al. (2014). Attitudes and knowledge levels of nurses and residents caring for adolescents with an eating disorder. International Journal of Adolescent Medicine and Health, 26(1), 131–136. https://doi.org/10.1515/ijamh-2013-0015

Rouleau, C. R., & von Ranson, K. M. (2011). Potential risks of pro-eating disorder websites. Clinical Psychology Review, 31(4), 525–531. https://doi.org/10.1016/j.cpr.2010.12.005

Sangha, S., et al. (2019). Eating disorders in males: How primary care providers can improve recognition, diagnosis, and treatment. American Journal of Men's Health, 13(3), 155798831985742. https://doi.org/10.1177/1557988319857424

Sonneville, K. R., & Lipson, S. K. (2018). Disparities in eating disorder diagnosis and treatment according to weight status, race/ethnicity, socioeconomic background, and sex among college students. International Journal of Eating Disorders, 51(6), 518–526. https://doi.org/10.1002/eat.22846

Strand, M., et al. (2020). Identifying research priorities for the study of atypical anorexia nervosa: A delphi study. International Journal of Eating Disorders, 53(10), 1729–1738. https://doi.org/10.1002/eat.23358

Thompson, C., & Park, S. (2016). Barriers to access and utilization of eating disorder treatment among women. Archives of Women's Mental Health, 19(5), 753–760. https://doi.org/10.1007/s00737-016-0618-4

US Preventive Services Task Force. Screening for Eating Disorders in Adolescents and Adults: US Preventive Services Task Force Recommendation Statement. JAMA. 2022;327(11):1061–1067. doi:10.1001/jama.2022.1806

van Eeden, A. E., van Hoeken, D., & Hoek, H. W. (2021). Incidence, prevalence and mortality of anorexia nervosa and bulimia nervosa. Current Opinion in Psychiatry, 34(6), 515–524. https://doi.org/10.1097/yco.0000000000000739

Waller, G., et al. (2009). Problems across care pathways in specialist Adult eating disorder services. Psychiatric Bulletin, 33(1), 26–29. https://doi.org/10.1192/pb.bp.107.018325

Webb, H., & Schmidt, U. (2020). Facilitators and barriers to supporting young people with eating disorders during their transition to, and time at, university: An exploration of clinicians' perspectives. European Eating Disorders Review, 29(3), 443–457. https://doi.org/10.1002/erv.2795

Whitney, J., et al. (2007). Caring for people with eating disorders: Factors associated with psychological distress and negative caregiving appraisals in carers of people with eating disorders. British Journal of Clinical Psychology, 46(4), 413–428. https://doi.org/10.1348/014466507x173781

Journal of Research High School

Vol. 2023 (2) 78 – 83
ISSN 2688-3651 [online]

Stem Cell Therapies for Lung Cancer Patients with a History of Tobacco Use

Ruhika Muralidhar[1] *

[1]Dublin High School, Dublin, CA, USA
*Corresponding Author: ruhika24@gmail.com

Advisor: Joseph Bruns, joseph.bruns93@gmail.com

Received November 27, 2022; Revised June 28, 2023; Accepted, July 19, 2023

Abstract

Lung cancer is the most commonly diagnosed cancer in both men and women. There are two main types of lung cancer, small cell carcinoma, and non-small cell carcinoma. Non-small cell carcinoma is more common and less aggressive than small cell carcinoma. Lung cancer is more commonly diagnosed in its later stages, making it harder to treat in the long run. The leading cause of small cell carcinoma is smoking and using tobacco products. The current treatments generally do not have a high success rate, specifically for patients with small cell lung cancer. A new treatment method that can change the course of cancer treatment in the future: stem cell therapy. Stem cell therapy can repair the lungs of small cell carcinoma patients and provide them with the needed cells for their lungs to return to normal function. The stem cells would be removed from the skin or blood of the patient and genetically transformed into embryonic stem cells, which provide more utility for the stem cells in the body. Once they are modified, these stem cells can be transferred to the body through in vitro scaffolds, tissue-engineered scaffolds, and local injections. After the transfer has occurred, the stem cells would be able to rebuild the lungs and help the body return to normal conditions.

Keywords: Lungs, Cancer, Stem Cells, Tobacco, Smoking

1. Introduction/Background

Lung cancer is caused by a malignant tumor that is formed in the bronchi, bronchioles, or alveolus of the lung. Many lung cancer patients experience symptoms in the later stages of the disease's development and may experience perpetual coughing with blood and phlegm along with a lack of energy and constant fatigue. For over 70 years, lung cancer has been the most common cancer between both women and men, with breast cancer being the most frequently diagnosed cancer among women and prostate cancer being common among men. Once diagnosed with this cancer, only 40% of patients survive for a year or more and only 15% are able to survive for 5 years or more (American Cancer Society, 2022). This is because lung cancer is more commonly diagnosed in the later stages of development, making it harder to treat. In 2022, there are estimated to be 236,740 more cases of lung cancer and 130,180 deaths from lung cancer in the United States. As shown in Figure 1, lung cancer makes up 25% of all cancer deaths, being higher than the breast, colon, and prostate cancer deaths combined. Lung cancer is more frequently diagnosed in patients that are 65 years old and older. (American Cancer Society, 2022).

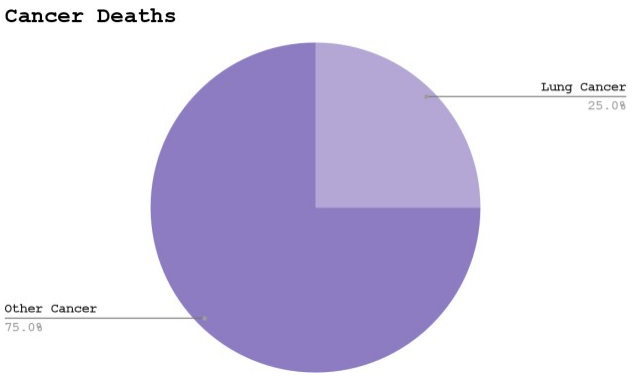


Figure 1. This image models the percentage of cancer deaths that are associated with lung cancer compared to other cancers.

Vol. 2023 (2) 78 – 83
ISSN 2688-3651 [online]

1.1 Types of Lung Cancer

There are two main types of lung cancer, small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). Small cell lung cancer, or small cell carcinoma, begins in the bronchi, where it rapidly grows and spreads throughout the entire body. On the other hand, non-small cell lung cancer is usually less aggressive than small cell carcinoma and is divided into three main groups, adenocarcinoma, squamous cell carcinoma, and large cell carcinoma (Spiro and Silvestri, 2005). All three types of non- small cell lung cancer spread moderately compared to small cell lung cancer and occur in different parts of the lung. While non-small cell carcinoma can occur to anyone, small cell carcinoma is caused from smoking and is in more need of a cure due to its aggressive nature. This review will focus on the treatment of small cell carcinoma, as it is directly connected with smoking and has a high mortality rate.

1.2 Smoking and its Impacts

Smoking is known to be the main cause of small cell carcinoma because of the many carcinogens, or cancer-causing agents, located inside a single cigarette. There is a total of 7,000 chemicals found in one cigarette, with 70 of these chemicals known to be carcinogens. Examples of such carcinogens are benzene, aldehydes, ethylene oxide, aromatic amines, and N-nitrosamines. When these chemicals are inhaled into the body and transported through the lungs, they can damage the epithelial cells located in the lungs and lead to cell mutations that form into cancerous cells. These cancerous cells can mutate and form tumors that lead to small cell carcinoma.

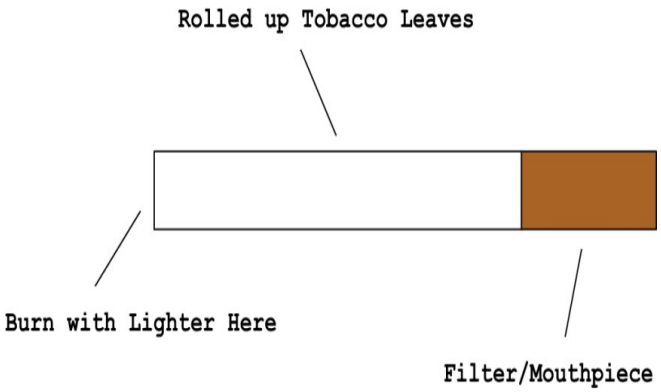


Figure 2. This is a model of the specific parts of a cigarette. It showcases how a person would use a cigarette and where the tobacco is located

There are multiple devices that an individual can use to smoke and consume tobacco. The most common throughout history has been cigarettes. When a cigarette is lit, the tobacco is burned and travels through the cigarette, where the tobacco leaves are filtered, and the smoke enters the mouth of the consumer. Figure 2 shows the specific parts of a cigarette. E-cigarettes and vapes have been more commonly used today. E-cigarettes contain a battery-powered heating device that vaporizes the chemicals in the device for aerosol to be produced. This aerosol is then inhaled by the smoker.

1.3 Current Treatments for Cancer

Currently, there are three main treatments for cancer, surgery, radiation therapy, and chemotherapy. Surgery removes the tumor from the affected location to prevent cancer from spreading. This has its drawbacks, as it is only beneficial for patients with tumors close to the skin and with cancer that has not likely spread. There is also a high mortality rate for the surgery and the median survival rate post-surgery is only around 3 years. Radiation therapy uses extreme amounts of radiation to target and kills the cancer cells to shrink the tumor. The radiation targets the DNA of the cancer cells and alters the sequence, which leads to apoptosis. Radiation can be harmful to the body if exposed for too long, which can lead to hair loss, problems in the heart and brain, and trouble with speech and memory. Similar to radiation therapy, chemotherapy uses specific anti-cancer chemicals that can kill cancer cells to diminish the tumor.

Chemotherapy also has a low success rate of 16% when used by itself, but does have a higher success rate of 35% when paired with radiation therapy. The current treatments for cancer are not sustainable for patients with small cell carcinoma as they only extend their lives by 1-4 years and leave the patients with many side effects from the treatments.

Vol. 2023 (2) 78 – 83
ISSN 2688-3651 [online]

1.4 What is Stem Cell Therapy?

There is a possible treatment method that can potentially change treatment methods for small cell carcinoma. This method is stem cell therapy. Stem cells are unspecialized cells located in specific parts of the body. The stem cells can then be used to build tissues in the body and maintain needed functions for survival. There are two main types of stem cells, embryonic stem cells and adult stem cells. Embryonic stem cells are the stem cells during the developmental processes of an embryo. They can become specialized into any cells for function in a newborn baby (Lo and Parham, 2009). On the other hand, adult stem cells have specific functions they will end up performing, meaning they will become specialized into cells based on their location. For instance, if a stem cell was located in the brain, it would only be able to become a brain cell. Stem cell therapy, also known as regenerative medicine, uses the idea of stem cells to replace cancerous cells with functioning stem cells to rebuild the organ and treat the patient (Lo and Parham, 2009). This way of treatment would be beneficial as it provides patients with the needed cells to rid the body of cancer as quickly as possible. It also allows the body to rebuild the damaged organ and get back its original structure and functions.

There was a main ethical concern regarding the use of stem cell therapy. Because it was hard to obtain specific stem cells for different kinds of cancer, scientists would use the stem cells from unused embryos in order to find stem cells that could be used for any part of the body. Many people were concerned with this way of extraction and with using embryos from in vitro fertilization (Lo and Parham, 2009). Luckily, in 2012, this problem was solved when Shinya Yamanaka and John Gurdon were awarded the Nobel Peace Prize for their influential discovery in this matter. Yamanaka and Gurdon were able to reprogram adult stem cells into embryonic stem cells using important genes in the cells (Gurdon and Yamanaka, 2019). This discovery was extremely important because scientists are now able to extract stem cells from the blood of the patient and reprogram the blood stem cells into embryonic stem cells, where they can be used for various treatments for any of their disorders.

1.5 Purpose

The hypothesis made is that stem cell therapy can be used as a potential cure for small cell carcinoma patients with a history of tobacco use. The purpose of this review is to prove whether or not this theory is true. Small cell carcinoma is largely associated with smoking, but with an increase in the variety of tobacco products, more people are developing addictions to cigarettes. Because of this, there is no known method to prevent people from smoking, and as small cell carcinoma cases arise, finding the best treatment possible will help reduce the deaths that are caused by smoking. This review showcases a treatment method that has been proven to be successful for other diseases, and developments will be made on whether or not scientists should experiment with stem cell therapy as a treatment method.

2. Stem Cell Therapies

There are four types of stem cell therapies, hematopoietic stem cells (HSCs), skin stem cells (SSCs), neural stem cells (NSCs), and mesenchymal stem cells (MSCs). HSCs are stem cells that are able to develop into any kind of blood cells, such as white blood cells, red blood cells, and platelets. They are most commonly found in the blood and bone marrow. SSCs are adult stem cells that are present in the skin. They are able to differentiate into multiple cell types of the skin and can be used for skin renewal and repair. NSCs are multipotent cells that form the radial glial progenitor cells. These cells create neurons and glia that help the nervous system function during the development of the embryo. Finally, MSCs are stem cells that are found in the bone marrow. They are important for the maintenance of the skeletal system, such as bone, fat, and cartilage.

The most common kinds of stem cell therapies are the use of HSCs and MSCs. For HSCs, the blood is removed from the body, the stem cells are extracted from the blood sample, and the blood is returned back to the body for normal function. In order to stimulate the growth of more stem cells, many patients take medications, such as filgrastim, to increase the bone marrow stem cell count days before the extraction is completed. Stem cell treatment

Vol. 2023 (2) 78 – 83
ISSN 2688-3651 [online]

is also commonly used with chemotherapy in order to remove cancerous cells from the body while providing the organ with new and healthy stem cells. Once the stem cells are retrieved from the patient, the adult stem cells have to be reprogrammed into embryonic stem cells in order to be able to be beneficial for the treatment of the patient. In order to do so, scientists most commonly inject the stem cells with viruses that contain four specific genes, SOX2, Oct4, KLF4, and cMyc. These genes modify the stem cells in order for them to obtain the characteristics of an embryonic stem cell without raising any ethical concerns. Finally, the stem cells are transferred back to the body using three common delivery methods: local injections, in vitro scaffolds, and tissue-engineered scaffolds. Many studies have been conducted in order to test the effectiveness of each kind of delivery method on patients with different disorders.

Local injections are usually conducted in a region close to the affected area of the patient. The treatment process is commonly split into three injection rounds in a span of 2-5 days. The first injection is a solution of simple sugars called dextrose that prepares the body for the stem cells to enter. The second injection is done with the extracted stem cells for them to travel to the region that needs to be treated. The final injection is platelets that will allow the stem cells to function in the body. Multiple studies have been conducted to show the effectiveness of local injection stem cell therapy.

In one study, done by Saieh Hajighasemlou et al., scientists conducted stem cell treatment through local injections on patients with hepatocellular carcinoma (HCC), the most common form of liver cancer. The cells derived from the bone marrow (MSCs) are used for treatment. Hajighasemlou et al. conducted this treatment on 18 nude mice with HCC and split them into 6 groups, chemotherapy, stem cells through intravenous therapy (IV), stem cells through local injection, chemotherapy and stem cells through IV, chemotherapy and stem cells through IV, and no treatment. The results were able to showcase that using a local injection rather than an IV worked better for the treatment of nude mice. The tumor was also able to be diminished with a combination of chemotherapy and stem cell treatments using the mesenchymal cells under a local injection. This is because the chemotherapy treatment kills the cancerous cells and leaves more room for the stem cells to help treat the patient.

Another study, conducted by Jignesh Dalal et al., discusses the benefits of mesenchymal stem cells on patients with Crohn's disease. Crohn's disease is a type of inflammatory bowel disease that occurs in the lining of the digestive tract. It commonly causes abdominal pain, diarrhea, fatigue, and weight loss. Patients with Crohn's disease also may experience perianal fistula, an infected tunnel between the skin and anus, or enterocutaneous fistula, an infected tunnel between the skin and stomach. Dalal et al. were able to explain how mesenchymal stem cells are able to be used as a treatment for Crohn's disease and suppress the rapid growth of the affected cells using local injections. In order to test this theory, they derived MSCs from 9 patients with perianal fistulas and 1 patient with enterocutaneous fistulas. The injections were done every 4 weeks until there was a response in the body. The results showcased a benefit in the fistulizing lesions with the use of local injections as the cells were shown to have had a major contribution to the repair process of the lesions.

Because of the shown effectiveness of locally injected stem cells for patients with hepatocellular carcinoma and Crohn's disease, this method of treatment can also be a beneficial treatment for patients with small cell carcinoma. The lungs, intestines, and liver all play extremely crucial roles in the human body and need proper treatments for them to function well. The human body would not be able to survive without these vital organs. If stem cells have advantages for the treatment of liver cancer and the lesions formed by the fistulas, they can play a large role in the treatment of small cell carcinoma and other cancers in the future. More research and experiments can be done to prove this theory.

Another method of stem cell treatment is scaffolding. A scaffold is a medical technique in which a framework has been produced in order to create an environment that helps with the formation of desired tissue for patients with degenerative diseases. The new tissue can then be introduced back into the patient to help the region that was impacted. In order to do so, stem cells (programmed to become lung cells) are placed on the scaffolds to be guided in their growth of new tissue. There are two types of scaffolding techniques, in vitro scaffolds and tissue-engineered scaffolds. In vitro scaffolds are done in Petri dishes to help initiate the growth of tissue. Tissue-engineered scaffolds use materials such as collagen and polylactic acid to create an environment for new tissue to grow. Many studies have been conducted in order to test the effectiveness of in vitro scaffolds on patients with tissue degenerative

Vol. 2023 (2) 78 – 83
ISSN 2688-3651 [online]

diseases.

One study, done by Andrew W. Holle et al., explains the general pattern of growth that has been experienced in the treatment of cancers in the body due to in vitro models. Holle et al. discuss specific drugs located in the cells that are more efficient in in vitro scaffolding methods, making their research more accurate and reliable. He also explains that since new methods for culturing cancer cells in scaffolds have been created, there are more areas in the future that can use this testing to their advantage. The researchers were able to conclude that the technique of scaffolding can be significant in the future treatment of cancer as they were able to help detect and reduce cancer cell tumors created in the body.

The use of scaffolding can also help with the growth of drug discoveries in this field of work to expand treatment methods to more than chemotherapy and radiation therapy.

Another study, conducted by Yongxiang Luo et al., analyzes bifunctional scaffolds as a potential treatment for breast cancer. The researchers decided to perform this experiment because scaffolding showcased similar results to the function and look of normal breast tissue, as it was very flexible and had similar behaviors to breast tissue. They build a scaffold using synthetic materials to create better results for treatment. The scaffold was able to show positive effects for cancer therapy and did match the properties of the tissue very well. The scaffold also was able to continue mitosis in the breast tissue to help the tissue maintain its health and overall well-being. All in all, the scaffold was shown to be a great option for the treatment of breast cancer and for the repairing of damaged tissues in the body.

The general methods that have been created for scaffolding show major advantages that can be used to determine how it can treat small cell carcinoma. Using the scaffolding technique with stem cells can also be used to help find more treatments for small cell carcinoma using different types of stem cells and different drugs. Scaffolds also have the ability to act similarly to normal breast tissue, making them beneficial for the treatment of breast cancer. The lung tissue also has to be flexible in order to contract and relax during inspiration and expiration. This means that the scaffolding technique can also be used for treating small cell carcinoma and can help continue the growth of the tissue to maintain a healthy lung.

More research is currently being conducted to test the effect of scaffolding on patients with more invasive forms of cancer.

3. Conclusion

The first and most beneficial way to avoid developing small cell carcinoma is to quit or avoid smoking entirely. By doing so, the lungs won't be easily damaged and have as high of a chance of developing any sort of lung cancer. It also avoids the possibility of developing any other kinds of lung diseases, such as emphysema, damaged air sacs, and chronic obstructive pulmonary disease (COPD), the blockage of the airways. However, with the increase in tobacco use across the country, it is becoming harder to stop people from smoking and increasing their risk for developing lung cancer. Finding the most beneficial treatment method will help limit the number of deaths associated with smoking and small cell carcinoma.

This review helped prove that stem cell therapy can extremely beneficial in the treatment of this disease. Stem cells can be used to initiate the growth of new tissue and also quicken the recovery process of small cell carcinoma patients. When paired with chemotherapy, the chemicals from the chemotherapy can help diminish the tumor and prevent the growth of cancerous cells while stem cell therapy can help initiate the development of healthy cells to quicken the healing process. Future research can be done to test this theory and discover the true potential of stem cell therapy for small cell carcinoma with a history of tobacco use.

References

Anton, B. (2005). "Stem Cells for Lung Cancer? Cell Press, 121(6), 811-813. https://www.sciencedirect.com/science/article/pii/S0092867405005477

Dalal, J., Gandy, K., & Domen, Jos. (2012). Role of mesenchymal stem cell therapy in Crohn's disease. Pediatric Research, 71, 445–451. https://www.nature.com/articles/pr201156

Vol. 2023 (2) 78 – 83
ISSN 2688-3651 [online]

Hajighasemlou, S., et. al. (2020). Sorafenib and Mesenchymal Stem Cell Therapy: A Promising Approach for Treatment of HCC. Evidence-Based Complementary and Alternative Medicine, 2020. https://www.hindawi.com/journals/ecam/2020/9602728/

Holle, A. W., Young, J. L., & Spatz, J. P. (2016). In vitro cancer cell–ECM interactions inform in vivo cancer treatment. Advanced Drug Delivery Reviews, 97, 270-279. https://www.sciencedirect.com/science/article/pii/S0169409X1500232X

How Induced Pluripotency Changed Stem Cell Science. (2019, March 7). A Closer Look at Stem Cells. Retrieved November 17, 2022, from https://www.closerlookatstemcells.org/2019/03/07/how-induced- pluripotency-changed-stem-cell- science/#:~:text=An%20important%20achievement%20in%20stem,Shinya%20Yamanaka%20and%20John%20Gurdon

Key Statistics for Lung Cancer. (2022, February 14). American Cancer Society. Retrieved November 16, 2022, from https://www.cancer.org/cancer/lung-cancer/about/key-statistics.html
Lo, B., & Parham, L. (2009). Ethical Issues in Stem Cell Research. Endocrine Reviews, 30(3), 204-213. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2726839/

Luo, Y., et. al. (2019). 3D printing of hydrogel scaffolds for future application in photothermal therapy of breast cancer and tissue repair. Acta Biomaterialia, 92, 37-47. https://www.sciencedirect.com/science/article/abs/pii/S1742706119303666

Spiro, S. G., & Silvestri, G. A. (2005). One Hundred Years of Lung Cancer. American Journal of Respiratory and Critical Care Medicine, 172(5). https://www.atsjournals.org/doi/full/10.1164/rccm.200504-531OE

Journal of Research High School

Vol. 2023 (2) 84 – 90
ISSN 2688-3651 [online]

Measuring the Effect of the Pandemic on Music Practicing Time and Motivation

Darian Liu[1*]

[1]Lynbrook High School, San Jose, CA, USA
*Corresponding Author: darianliu2014@gmail.com

Advisor: Michael Pakaluk, michael_pakaluk@fuhsd.org

Received September 1, 2022; Revised June 1, 2023; Accepted, June 28, 2023

Abstract

The COVID-19 pandemic had a large impact on the world as a whole. The pandemic had a particular impact on musicians, replacing live performances and in-person lessons with recorded performances and remote lessons during lockdown. Disruptions and uncertainty caused by the pandemic have created great change in the lives and schedules of music students, which has been observed by several recent studies. Such studies focus on adult classical musicians, but this study aims to observe school-age music students. In the current study, American music students filled out a survey measuring demographic data and descriptive data about practice time, feelings of threat, and feelings of uncertainty. I hypothesized that practice time would be affected by the pandemic, as well as feelings of motivation, threat, and uncertainty. I found that none of the above held a strong correlation with practice time. However, I also found a significant relationship ($p<0.05$) between years played and motivation to practice. Changes in lesson format positively affected student motivation and practice times, and overall motivation and effectiveness of lessons were positively affected by years of musical experience.

Keywords: M*usic students, Lesson format, Pandemic, Practicing time, Effectiveness, External motivation, Internal motivation.*

1. Introduction

The 2019 coronavirus disease (COVID-19) pandemic caused significant stress and uncertainty among everyone. Students were especially affected by the pandemic, as lockdown procedures caused school closures and restricted in-person gatherings. Students had to adapt to the new teaching structure of asynchronous lessons and material, while also developing skills to maintain and manage new schedules. Music students were even more so affected by such severe disruptions, with live performances and in-person lessons being important parts of their curriculum.

This research project responds to the gap in the literature regarding classical music students over the pandemic and seeks to further examine the impact of the pandemic on musicians. My research question asks whether the COVID-19 pandemic increased or decreased the length of music students' practice times. I hypothesized that the COVID-19 pandemic decreased the length of music students' practice times for students who had increased anxiety or decreased motivation over the pandemic. My second research question was whether changes in motivation affected the amount of practice time. I hypothesized that changes in motivation will decrease the amount of practice time. Lastly, my third research question was whether feelings of threat and uncertainty affect the amount of practice time. I hypothesized that an increase in feelings of threat and uncertainty will result in a decrease in practice time. Predicted findings would be that practice time and motivation would go down during the pandemic. Results could help us determine whether the pandemic had a positive, negative, or neutral effect on the musician population. By assessing the pandemic's effect on the musician population, researchers can understand what factors impacted the musicians the most and maximize musical productivity.

Vol. 2023 (2) 84 – 90
ISSN 2688-3651 [online]

2. Literature Review

Recent literature suggests that students have been heavily impacted by the pandemic. Garcia and Weiss (2020) found that the pandemic affected teaching and learning due to lacking the needed requirements for effectiveness. With the general student population in such a state, it is prudent to analyze how other demographics were affected by the pandemic. By observing how certain demographics were affected by the pandemic, researchers can narrow down which aspects of the pandemic were most impactful. Artists, for instance, were negatively impacted by the pandemic in particular ways. In 2020, many artists were economically displaced due to the pandemic and experienced a large decline in revenue (Guibert and Hyde, 2021).

Fram, N. R., Goudarzi, V., et al. (2021) utilized an internet survey on musical practice and collaboration and received responses from US musicians. They found that musicians are resilient to extrinsic shocks, as musical practice and correlation numbers did not see a large differential over the pandemic.

Studies that examined music students also highlighted the pervasive impact of COVID-19. Rossen, Baumann, and Altenmuller (2021) surveyed 75 German university students in 2019 and 2020, and found that 76% of students indicated a change in practicing hours during the pandemic: 31% of students reported an increase in daily practicing hours, while 45% reported a decrease. This may show that there are other factors other than the pandemic that influences practice habits, and also that not all musicians were affected in the same way by the pandemic.

Other researchers, such as López-Íñiguez, McPherson, et al. (2022), measured the effects of internal and external motivation on professional classical musicians before, during, and after the pandemic. Internal motivation was classified as motivation that comes from internal thoughts and feelings (e.g. “I want to practice because it is important to me”). In comparison, external motivation is motivation that comes from outside factors (e.g. “I practice because my parents want me to”) They found that externally motivated musicians were more affected by the pandemic, while self-motivated students were less affected. They also measured and found an inverse correlation between threat and self-motivation, with increased feelings of threat resulting in a decrease in self-motivation. Feelings of threat referred to the amount that the participants felt they were threatened by the pandemic and COVID-19.

Nusseck and Spahn (2021) measured the self-motivation of musicians as well, measuring German university music students. They collected data on the practice environment, physical activity (exercising and sports), and practice time of the students. They found that while the mean practice time between before and during the pandemic stayed the same, there were notable differences in the practice times of music performance and music education majors: music performance majors practiced less over the pandemic, while music education majors practiced more than before the pandemic.

Howard, Green, et al. (2021), who studied young musicians from Australia, England, and Portugal, found similar results, noting an increase in practice time and music production over the pandemic. They concluded this was due to self-motivated young musicians being able to turn inward and find more time to practice.

Overall, recent research indicates that whether positive or negative, the impacts of the COVID-19 pandemic were significant for musicians. This study seeks to explore this assumption in an unexamined population: younger music students.

3. Materials and Methods

3.1 Participants

The study actively recruited participants from the San Francisco Conservatory of Music and Lynbrook High School, as well as other students from various states, though to a much lesser degree. The participants were largely from these two areas as they were most receptive to the author’s recruitment. There were 85 participants in total, of which 35 were male, 47 were female, 1 was non-binary, and 2 preferred not to state their gender. Out of the participants, 3 played percussion, 31 played keyboard instruments, 25 played string instruments, and 26 played wind instruments. Participants were divided into three age groups:1-15 years old (N = 35), 16-21 years old (N = 42), and 22-66 years old (N = 8).

3.2 Design and Materials

Vol. 2023 (2) 84 – 90
ISSN 2688-3651 [online]

This study used an ex-post facto design. Due to the pandemic's effects being lessened by the time of the study, an ex-post facto design was ideal for observing potential relationships and correlations during the time of the pandemic. A survey was designed to investigate the potential relationships and correlations between various factors (e.g. remote or in-person lessons) and measure overall practice time for classical music students and professors in the United States during the COVID-19 pandemic. The pandemic was divided into three periods: before January 30, 2020, when WHO declared COVID-19 a pandemic, January 30, 2020 to June 15, 2021, when lockdown restrictions were lifted in California, and June 15, 2021 to the present. The data was collected after the three periods of the pandemic to ensure all the periods were taken into account for comparison. The survey measured feelings of threat, uncertainty, and motivation of musicians to practice over the three periods of the pandemic. Additionally, surveys were chosen as the primary source of data collection due to convenience of making a survey and the ease of gathering responses to a relatively short survey compared to the time required for a full-fledged experiment. In the survey, feelings of threat and uncertainty were measured by how much the participants felt they were threatened by the pandemic situation, rating agreeability on statements such as "I felt threatened by a possible viral infection and disease of my loved ones." Motivation was measured using similarly styled survey questions to determine the participants' motivation for practice and whether they had external or internal motivation. The threat and motivation scores were calculated by averaging all appropriate item responses.

The survey was formatted with seven-point Likert scales for each question. The seven-point Likert scale system would allow for a numerical measure of the measuring of abstract concepts, which makes statistical testing more feasible. The questions measured demographics, self and external motivation for practice, and descriptive data about changes in practice time. Levels of threat and uncertainty were measured using a COVID-19 adaptation by Vermote et al. (2021) of the original *Feelings of Uncertainty and Threat Scale* from Chen et al. (2015). Motivation was measured using the *Multidimensional Motivation at Work Scale* by Gagne et al. (2015). The two scales are different since they were sourced from existing research papers. As the scales were created for the express purpose of measuring each variable, using the different scales should provide clearer results.

3.3 Data Analysis

In the data analysis process, frequency charts were first used to determine the overall spread of the data, ensuring that there were not any egregious outliers. The data was then tested through t-tests, assessed through correlational tests, and inserted into regression models. Significance was determined by examining the p-values and R-squared values found in the data and assessing if they were under 0.05.

4. Results

4.1 Changes in Practice Time Before, During, and After the Pandemic

To answer if the pandemic had an effect on practice times, the data for each time period was first averaged and compared for any trends. I found that there were not any strong trends regarding practice time before, during, and after the pandemic, as all three were relatively similar (Figure 1).

The following percentages, r-values, and p-values were all found through the input of data in Microsoft Excel. I analyzed the practice times of different instrument types to ensure that this trend applied to all instrument types. As different instrument types likely had different experiences during the pandemic (ie. wind instruments being unable to play in public due to mask restrictions), thus, it is important to observe the results by instrument type to ensure

Vol. 2023 (2) 84 – 90
ISSN 2688-3651 [online]

my data is applicable to musicians as a whole. There was a fairly even distribution of instruments played by the participants of the study (3.5% percussion, 36.5% keyboard, 29.4% strings, and 30.6% winds). Because the instrument type was approximately equally distributed, we were able to attain data that was more applicable to musicians of all instrument types. Using the demographic data attained, we were able to sort practice times by instrument (Figure 2). Trends in the data showed that winds tended to practice the least, while pianists practiced the most. There was not a strong trend found between practice times before, during, and after the pandemic, but it can be noted that percussionists experienced gradual increase in practice hours from pre- to post- COVID, while the inverse was the case for wind instrumentalists, who experienced a gradual decrease.

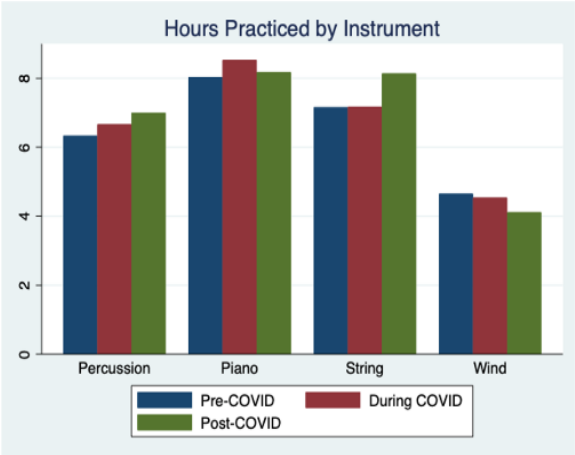


Figure 2. Mean Number of Hours Practiced Before, During, and After the Pandemic Across Different Instrument Groups

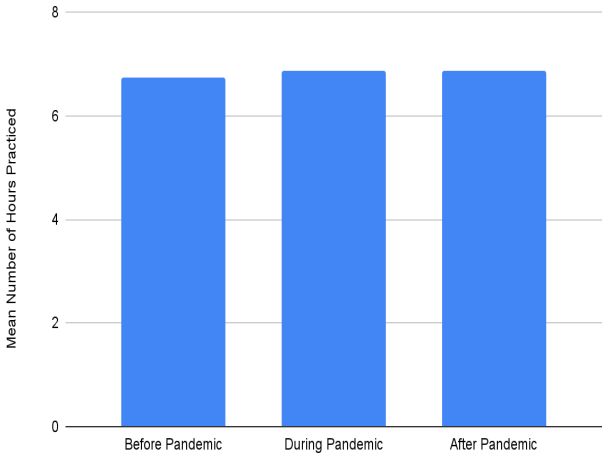


Figure 1. Mean Number of Hours Practiced Before, During, and After the Pandemic

String instrumentalists stayed the same between pre- and during COVID, but times by instrument experienced an increase after COVID, while keyboardists had a dramatic increase during COVID, but afterward receded to similar numbers as before COVID.

4.2 Effects of Motivation on Practice Time During the Pandemic

I found that there were mixed results regarding motivation in my study. Although 47.1% of students saw a decrease in motivation, a slight majority reported no change or an increase in motivation, with 31.8% reporting no decrease in motivation and 21.2% reporting a possible decrease in motivation. Additionally, motivation was weakly positively correlated with mean hours of practice during the pandemic, but the correlation was not significant ($r = 0.193$, $p = 0.077$).

4.3 Effects of Feelings of Threat and Uncertainty During the Pandemic

On average, participants had a mean threat score of 4.3, indicating that there was not a strong sense of threat felt during the pandemic. This is supported by the weak correlation between the mean hours of practice during the pandemic, and the correlation was not significant as well ($r = 0.125$, $p = 0.256$).

4.4 Further Investigation

We conducted additional analysis based on the results we received in the survey. From the 85 participants, 47 were female, 35 were male, and 3 were otherwise identified. Descriptive analysis and subsequent t-tests found no significant difference in the number of instruments based on gender. Survey questions were answered by those who took lessons under the asked formats. Descriptive analysis indicated that in-person lessons were widely regarded as more effective than remote lessons. On a scale of 1 to 5, in-person lessons were given a mean score of 4.695 compared to remote lessons' mean score of 2.725 (Figure 3).

Vol. 2023 (2) 84 – 90
ISSN 2688-3651 [online]

A t-test allowed us to find a negative correlation between years and motivation, in which participants that have more years of musical experience experienced lower motivation during the pandemic (r = -0.218, p=0.045). Lesson format was classified in two categories: remote and in-person. Remote lessons are virtual lessons held through services such as Zoom or Google Meets, while in-person lessons are lessons held in a physical setting where both the instructor and the student are in close proximity. A positive correlation was found between lesson format during COVID and feelings of threat and uncertainty (r=0.2206, p=0.0424).

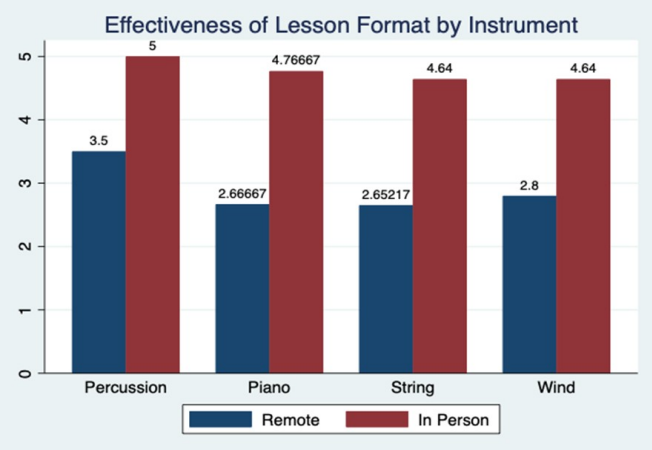


Figure 3. Effectiveness of Lesson Format by Instrument Group

Similarly, p-values of less than 0.05 made up every test regarding hours of practice and effectiveness of lesson formats for both remote and in-person during all time frames, allowing us to determine that hours of practice and effectiveness of lesson formats are correlated. This correlational model also supports this data, as the p-values found are less than 0.05, meaning more hours of practice result in a higher rating of lesson effectiveness.

Age and gender, on the other hand, had a lesser effect on the independent variable of practice time (p=0.069). Thus, we fail to reject that age and gender do not have an effect on practice time. When examining regression models, measures of threat and uncertainty also did not have a strong effect on results (p=0.473, 0.582).

5. Discussion

In this study, 85 classical musicians were surveyed about their practice times, lesson format, and feelings of threat and uncertainty amid a global pandemic. While no significant trends were found in terms of practice hours before, during, and after COVID, it is clear that musicians tend to favor in-person lessons over remote ones. The lack of discernible pattern in practice hours was also shown by Nusseck and Spahn (2021) and López-Íñiguez, McPherson, and Alzugaray (2022). Both found that when looking at the average of all the practice times submitted, there was not a large increase or decrease found in overall practice time over the pandemic from university music students. Thus, my findings regarding practice time also contradict the findings of Howard, Green, et al. (2021) who posit that practice times during the pandemic increased due to musicians being able to turn inward and find self-motivation to practice. This discrepancy could occur because of the difference in participant demographics, as this paper researched musicians of all ages while Howard, Green, et al. (2021) conducted research exclusively on professional musicians who would have more motivation because their profession is in music.

Feelings of threat were found to not be strongly correlated to any of the variables. For motivation, however, a negative relationship was found between years of playing and motivation, as musicians with more years of experience reported less motivation overall. This was one factor no other research paper has taken into account thus far, as surveys did not ask for the participants' previous music experience. Another factor previously unaddressed was the perceived effectiveness of lesson formats. Doing so, we found that in-person lessons were significantly rated as more effective than remote lessons across instrument types and age groups – when rated on a scale of one to five, the mean of in-person lessons was 4.695 while the mean of remote lessons was 2.725.

5.1 Limitations

Several limitations to this project must be noted. First, this study utilized a voluntary response sample, which creates potential for nonresponse bias or risk of an unrepresentative sample. With only willing musicians answering the survey, perhaps the unrepresented musicians would have provided different answers. The responses were also self-

Vol. 2023 (2) 84 – 90
ISSN 2688-3651 [online]

reported, so many factors such as effectiveness of lesson format may not be accurate due to it being based on human opinion. Additionally, although data was not collected from a single school, location, and instrumentalist, the overall sample size was fairly small, which engenders a lack of generalizability, as most responses are concentrated within a certain demographic, whether it be age or location. This limits the applicability of the findings to a larger population. The population was also largely split between two different demographics – Lynbrook High School and the San Francisco Conservatory of Music. Due to differences in dedication towards music, as SFCM students are likely more devoted and spend more time practicing, my data may be less applicable to a wider population.

5.2 Future research

Future research should seek to engage a larger sample size and attain data on these variables from a wider, more varied population. The findings indicate there is value in probing more specific questions regarding motivation to practice and practice hours, since my survey data did not demonstrate a correlation between those two factors; however, more specific questions designed to determine if the two are related might yield a more definitive answer. Future studies might be able to look more deeply into the variables of musical experience and lesson format, as my paper was the only one to report data related to effectiveness of lesson format and the impact of musical experience on results. With more questions or data related to these variables, greater understanding of the importance of them can be determined. Additional relevant variables that could be explored include sources of external and internal motivation to practice, and demographic variables such as geographic location, level of education, and annual family income.

6. Conclusion

This research project was conceived to address the gap in extant literature on the impact of the COVID-19 pandemic on musicians, which focused on professional or college musicians. In contrast, the sample size of this project primarily centers around students from middle to high school age, in order to target a demographic otherwise less explored. It explores the variables of practice time, motivation, and feelings of threat and uncertainty. The three initial research questions were answered throughout the investigation. The pandemic did not have a significant correlation on practice time. Moreover, feelings of motivation also did not have a significant correlation on practice time. Lastly, the participants' feelings of threat and uncertainty also did not have a significant correlation with practice time. Aside from the initial research questions, more of my survey data was analyzed to determine other correlations between variables. The findings suggest that in-person lessons are the most effective across all instrument types pre- and post-pandemic. Further, previous musical experience was positively correlated with level of motivation to practice throughout the pandemic. By providing data from an untested demographic, these results can be added to the existing body of research towards better conceptualizing the differences and similarities between child or teen musicians to adult musicians and the impact of the pandemic.

References

García, E., & Weiss, E. (2020). *Covid-19 and student performance, equity, and U.S. education policy: Lessons from pre-pandemic research to inform relief, recovery, and rebuilding*. Economic Policy Institute. https://www.epi.org/publication/the-consequences-of-the-covid-19-pandemic-for-education-performance-and-equity-in-the-united-states-what-can-we-learn-from-pre-pandemic-research-to-inform-relief-recovery-and-re building/

Fram, N. R., et al. (2021). *Collaborating in isolation: Assessing the effects of the COVID-19 pandemic on patterns of collaborative behavior among working musicians*. Frontiers in psychology, Vol. 12, https://doi.org/10.3389/fpsyg.2021.674246

Howard F., et al. (2021) *"It's Turned me From a Professional to a 'Bedroom DJ' Once Again": COVID-19 and New Forms of Inequality for Young Music-Makers.* Sage Journals, Vol. 29 https://doi.org/10.1177%2F1103308821998542

Vol. 2023 (2) 84 – 90
ISSN 2688-3651 [online]

Rosset, M., et al. (2021). *Studying music during the coronavirus pandemic: Conditions of studying and health-related challenges*. Frontiers in psychology, Vol. 12 https://doi.org/10.3389/fpsyg.2021.651393

Nusseck, M., & Spahn, C. (2021). *Musical practice in music students during COVID-19 lockdown*. Frontiers in psychology, Vol 12 https://doi.org/10.3389/fpsyg.2021.643177

Guibert, G., & Hyde, I. (2021). *Analysis: COVID-19's Impacts on Arts and Cultures*. Argonne. https://www.arts.gov/sites/default/files/COVID-Outlook-Week-of-1.4.2021.pdf

López-Íñiguez, G., et al. (2022). *Effects of Threat and Motivation on Classical Musicians' Professional Performance Practice During the COVID-19 Pandemic*. In Frontiers in Psychology (Vol. 13). https://doi.org/10.3389/fpsyg.2022.834666

Journal of Research High School

Vol. 2023 (2) 91 – 95
ISSN 2688-3651 [online]

The Importance of Social Class and Gender in Identity

Maia Getz,[1] * Conrad Warner[1], Glen Getz[2]

[1]North Allegheny Senior High School, Wexford, PA, USA
[2]Allegheny Health Network, Pittsburgh, PA, USA
*Corresponding Author: mgetz@nastudents.org

Advisor: Glen Getz (Ph.D.), glen.getz@ahn.org

Received February 10, 2023; Revised June 29, 2023; Accepted, July 21, 2023

Abstract

Identity plays a predominant role in today's society. According to the Oxford Dictionary, identity is "the fact of being who or what a person or thing is." Despite its rather simple definition, identity is extremely important in society because it drives political affiliation, school selection, occupational paths, and other important components of life. Social class and gender are two important aspects of identity. An individual's social class is correlated with political views, finances, education, and opinions about minorities and immigrants. Divisions created by gender also influence identity. Historically, women have often been discriminated against because of society's views of gender roles. Although there are multiple examples of women being treated differently than men, this is particularly prevalent in politics. This article examined how social class and gender can have a negative or positive impact on an individual and society in terms of various opportunities ranging from education, finances, and politics. It will be necessary for changes to continue to occur in social class and gender inequalities in order to improve upon an individual's identity, resulting in societal inclusion and evolution. We will discuss ideas to resolve these socioeconomic inequalities and the lack of political opportunities for women.

Keywords: Identity, Gender, Socioeconomic class, Education, Politics

1. Introduction

Identity is a concept of how people, individually and in a society, define themselves, and how others view them. There are many factors influencing identity, whether it is something as trivial as how we dress, to something as broad and complicated as social class. Throughout the 20th century and early in the 21st century, identity continues to impact individual decisions, influencing cultural outcomes. Social class and gender dictate people's self-identification, political views, and ability to change. Although there are several other important and well-studied factors, particularly race, that play a role relating to identity (Cole & Omari, 2003), this article will focus specifically on social class and gender. The objectives of this article are to highlight and explore identity when relating to social class's impact on education and views on others as well as to examine how gender plays a role in political opportunities, while also offering ideas to resolve this problem. Researching and linking these factors can assist with understanding disparities which could guide positive social change. We hypothesize that decreasing socioeconomic inequalities and increasing political opportunities for women would result in improved identity resulting in the betterment of society.

2. Social Class Relates to Identity

Social class has been demonstrated to be highly related to identity. First and foremost, social class is evident in nearly every society. Social class has been a prominent part of everyday life from some of the first societies in Mesopotamia and Egypt through the Roman Era and Dark Ages to our modern world. One historically known factor that impacts social class is income.

Vol. 2023 (2) 91 – 95
ISSN 2688-3651 [online]

Income inequality within different cultures impacts people's perceptions of societal class placement. Specifically, in today's society, there is a significant gap that continues to grow between the lower class and the upper class. In the United Kingdom (UK), the top one-fifth of households have 40% of the national income, while the bottom one-fifth have only 8% (Evans & Mellon, 2016; Social Mobility Commission, 2017). This wage gap is not uncommon. According to Nijman (2020), this disparity is the most extreme in the United States. For example, "in 2016, the top 1% of income earners in New York, took 40% of all income while the bottom half had to make do with 6%," (Nijman 2020).

Interestingly, it is likely that self-perception of social class is not consistent with actual earned income. For example, about 14% of the UK's population is lower class. However, 49-60% of UK citizens define themselves as lower class (Evans & Mellon, 2016). Part of this might be because people don't only determine their social class by income, but by other factors such as education or occupation. This is socioeconomic status (SES). One influencing factor of SES is an individual's level of education.

2.1 Social Class and Education

Education level has been demonstrated to play an essential part in social class attainment (Assari, 2018). For example, an interview-based study with American college students found that the wealthier the students were, the more importance they put on that wealth (Manstead, 2018). SES also makes it harder for people-- mostly lower class-- to leave the SES that they were born into. Children from low-SES families enter high school with an average of literacy skills being 5 years behind those of high-income students (Rearson et al., 2013). This is further demonstrated by lower-class students historically having more difficulty getting into and attending college. The cost of college is a burden for many lower SES families, making it sometimes improbable for lower SES individuals to advance in social class. Over the last two decades, college costs have significantly increased and lower-class students may not be able to afford the cost, may not want to incur tremendous debt, or can't move far from home because they have to take care of family members. Some may also have to pay for their own college and they may have jobs that they can't afford to lose if they went to college. Even for those who attend college, SES has been demonstrated to have lower incomes, educational attainment, and graduate school attendance than higher SES students (Walpole, 2003).

It is likely that the school system promotes ideals that middle-class and upper-class children are more familiar with and more likely to benefit from. The design of the American education system structure helps upper and middle-class students outperform their lower-class peers, increasing opportunities later in life. Lower-class children may also be limited in the college that they can attend. Two studies were conducted to show that "highly able students from socially disadvantaged backgrounds are more likely to settle for less prestigious universities" (Manstead, 2018). On top of that, lower-class students generally have more to worry about than upper-class students. Even if a disadvantaged student has high enough grades and test scores to get into a prestigious college, they opt instead to either not go to college, or go to a community college (Marcus & Hacker 2015). This is likely not due to a lack of motivation, but limited opportunities that stem from being in a lower SES.

2.2 Social Class and Views on Others

One's own social class has been demonstrated to impact views on other people's financial situations. For example, there is also a rising trend for middle or upper-class citizens to believe that those who live in poverty do so because of laziness instead of social injustice (Clery et al., 2013). This concept extends beyond socioeconomic class. For example, higher socioeconomic classes aren't the only ones who judge those below them. White lower-class people are prejudiced, especially towards ethnic minorities and immigrants.

On the other hand, in multiple studies, it was found that these working and lower-class individuals also have more empathy for others (Piff et al., 2010). This poses an interesting question: Why is it that those in a lower class experience more empathy, but are also possibly intolerant of immigrants who may be coming from the same class as them? Two highly likely reasons for this disconnect are ethnicity and fear.

Vol. 2023 (2) 91 – 95
ISSN 2688-3651 [online]

While social class is a very important part of our identities, ethnicity plays a large role. It is possible that lower-class people may have more empathy, but the empathy is primarily for people from their own ethnicity. Immigrants are more likely to have jobs similar to the lower class, so those lower-class citizens see immigrants as threats who are capable of "stealing" their job (Manstead, 2018). In other words, this bias toward others stems from irrational fears and concerns about other ethnicities. People in all SES ranges exhibit these biased behaviors. However, the behaviors of different SES do vary. It is possible that lower-class individuals blame outside factors for their shortcomings, and subsequently will therefore blame other individuals--such as immigrants--when they are unable to advance their own SES status. Meanwhile, the upper class generally believes that whatever happens to them is not so much outside factors and more their own behaviors and accomplishments. The differences in locus of control and success have been well studied in psychological research (Ladau, 1995).

3. Gender Relates to Identity

Gender has been proven to be highly related to identity. In nearly every society, gender roles have influenced the way women are treated. Historically, women are expected to be nurturers, kind and compassionate. When women act in a more masculine way, it results in increased negative reactions due to stereotypical expectations (Rincker, 2009). This often translates to being involved in caring-type occupational positions, such as homemakers, nurses, teachers, and therapists, which are viewed as more feminine occupations. The lack of women in more "masculine" occupations has been demonstrated consistently in politics, a classically male-dominated field.

3.1 Gender and Political Opportunities

Women have historically had to work much harder than men to be involved in leadership positions across different levels and different countries in order to have their thoughts heard and taken seriously. Women have had fewer political leadership positions, despite demonstrating success when in those positions. Even in the United States of America, there has never been a woman president, and the first woman vice president was finally elected in 2020. As a result, women often face the dilemma of being "mascinulized or marginalized" (Rincker, 2009). This means that female politicians have the challenge of either having to change their views to match with their male counterparts or continue to advocate for what they want to be changed, but in doing so, risk being ignored or dismissed, mainly because of their gender.

Furthermore, male politicians often respond to female politicians by acting more aggressively toward them or giving little to no importance to women's political accomplishments. This might be because they view women as outsiders. Historically, most governments are classically male dominated. It has been shown that in more right-wing governments, fewer women are in office, there are fewer feminist groups, and women tend to have less power (Rincker & Ortbals, 2007).

Interestingly, it has been proven that governments where women have more power often have more peaceful policies and less civil conflict (Piccone, 2017). This influences foreign policy, and gender equality throughout the country, and lessens violence against women. It was also found that when women have equal rights in politics as men, the government has a positive increase in democracy. Women in government also pay more attention to the stereotypically more "feminine" concerns such as social welfare, legal protection, and transparency in government and business. Men often view these as "soft" issues of security and focus less on them, when, in reality, they are extremely important (Piccone, 2017). This demonstrates the importance of women being involved in political leadership for the consideration and betterment of societal issues.

4. Conclusion

Historically, we are able to pinpoint the major effects that identity has on society. Self-identity leads to actual and perceived opportunities that exist in all parts of life, including educational attainment, occupation, and political aspirations. Although our world is developing and advancing in many ways, several identity-based inequalities still

Vol. 2023 (2) 91 – 95
ISSN 2688-3651 [online]

haunt society. Identity is an important part of who we are and influences a person's opportunities. While there is a potential for positive aspects of one's identity, it can oftentimes lead to disadvantages as well. Identity is a reason why people are treated unequally. Equality in terms of education, socioeconomic status, and acceptance of others as well as increased inclusion of women in politics would not only improve an individual's self-identity but also result in improved societal well-being. While these issues have evolved, there is room for growth that could lead to better health, decreased disparity, and more positive outcomes. Specifically, possible solutions could include increased public funding for lower-income public education to close the gap and decrease socioeconomic disparity through free job training and vocational placement. Furthermore, to increase women's political involvement, the creation of initiatives similar to those which have recently occurred in women's involvement in the Sciences, Technology, Engineering, and Math (STEM) may be effective. Specifically, explicitly focusing and programming throughout school, beginning in elementary school on encouraging and advocating for young girls to be exposed to and learn about political science and government could create a healthy environment in which women would be more involved in politics. This could also result in men being more accustomed to having female counterparts in politics, potentially allowing for more dialogue and less marginalization of women in politics. These possible solutions could ultimately begin to decrease social class and gender disparity by evolving society through increased inclusion where necessary.

Disclosure

Mr. Conrad Warner is the high school teacher who initially provided an assignment in which the lead author completed the research and initial draft of an article. Mr. Warner provided initial review and guidance for the assignment and initial draft. Mr. Warner was unable to provide guidance after the school year ended. The current article underwent multiple revisions and changes under the research advisor, prior to submission.

References

Assari, S., Preiser, B., & Kelly, M. (2018). Education and Income Predict Future Emotional Well-Being of Whites but Not Blacks: A Ten-Year Cohort. *Brain Science*, 29 8(7):122.
doi: 10.3390/brainsci8070122. PMID: 29966278; PMCID: PMC6070982.

Clery, L., Lee, L., & Kunz, S. (2013). Public attitudes to poverty and welfare, 1983-2011. Analysis using British Social Attitudes data. *NatCen Social Research.*

Cole, E. R., & Omari, S. R. (2003). Race, Class and the Dilemmas of Upward Mobility for African Americans. *Journal of Social Issues, 59*(4), 785–802.

Evans, G., & Mellon, J. (2016). Social Class: Identity, awareness, and political attitudes: why are we still working class? *British Social Attitudes*, *1-19*. https://www.bsa.natcen.ac.uk/media/39094/bsa33_social-class_v5.pdf

Landau, R. (1995). Locus of control and socioeconomic status: Does internal locus of control reflect real resources and opportunities or personal coping abilities? *Social Science & Medicine, 41*(11), 1499–1505. https://doi.org/10.1016/0277-9536(95)00020-8

Manstead, A.S.R. (2018), The psychology of social class: How socioeconomic status impacts thought, feelings, and behaviour. *British Journal of Social Psychology,* 57, 267-291. https://doi.org/10.1111/bjso.12251

Marcus J. & Hacker H. (2015, December). "The Rich-poor Divide on America's College Campuses Is Getting Wider, Fast." The Hechinger Report hechingerreport.org/the-socioeconomic-divide-on-americas-college-campuses-is-getting-wider-fast/. Accessed 2 Jan. 2023.

Nijman J. & Wei Y.D. (2020). Urban inequalities in the 21st-century economy. Applied Geography. Apr 117:102188. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7124478/

Vol. 2023 (2) 91 – 95
ISSN 2688-3651 [online]

Ortbals, C. & Rincker, M. (2009). Fieldwork, Identities, and Intersectionality: Negotiating Gender, Race, Class, Religion, Nationality, and Age in the Research Field Abroad: Editors' Introduction. *Political Science & Politics,* 42(2), 287-290. doi:10.1017/S104909650909057X

Piccone, T. (2017). Democracy, Gender Equality, and Security. Policy briefs. https://www.brookings.edu/wp-content/uploads/2017/08/fp_20170905_democracy_gender_security.pdf

Piff, P.K., et al. (2010). Having less, giving more: The influence of social class on prosocial behavior. *Journal of Personality and Social Psychology, 99*(5), 771–784.
https://doi.org/10.1037/a0020092

Reardon, S.F., et al. (2013). Patterns and Trends in Racial Academic Achievement Gaps Among States, 1999-2011. Stanford University supported by the Institute of Education Sciences, U.S. Department of Education
https://cepa.stanford.edu/sites/default/files/reardon%20et%20al%20state%20achievement%20gaps%20aug2013.pdf

Rincker, M.E. (2009). Masculinized or Marginalized: Decentralization and Women's Status in Regional Polish Institutions*, Journal of Women, Politics & Policy,* 30(1), 46-69.

Social Mobility Commission. (2017). State of the Nation 2017: Social Mobility in Great Britain. *Open Government Licence*.https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/662744/State_of_t he_Nation_2017_-_Social_Mobility_in_Great_Britain.pdf

Walpole, M. (2003). Socioeconomic Status and College: How SES Affects College Experiences and Outcomes. *Review of Higher Education: Journal of the Association for the Study of Higher Education, 27*(1), 45–73.
https://doi.org/10.1353/rhe.2003.0044

Journal of Research High School

Vol. 2023 (2) 96 – 105
ISSN 2688-3651 [online]

Lyme Disease: A Grand Challenge

Roman Frankel[1] *

[1]Score Academy Boca Raton, Boca Raton, FL, USA
*Corresponding Author: romanfrankel.school@gmail.com

Advisor: Karl Anderson, ande3701@crk.umn.edu

Received October 6, 2022; Revised July 1, 2023; Accepted, July 24, 2023

Abstract

Lyme disease (Lyme borreliosis), first discovered in 1908, is the most common vector-borne disease in the United States, with around 476,000 new cases per year. As a global disease, the impact of Lyme disease is such that it should be thought of as one of the Grand Challenges in Global Health. Despite its pervasiveness, there are very few approved tests that are accurate in detecting it. Similarly, there are even fewer treatments for Lyme disease after it passes its initial stages. It is vital for the medical community and the general public to further educate themselves on the early symptoms of Lyme disease and the overall prevalence of vector-borne illnesses so that they can be treated before they suffer immense repercussions. Moreover, further research needs to be done into treatments, testing, and the impact of climate change on disease risk.

Keywords: Identity, Gender, Socioeconomic class, Education, Politics

1. Introduction

1.1 Objectives

Aims of this review article:

a) To raise awareness about the prevalence and impact of Lyme disease as a Grand Challenge in Global Health.
b) To highlight the need for accurate diagnostic tests for Lyme disease.
c) To emphasize the importance of early detection and treatment of Lyme disease.
d) To identify the gaps in research related to treatments, testing, and the impact of climate change on disease risk.

1.2 Hypotheses/Implications

The prevalence of Lyme disease is significantly underestimated, and there is an increased need for education and awareness related to the early symptoms and ubiquity of it. Current diagnostic tests for Lyme disease, such as the two-tier serological test, have limitations and may lead to false results, indicating a need for the development of more specific and sensitive tests. The early detection and prompt treatment of this disease can markedly reduce the risks of long-term complications and patient outcomes. Further research on treatments, testing methods, and the impact of climate change on disease risk contribute to the development of more effective strategies for Lyme disease prevention and treatment.

This review provides a broad synopsis of Lyme disease and highlights important areas for further examination. The details discussed, and information provided support the provided objectives.

1.3 Background

The discovery of diseases is not usually accredited to dermatologists; however, this has been known to occur in specific instances. Dermatologists, as its prefix “derma” would imply, specialize in skin-based conditions, putting

Vol. 2023 (2) 96 – 105
ISSN 2688-3651 [online]

them at the forefront of uncovering diseases or markings indicative of a particular disease. One of these distinctive lesions is the erythema migrans (EM), a tell-tale sign of Lyme Disease (LD) (Radolf, et al., 2020; Bockenstedt and Wormser, 2014; Bratton, et al., 2008), which a Swedish dermatologist named Arvid Afzelius first observed in 1908 (Burgdorfer, et al., 1983; Afzelius, 1921) and was soon associated as a symptom of a tick bite (Bratton, et al., 2008). Despite being observed in Europe during the early 1900s, erythema migrans would not be reported in the United States until 1970 by a Wisconsin dermatologist named Rudolph Scrimenti (Radolf, et al., 2020). Shortly after this, physicians in New England identified groups of children in areas near Lyme, CT, who shared a strange rash that resembled EM and had other symptoms that had previously been associated with it. Due to the close geographical proximity of the children and the symptoms that they had in common, the belief was that the children all had a similar illness, likely transmitted by a tick. In 1977, more than half a decade after EM was first observed, the condition the children suffered from was named Lyme Disease (Bratton, et al., 2008).

Following the discovery of LD in the United States, scientists began researching potential causes of the condition. Soon, in 1981, Willy Burgdorfer and colleagues uncovered a previously unknown spirochete known today as *Borrelia burgdorferi* (Steere, et al., 2004; Radolf, et al., 2020). Later, *Borrelia burgdorferi* spirochetes were found in the intestinal tracts of *Ixodes dammini* ticks, known today as *Ixodes scapularis*. Subsequently, in 1982, The Centers for Disease Control and Prevention (CDC) began surveillance for Lyme disease; by 1991, Lyme disease was classified as a nationally reportable disease. After starting surveillance, researchers noticed a gradual uptick in cases each year. By 1992, there had been 9677 confirmed cases of Lyme Disease across 47 states, a slight increase from the year prior (Lyme Disease 4, n.d., 1993). Most cases were reported from the Northeast, Mid-Atlantic, and Pacific Coastal regions (Lyme Disease 3, n.d., 1994). This steady increase in cases would continue into the new millennium, with cases going from around 11,000 in 1995 to approximately 18,000 in 2000 (Lyme Disease 1, n.d., 1997; Lyme Disease 2, n.d., 2002). Figure 1 demonstrates the yearly increase in Lyme disease cases in the United States.

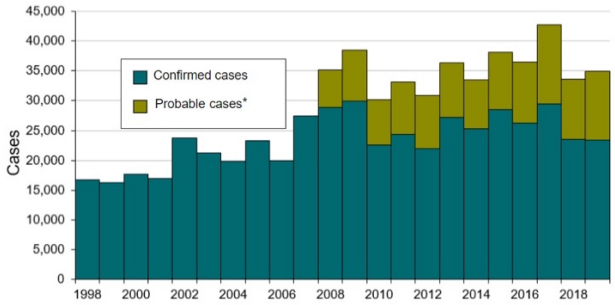


Figure 1. Reported cases of Lyme disease in the United States by year; 1998-2019. (Surveillance Data, 2022)

Today, Lyme disease is the most reported vector-borne illness in the United States, having been reported in all 50 states (NCEZID: Vector-Borne Diseases, 2019). Although state and local health departments estimate that there are around 35,000 new cases of Lyme disease each year, the CDC estimates that about 476,000 Americans are diagnosed with Lyme and receive treatment for it each year. Despite the high number of new cases reported to the CDC each year, the National Institutes of Health (NIH) investment in research into Lyme disease remains relatively small compared to other infectious diseases (Bobe, et al., 2021). In addition to the increase in cases each year since its discovery, Lyme disease case numbers in each state have also been increasing accordingly. Much like that of the United States, the incidence of Lyme Disease continues to grow in both Canada and Europe, with Western Europe having greater than 200,000 cases per year (Marques, et al., 2021), and Canada has its highest number of cases to date in 2021, with nearly 3,000 cases (Lyme disease: Monitoring, 2022). The incidence of LD in Europe is highest in Northern Europe, most notably in Scandinavian states. Additionally, LD is prevalent in Central Europe. In particular, Czechia, Germany, Austria, and nations in the Balkan Peninsula. Interestingly, *B. burgdorferi* is not found in *I. ricinus and I. persulcatus;* instead, LD is caused by *B. garinii and B. afzelii,* with most cases of LD in Europe being caused by *B. afzelii. To date,* neither of these *Borrelia* strands is found in the United States (Marques, et al., 2021; Vanderkerckhove, et al., 2019). Much like their Western counterparts, China also suffers from endemic Lyme disease, with confirmed cases in 29 out of 31 provinces. (Wu, et al., 2013) In the four northernmost provinces, approximately 3 million are bitten by ticks, with around 30,000 people infected with Lyme borreliosis (Wu, et al., 2013). Inexplicably, LD in China is more complex than LD in other regions. There is a noticeable difference in the species of tick and Borrelia between Northern and Southern China. Northern China is the

Vol. 2023 (2) 96 – 105
ISSN 2688-3651 [online]

less complex region, with I. persulcatus acting as the primary vector, transmitting both B. afzelii and B. garinii. However, in Southern China, it has yet to be entirely known what the most prominent vector is; however, it is likely one of I. granulatus or H. bispinosa. Despite this, the peak of incidence is the same in both regions between June and August (Wu et al., 2013; Hao et al., 2011). With the global reach of Lyme disease and the classification of vector-borne illnesses as Grand Challenges in Global Health, Lyme disease should be thought of as one of these Grand Challenges that the world must invest funding into in attempts to find a means of prevention and a potential cure.

Table 1: Disease expenditures and number of reported cases.

Disease	NIH Funding FY 2021 (in millions)	Most recent number of reported cases in the USA	Funding reported per case	Deaths per year in the United States
HIV/AIDS	$3,082	34,800[b]	$88,563	~13,000
Malaria	$229	~2,000[c]	~$114,500	~5
Tuberculosis	$594	7,174[d]	~$82,800	526
Hepatitis-B	$149	~3,000[e]	$49,666	~3,000
Lyme disease	$39	~ 34,945[f] (~476,000 estimated cases)[g]	$1,116 ($81.93 per estimated case)	N/A

a. https://report.nih.gov/funding/categorical-spending#/
b. https://www.hiv.gov/hiv-basics/overview/data-and-trends/statistics
c https://www.cdc.gov/parasites/malaria/index.html
d https://www.cdc.gov/tb/statistics/tbcases.htm
e https://www.cdc.gov/hepatitis/statistics/2018surveillance/HepB.htm
f. https://wonder.cdc.gov/nndss/static/2019/annual/2019-table2j.html
g. https://www.cdc.gov/lyme/stats/humancases.html

2. Clinical Manifestations

In most cases, patients experience the onset of LD in stages in reaction to the immune response of the *Borrelia burgdorferi* spirochete. The first stage, known as early localized LD, is estimated to occur 3-30 days after being bitten and presents in most cases as an erythematous, annular-shaped lesion that often contains a defined center and expands outwards. This lesion, colloquially known as a "bullseye rash" due to its likening to a bullseye, is the EM lesion that Afzelius discovered in 1908 (shown in figure 2). Despite a widespread belief that one must display EM to have Lyme disease, EM is not present in around 15-30% of patients (CDC, 2022; Aucott, et al., 2009; Bratton, et al., 2008). In addition to EM, most patients experience "flu-like symptoms," including but not limited to fever, chills, fatigue, joint pain, headache, and myalgia. In some cases where there is a lack of EM, patients can experience swelling of the lymph nodes. Occasionally, there can be a tingling or burning sensation associated with EM. Intriguingly, many cases of LD in Europe display clinical manifestations that have not been observed in other regions: Acrodermatitis chronica atrophicans (ACA) and Borrelial lymphocytoma. ACA is a cutaneous manifestation of LD primarily found on the hands and feet. It begins with a reddish-blue discoloration as well as swelling of the skin. This swelling continues to enlarge and can be followed by atrophic changes several months to years later. Borrelial

Figure 2. A, B, and C. Erythema migrans presenting as a single lesion (A, B) and as multiple lesions (C). (Lyme Disease Rashes and Look-Alikes, 2022)

Vol. 2023 (2) 96 – 105
ISSN 2688-3651 [online]

lymphocytoma first appears as a small area of skin induration which slowly grows into a solitary nodule. It is primarily found on the breast in adults and on the earlobes of children. Due to the lack of a presence of these manifestations in the United States, they are likely caused by the *B. afezlii* infection (Marques, et al., 2021).

Following the conclusion of the early localized infection, the disease progresses to the early disseminated stage, occurring days to weeks after infection. As spirochetes spread from the site of infection, additional EMs can appear (Bobe, et al., 2021). Furthermore, patients can develop various musculoskeletal symptoms such as migratory muscle or joint pain, swelling of the joints, and the development of arthritis. Neurological issues can also appear weeks or months after infection, most often as potentially bilateral seventh cranial nerve palsy. Meningitis, radicular neuropathies, and neuroborreliosis can occur with or without nerve palsy. If left untreated after this stage, LD progresses to its final stage, late disseminated infection. Symptoms of this stage can include intermittent pain and swelling of one or more joints, typically the knees and hips. Like in the early disseminated stage, neurological manifestations such as polyneuropathy and encephalomyelitis can develop (Radolf, et al., 2020; Bockenstedt and Wormer, 2014). Encephalomyelitis specifically can cause somewhat debilitating effects, including insomnia, changes in personality, and impaired mental ability, leading to memory issues and brain fog.

Arguably the most severe symptom of LD is Lyme carditis. Despite being present in only 1% of cases reported to the CDC (Bockenstedt and Wormer, 2014), Lyme carditis is unlike other LD symptoms because it can be deadly. Patients might experience heart palpitations, shortness of breath, and lightheadedness, common symptoms of an atrioventricular nodal block. If left untreated, Lyme carditis can progress to a complete heart block, which could cause a higher probability of sudden death. This symptom can manifest in any of the stages discussed above. It is believed to occur because of *Borrelia* spp. coming in contact with heart tissue. This action causes an exaggerated macrophagic and lymphocytic response within cardiac tissues (Lyme carditis: A can't miss diagnosis., 2020). Despite the relatively small number of spirochetes found within the heart tissue of this Lyme carditis, there tends to be an excessive inflammatory response. Despite the severity of the symptom, it can be fully treated with the appropriate antibiotics and therapy (Lyme carditis: A can't miss diagnosis., 2020).

3. Diagnosis & Testing

In most cases, LD is diagnosed based on the presence of clinical manifestations such as an EM lesion. Keeping in track with other tick-borne illnesses, anywhere from 30-50% of patients do not recall a tick bite, usually because deer tick nymphs are small and can go unnoticed. Typically, standard laboratory tests are not well equipped to diagnose LD as they cannot distinguish it from other entities (Bockenstedt and Wormser., 2014; Bratton et al., 2008). White blood cell counts have the potential to be elevated or standard, and the results of hemoglobin, hematocrit, creatinine, and urinalysis testing all tend to be within expected limits. Furthermore, PCR testing of blood, serum, or plasma is unreliable for diagnosing LD because spirochetes are transient and have a low copy number (Bobe, et al., 2021). However, in the early stages of infection, liver function tests may reveal mild elevations (Bockenstedt and Worsmer., 2014). A biopsy of the EM lesion is a more helpful tool for diagnosis, but this step is unnecessary, as those with an EM lesion should begin antibiotic treatment as soon as diagnosed.

The mainstay of laboratory diagnosis for LD is a two-tier serological test consisting of first an enzyme-linked immunoblot assay (ELISA) followed by a separate Western immunoblot test if the ELISA result is positive or equivocal (Bobe, et al., 2021; Sanchez et al., 2016; Bockenstedt and Wormser, 2014; Stricker and Johnson., 2011; Bratton, et al., 2008). The reasoning for the two-tiered structure of testing is that the ELISA test detects antibodies but does not test specifically for *B. burgdorferi*. In many cases, this can provide false positives, but if it is negative, it is unlikely that the patient has LD. A concern arises with false negatives, as in early cases of LD, it is improbable that *B. burgdorferi* antibodies will be present. If the test is positive or inconclusive, it is then recommended for the patient to undergo a Western blot test. If this result is positive, the patient has likely contracted LD. If the result is negative, it suggests that the patient does not have LD and the ELISA test was a false positive (Lyme Disease Diagnostics Research, 2022). In addition, serological tests are incapable of distinguishing prior exposure to *B. burgdorferi* from an active infection (Bobe, et al., 2021). However, this test is not necessarily essential to diagnosing LD and primarily acts as a tool to further confirm diagnosis after the presence of an EM lesion.

Vol. 2023 (2) 96 – 105
ISSN 2688-3651 [online]

To date, there are very few FDA-approved tests for diagnosing LD. The two-tiered serological testing remains the most common form of LD test and is primarily recommended by the CDC. However, this test is not without its faults, and those who seek it should be wary of false results. Naturally, this should be of concern since not all LD patients develop an EM lesion. There is a need for a highly sensitive and specific test that can reliably detect infection of multiple strains of *Borrelia* at all stages of infection.

4. Coinfections

Despite the prevalence of LD in the US, it is not widely known to physicians that several coinfections can be found along with it that can cause a variety of symptoms and make treatment for LD more complicated. Of these coinfections, Babesiosis, Human Granulocytic Anaplasmosis (HGA), and Bartonella are the most commonly found in LD patients.

4.1 Babesiosis

Babesiosis, a worldwide tick-borne infection caused by hemoprotozoan parasites of the genus Babesia, has a similar geographic expansion to that of LD, albeit more restricted. This infection is spread via the bites of Ixodes ticks, a commonality between LD and its coinfections. The bacteria that cause this condition is *Babesia microti*, and the prevalence of *B. microti* infection in nymphal *I. scapularis* ticks can range from 1% in areas where it has recently become endemic to around 20% in well-established areas (Vannier, et al., 2015). Babesiosis has a lower incidence than LD, despite an exponential increase in the past five decades, a more significant proportion of asymptomatic infection, and a more remarkable lack of physician awareness.

Diagnosis of active babesiosis is usually made via visualization of Babesia parasites on Giemsa- or Wright-stained thin blood smears (Sanchez, et al., 2016). Due to the small size of the Babesia parasites, thick blood smears are not recommended. Symptoms can include chills, sweats, headaches, body aches, nausea, fatigue, and loss of appetite. Since *Babesia* parasites infect red blood cells, babesiosis can lead to hemolytic anemia, a disorder in which red blood cells are destroyed faster than they can be produced (CDC, 2022; Hemolytic Anemia, 2022). A one-week course of antibiotics should be considered for those with an asymptomatic Babesia infection. For those suffering from mild to moderate babesiosis, a 7–10-day course of antibiotics such as atovaquone and azithromycin or clindamycin and quinine is recommended for severe babesiosis, which typically develops in patients who suffer from external risk factors, intravenous treatments in hospitals are recommended.

4.2 HGA

Human Granulocytic Anaplasmosis (HGA) is a deer tick-transmitted rickettsial infection commonly found in the Northeast and Upper Midwest and a common cause of fever. Although the disease usually resolves in most cases, as many as 3% of patients may develop life-threatening complications, and almost 1% succumb to these complications (Bakken and Dumler., 2015). The incidence of HGA increased 12-fold between 2001 and 2011 (Bakken and Dumler., 2015). HGA is carried by *Ixodes scapularis* ticks in the Northeast and Midwestern United States and by the *Ixodes pacificus* throughout the West Coast (Transmission, 2022). The usual symptoms of HGA include fever, chills, severe headache, nausea, vomiting, diarrhea, and muscle aches. However, if left untreated, HGA can cause severe illness, including bleeding problems, respiratory failure, organ failure, and death. (CDC, 2022). Much like LD, HGA is commonly treated with antibiotics, most commonly doxycycline. Doxycycline treatment is highly efficacious, and there tends to be a marked improvement in symptoms within 24-72 hours (CDC, 2022)

4.3 Bartonella

Bartonellosis, also known as cat scratch disease, is a condition carried by vectors, primarily fleas and animal bites. It is caused by the *Bartonella henselae* bacterium, an intracellular parasite that prefers red blood cells and endothelial

Vol. 2023 (2) 96 – 105
ISSN 2688-3651 [online]

cells (Bartonellosis, 2022). It has previously been thought that *B. henselae* is not transmittable via tick bite. However, studies have shown that *Ixodes* ticks can potentially be vectors for *B. henselae* (Cotté, et al., 2008). Bartonellosis first appears as a papule which eventually develops into a pustule. In patients with competent immune systems, symptoms are typically limited to region adenopathy and occasional fevers. However, in immunocompromised patients, more severe symptoms such as bacillary angiomatosis, a condition in which tumor-like masses form due to the proliferation of blood vessels, and endocarditis (Bartonellosis, 2022). Other manifestations of Bartonellosis can include but are not limited to depression, reduced impulse control, sleep disorders, Guillain-Barré syndrome, and osteomyelitis (Berghoff., 2012). Antibiotic treatment is not usually recommended for those with competent immune systems, as the infection should resolve independently. However, antibiotic treatment, such as doxycycline and erythromycin, is recommended for those with more severe symptoms or a more complicated condition.

5. Treatments and preventions

If caught in its early stages, Lyme disease can be effectively treated with oral antibiotics such as doxycycline or amoxicillin for 14-21 days (Bobe, et al., 2021). However, a shorter treatment is less effective as the disease progresses, and a longer treatment time is likely recommended. In cases of neurologic-Lyme disease, such as meningitis or Lyme arthritis, oral antibiotics are less effective, and intravenous antibiotics are preferable (Johns Lyme Disease Treatment, 2022). Several studies have found that complementary natural and herbal products can help treat LD (Bobe, et al., 2021). These treatments have been shown to have activity against *Bb and Babesia*, but further trials are needed to evaluate their effectiveness. Approximately 7-30% of LD patients develop Jarisch-Herxheimer reaction (JHR) (Nykytyuk, et al., 2020; Bratton, et al., 2008) within 24 hours of starting therapy. This reaction appears as a worsening of LD symptoms such as fever, sweating, general malaise, and headache. The JHR is believed to be caused by the release of harmful toxins by dying microorganisms during antibiotic treatment, but the cause is not definitively known. The clinical manifestations of JHR can vary in expression, time of onset, and duration. As a result, physicians tend not to be aware of the presence of JHR and mistake it for an allergic reaction to antibiotic treatment (Nykytyuk, et al., 2020).

The concept of a vaccine for LD is one of much discussion. While a vaccine was developed and approved for use in the late 1990s, it was quickly made unavailable to the public. Limited efficacy, low demand, high price, and a potential association with the development of autoimmune-related arthritis led to the downfall of the original LD vaccine. In 2017, the French biotech company Valneva began developing an LD vaccine with the help of Pfizer. This vaccine, dubbed VLA15, has proved promising in early animal-based trials as it has shown efficacy against spirochetes that express the outer surface protein A (OspA), an abundant immunogenic lipoprotein of *Borrelia burgdorferi* (Comstedt, et al., 2017). Preclinical studies have shown that immunization with VLA15 has demonstrated protection from five types of OspA, and in addition to this, antibodies that contain VLA15 provide further protection upon passive immunization (Comstedt, et al., 2017)

Personal protection measures should be taken in areas where LD is expected to prevent infection. Such protection methods include but are not limited to protective clothing and insect repellants. Many health organizations recommend that those in areas where ticks are commonly found wear long-sleeved, light-colored clothing so that ticks are more visible if attached (Ticks and Lyme Disease, 2022). It is also recommended to avoid walking through tall bushes and other vegetation and instead walk in the center of trails. Moreover, it is recommended to tuck pant legs into socks or shoes and tuck shirts into the waistband of pants to prevent ticks from crawling through the space between articles of clothing to get to the skin. The Environmental Protection Agency (EPA) recommends using registered insect repellents that contain 20% or higher concentration of DEET for skin and products that contain 0.5% permethrin for clothing to prevent LD. (Prevent Lyme Disease, 2022; Symptoms and causes, 2022)

5.1 Ten suggestions that can help prevent the contraction of Lyme disease

- Know where to expect ticks.
- Seek to avoid brushy, woody, and grassy areas. Especially during regional periods of high tick activity

Vol. 2023 (2) 96 – 105
ISSN 2688-3651 [online]

- Wear long pants and long-sleeved shirts, as well as closed-toed shoes.
- Wear light-colored clothing for easier visibility of ticks present on clothing.
- Wear a hat
- Tuck pant legs into socks or shoes, and tuck shirts into pants
- Apply an insect repellant that contains DEET on clothing and uncovered skin
- Walk in the center of trails
- Carefully check for ticks after outdoor activities
- Look into potentially purchasing repellent-treated clothing.

6. Post-Lyme Disease Syndrome

Post-Lyme Disease Syndrome (PTLDS), also called chronic Lyme disease, is a term typically applied to people with otherwise-unexplained symptoms of LD lasting more than six months after completion of antibiotic treatment. PTLDS occurs in below 10% of patients diagnosed and treated for LD (Bockenstedt and Wormser., 2014.) Despite this, PTLDS can be debilitating with persistent and recurring symptoms of PTLDS, including severe fatigue, myalgias, cognitive disruptions, and arthralgias (without the presence of arthritis) (Bobe, et al., 2021; Rebman and Aucott, 2020; Bratton, et al., 2008). More symptoms are displayed in Figure 3, as shown below.

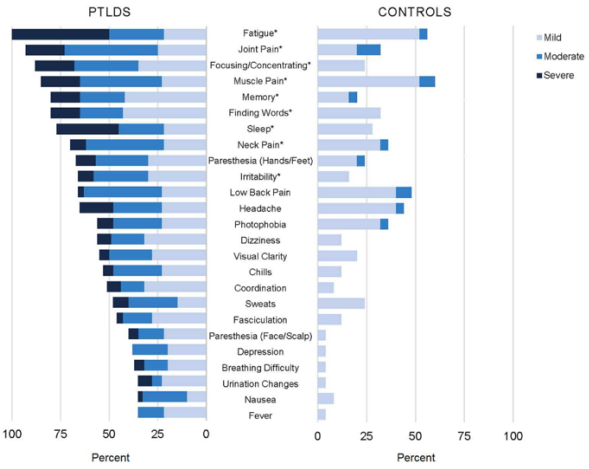


Figure 4. Participants with PTLDS and controls were asked about the presence and severity of 36 signs/symptoms over the past two weeks. The 25 signs/symptoms display a statistically significant difference in severity by group. (Teixeira, et al., 2017)

The reasons for the continuation of symptoms in some patients following treatment for LD are unknown. Residual tissue damage, slowly resolving inflammation, and a cytokine-induced illness because of raised levels of circulating cytokines before and during treatment of LD (Bockenstedt and Wormser, 2014; Bratton et al., 2008). These symptoms can last for months or years following treatment of LD, and there are no FDA-approved or commonly agreed-upon treatments for patients suffering from PTLDS (Bobe, et al., 2021; Rebman and Aucott, 2020).

Furthermore, no current test can objectively prove the presence of PTLDS, and there seems to be no commonality in antibody levels found in those who suffer from PTLDS (Bobe, et al, 2021; Bockenstedt and Wormser, 2014). According to the CDC, patients with PTLDS typically get better as time goes on; however, it can take many months or years to feel completely well.

7. Remaining Problems

To date, LD case numbers continue to increase at an increasing rate and do not show any signs of stopping. The issue of climate change should concern all sectors of the world, especially the medical and vector-borne illness community. According to the American Public Health Association (APHA), the development of ticks, and LD via proxy, is heavily influenced by climatic factors. More specifically, temperature, precipitation, and humidity. With the looming threat of global warming, areas with a high incidence of LD will likely experience milder winters and, as a result, an increase in both tick population and risk of contracting LD (APHA, 2022.) The life cycle and abundance of *Ixodes* ticks, particularly *scapularis*, are strongly influenced by temperature (Dumic and Severnini, 2018), with *I. scapularis* being most active in areas with temperatures above 45˚F and prospering in areas with at least 85% humidity

Vol. 2023 (2) 96 – 105
ISSN 2688-3651 [online]

(EPA, 2022). It is clear to see the link between rising temperatures and humidity and increased tick population. It has been falsely believed in the past that reforestation aided the expansion of the tick life cycle in the northeastern United States during the initial emergence of LD; new studies have shown that despite deforestation, *I. scapularis* population continues to increase and expand into new areas. It can be assumed that a considerable increase in the incidence of LD cases in endemic regions is at hand. This logic should alert public health agencies and officials, physicians, and everyday people to the imminent vector-borne illness crisis we will soon face. In times such as this, it is crucial for physicians and the general populace to educate themselves to be aware of the early symptoms of prevalent vector-borne illnesses, which can effectively be treated before debilitating consequences ensue. Moreover, vector-borne illnesses should be treated as a serious issue by public health authorities and thought of like other Grand Challenges in Global Health. As a result, there should be a vast increase in funding not only for Lyme disease but also for vector-borne illnesses as a whole.

References

Afzelius, A. (1921). *Erythema chronicum migrans*.

Bakken, J. S., & Dumler, J. S. (2015). *Human Granulocytic Anaplasmosis*. https://doi.org/10.1016/j.idc.2015.02.007

Bartonellosis | Lyme Disease. (n.d.). Retrieved July 8, 2022, from https://www.columbia-lyme.org/bartonellosis

Berghoff, W. (2012). Chronic lyme disease and co-infections: Differential diagnosis. *The Open Neurology Journal*, *6*(1), 158–178. https://doi.org/10.2174/1874205x01206010158

Bobe, J. R., et al. (2021). Recent Progress in Lyme Disease and Remaining Challenges. *Frontiers in Medicine*, *8*, 666554. https://doi.org/10.3389/FMED.2021.666554

Bockenstedt, L. K., & Wormser, G. P. (2014). Unraveling Lyme Disease. *Arthritis & Rheumatology (Hoboken, N.J.)*, *66*(9), 2313. https://doi.org/10.1002/ART.38756

Bratton, R. L., et al. (2008). Diagnosis and Treatment of Lyme Disease. *Mayo Clinic Proceedings*, *83*(5), 566–571. https://doi.org/10.4065/83.5.566

Burgdorfer, W., et al. (1983). Erythema chronicum migrans - a tickborne spirochetosis*. *Acta Trop*, *40*(1), 79–83. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85025834183&partnerID=40&md5=84a6211723ab734721e05760926ef180

CDC - Babesiosis - General Information - Frequently Asked Questions. (n.d.). Retrieved July 26, 2022, from https://www.cdc.gov/parasites/babesiosis/gen_info/faqs.html

Climate Change Indicators: Lyme Disease | US EPA. (n.d.). Retrieved July 8, 2022, from https://www.epa.gov/climate-indicators/climate-change-indicators-lyme-disease

Comstedt, P., et al. (2017). *The novel Lyme borreliosis vaccine VLA15 shows broad protection against Borrelia species expressing six different OspA serotypes*. https://doi.org/10.1371/journal.pone.0184357

Cotté, V., et al. (2008). Transmission of Bartonella henselae by Ixodes ricinus. *Emerging Infectious Diseases*, *14*(7), 1074. https://doi.org/10.3201/EID1407.071110

Diagnosis | Lyme Disease. (n.d.). Retrieved July 8, 2022, from https://www.columbia-lyme.org/diagnosis

Dumic, I., & Severnini, E. (2018). "ticking Bomb": The impact of climate change on the incidence of lyme disease. *Canadian Journal of Infectious Diseases and Medical Microbiology*, *2018*. https://doi.org/10.1155/2018/5719081

Vol. 2023 (2) 96 – 105
ISSN 2688-3651 [online]

Hemolytic Anemia | Johns Hopkins Medicine. (n.d.). Retrieved July 26, 2022, from https://www.hopkinsmedicine.org/health/conditions-and-diseases/hemolytic-anemia

Lyme carditis: A can't miss diagnosis | British Columbia Medical Journal. (n.d.). Retrieved July 6, 2022, from https://bcmj.org/articles/lyme-carditis-cant-miss-diagnosis

Lyme Disease 1 --- United States, 2000. (n.d.). Retrieved July 1, 2022, from https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5102a3.htm

Lyme Disease 2 -- United States, 1995. (n.d.). Retrieved July 1, 2022, from https://www.cdc.gov/mmwr/preview/mmwrhtml/00042555.htm

Lyme Disease 3 -- United States, 1993. (n.d.). Retrieved July 1, 2022, from *https://www.cdc.gov/mmwr/preview/mmwrhtml/00032251.htm*

Lyme Disease 4 -- United States, 1991-1992. (n.d.). Retrieved July 1, 2022, from https://www.cdc.gov/mmwr/preview/mmwrhtml/00020506.htm

Lyme Disease Diagnostics Research | NIH: National Institute of Allergy and Infectious Diseases. (n.d.). Retrieved July 8, 2022, from https://www.niaid.nih.gov/diseases-conditions/lyme-disease-diagnostics-research

Lyme disease: Monitoring - Canada.ca. (n.d.). Retrieved September 1, 2022, from https://www.canada.ca/en/public-health/services/diseases/lyme-disease/surveillance-lyme-disease.html

Lyme Disease Treatment | Prognosis for Lyme Patients after Treatment. (n.d.). Retrieved July 8, 2022, from https://www.hopkinslyme.org/lyme-disease/treatment-and-prognosis-of-lyme-disease/

Marques, A. R., Strle, F., & Wormser, G. P. (2021). Comparison of Lyme Disease in the United States and Europe. *Emerging Infectious Diseases*, *27*(8), 2017. https://doi.org/10.3201/EID2708.204763

Nykytyuk, S., et al. (2020). The Jarisch-Herxheimer reaction associated with doxycycline in a patient with Lyme arthritis. *Reumatologia*, *58*(5), 335. https://doi.org/10.5114/REUM.2020.99143

Post-Treatment Lyme Disease Syndrome | Lyme Disease | CDC. (n.d.). Retrieved July 28, 2022, from https://www.cdc.gov/lyme/postlds/index.html

Prevent Lyme Disease | Division of Vector-Borne Diseases | NCEZID | CDC. (n.d.). Retrieved July 8, 2022, from https://www.cdc.gov/ncezid/dvbd/media/lymedisease.html

Radolf, J. D., et al. (2020). Lyme Disease in Humans. *Current Issues in Molecular Biology 2021, Vol. 42, Pages 333-384*, *42*(1), 333–384. https://doi.org/10.21775/CIMB.042.333

Rebman, A. W., & Aucott, J. N. (2020). Post-treatment Lyme Disease as a Model for Persistent Symptoms in Lyme Disease. *Frontiers in Medicine*, *7*, 57. https://doi.org/10.3389/FMED.2020.00057

Sanchez, E., et al. (2016). Diagnosis, Treatment and Prevention of Lyme Disease, Human Granulocytic Anaplasmosis and Babesiosis. *JAMA*, *315*(16), 1767. https://doi.org/10.1001/JAMA.2016.2884

Schwartz, A. M., Hinckley, A. F., Mead, P. S., Hook, S. A., & Kugeler, K. J. (2017). Surveillance for Lyme Disease — United States, 2008–2015. *MMWR Surveillance Summaries*, *66*(22), 1. https://doi.org/10.15585/MMW.SS6622A1

Signs and Symptoms | Anaplasmosis | CDC. (n.d.). Retrieved July 26, 2022, from https://www.cdc.gov/anaplasmosis/symptoms/index.html

Vol. 2023 (2) 96 – 105
ISSN 2688-3651 [online]

Steere, A. C., Coburn, J., & Glickstein, L. (2004). The emergence of Lyme disease. *Journal of Clinical Investigation*, *113*(8), 1093–1101. https://doi.org/10.1172/JCI21681

Stone, B. L., Tourand, Y., & Brissette, C. A. (2017). Brave New Worlds: The Expanding Universe of Lyme Disease. *Vector Borne and Zoonotic Diseases*, *17*(9), 619. https://doi.org/10.1089/VBZ.2017.2127

Stricker, R. B., & Johnson, L. (2011). Lyme disease: the next decade. *Infection and Drug Resistance*, *4*(1), 1. https://doi.org/10.2147/IDR.S15653

Surveillance Data | Lyme Disease | CDC. (n.d.). Retrieved September 1, 2022, from https://www.cdc.gov/lyme/datasurveillance/surveillancedata.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Flyme%2Fstats%2Ftables.html

Symptoms and causes - Mayo Clinic. (n.d.). Retrieved July 26, 2022, from https://www.mayoclinic.org/diseases-conditions/lyme-disease/symptoms-causes/syc-20374651

Teixeira, M. C., et al. (2017). *The clinical, symptom, and Quality-of-life characterization of a Well-Defined group of Patients with Posttreatment lyme Disease syndrome*. *4*, 14. https://doi.org/10.3389/fmed.2017.00224

Ticks and Lyme Disease: Symptoms, Treatment, and Prevention | FDA. (n.d.). Retrieved July 8, 2022, from https://www.fda.gov/consumers/consumer-updates/ticks-and-lyme-disease-symptoms-treatment-and-prevention

Transmission | CDC. (n.d.). Retrieved July 26, 2022, from https://www.cdc.gov/anaplasmosis/transmission/index.html

Vannier, E. G., et al. (2015). Babesiosis. *Infectious Disease Clinics of North America*, *29*(2), 357. https://doi.org/10.1016/J.IDC.2015.02.008

Wu, X. B., et al. (2013). Distribution of tick-borne diseases in China. *Parasites & Vectors*, *6*(1), 119. https://doi.org/10.1186/1756-3305-6-119

Journal of Research High School

Vol. 2023 (2) 106 – 114
ISSN 2688-3651 [online]

Evaluating the Effect of Potential Career-Altering Injuries on NBA Athletes

Derek Gao[1] *

[1]Lexington High School, Lexington, MA, USA
*Corresponding Author: 27d.gao@gmail.com

Advisor: Dr. Ganesh Mani, ganeshm@andrew.cmu.edu

Received May 2, 2023; Revised June 15, 2023; Accepted, August 4, 2023

Abstract

This project aimed to evaluate the effect of three major injuries on the post-recovery performance of athletes in the National Basketball Association (NBA): anterior cruciate ligament (ACL) tears, Achilles tendon tears, and meniscus tears. All three are crucial in stabilizing the body during strenuous movements, such as pivoting, jumping, or accelerating. Recovery from these injuries takes many months and can change an athlete's career trajectory. A dataset with every NBA player who has suffered one of the three named injuries in the past twelve seasons was curated (n=113), with information about each player's overall impact on the court before and after injury and some characteristics that may determine post-injury performance. The data was run through various data analysis models, including multiple regression and decision tree classification. This study was the first to use newer statistical metrics such as LEBRON to evaluate player performances, as well as the first to combine the three injuries for direct comparison. The return rates for ACL, Achilles, and meniscus injuries were 85.7%, 96.3%, and 88.6%, respectively. Results indicated that age and average minutes per game pre-injury were the most important factors in determining a player's post-recovery performance. Furthermore, Achilles injuries harmed offensive impact more than ACL or meniscus injuries. However, they were more likely the result of over-usage and, therefore, less likely to end a player's career. Finally, those who previously suffered a similar injury performed significantly worse than those who did not.

Keywords: Anterior cruciate ligament, Achilles tendon, Meniscus, Basketball, Injury, Performance

1. Introduction

The National Basketball Association (NBA) is the professional basketball league of the United States, with players from all over the world competing in a 30-team league. The rigorous schedule of the season, which features an average of one game every three days, can take a massive toll on athletes' bodies, often in the form of injuries. Although many of these injuries permit recovery with a few days rest, some injuries, such as anterior cruciate ligament (ACL) tears, Achilles tendon tears, and meniscus tears, can be devastating and may even alter the course of a promising career. A tear in any of these areas will greatly impair an athlete's agility and speed, ravaging their ability to compete in a physical sport like basketball. The NBA has seen cases like Derrick Rose, a young star voted the league's Most Valuable Player in just his third season reduced to a bench player after multiple injuries, and others like Kevin Durant, who returned from an Achilles injury a year and a half later to play at the same all-star level.

The ACL is a ligament in the knee that runs from the back of the femur to the front of the tibia. It is a strong band of tissue that provides stability in the knee when one pivots or accelerates. Hence, ACL injuries most commonly occur during sports that require rapid changes in pace or direction, jumping, and landing, such as basketball, soccer, skiing, or football. A full tear, and often partial tears, of the ligament require surgical repair and a recovery journey of upwards of eight months (Mayo Foundation for Medical Education and Research, 2022).

The Achilles tendon is the largest tendon in the body, stretching from the bones of the heel to the calf muscles. It allows one to point their toes to the floor or raise up on their tiptoes. Achilles tendon tears are also common in activities that require rapid changes in speed or pivoting. They often occur when one suddenly starts moving, such as surging

Vol. 2023 (2) 106 – 114
ISSN 2688-3651 [online]

off the sprint block in a race. Recovery from an Achilles tear also requires surgery, while the return to sports takes anywhere between six and twelve months (Sachdev, 2021).

The meniscus is a piece of cartilage in the knee that cushions and stabilizes the joint, protecting the bones from wear and tear. Meniscus tears can happen on their own but often accompany ACL tears. The outer region of the meniscus has a good blood supply and can often heal on its own, but tears in the inner two-thirds usually require surgical repair. Still, full recovery from meniscus tears takes six to eight weeks (Wheeler, 2022).

Although injuries to the ACL and Achilles are much more devastating, the meniscus is also crucial to knee function and can cause players to miss a significant amount of playing time, usually around 8-10 weeks. The objective of this investigation is to see if there is a correlation between certain characteristics of an athlete and their performance after recovering from these three injuries using both simple calculations and machine learning models. Some hypotheses include age and previous injury having significant impacts, as older players or those who have already suffered a similar injury are less likely to recover fully. Achilles and ACL injuries are likely more severe than meniscus ones, especially for offensive play, as they require much longer rehabilitation periods and are more crucial to leg stability. The results in this study were supported with statistical methods and corroborated with previous similar studies to increase confidence in conclusions.

2. Dataset

The dataset used in this study was extracted from a larger dataset recording every injury in the NBA from the 2010 to 2020 seasons. It is available to the public on Kaggle (https://www.kaggle.com/datasets/ghopkins/nba-injuries-2010-2018). This set was filtered through to find every instance of a torn ACL, torn Achilles tendon, or a torn meniscus, and their respective players were added to a personal dataset (https://github.com/derekgao2/NBA-Injuries.git) with the following features:

- *Player* is the last name, and sometimes first initial, of the player that was injured.
- *Injury type* describes the injury with an integer: 0 for torn ACL, 1 for torn Achilles, and 2 for a torn meniscus. If a player tore both their ACL and meniscus, they were given a 0, as the ACL is the more significant injury.
- *Season* is the season that the injury occurred. For example, if a player was injured in the 2015-16 season, the year 2016 is written.
- *Age* indicates the age at which the athlete was injured
- *Height* indicates the height of the player in inches, to the nearest inch
- *Weight* indicates the weight of the player in pounds, to the nearest pound
- *Previous injury* describes the player's injury history with an integer: 0 if they have never suffered one of the three injuries at a previous moment in their careers, 1 if they have.
- *Avg minutes* indicates the average amount of minutes that player played per game in the season prior to injury.
- *LEBRON diff* is the difference in the average on the player's LEBRON score of the two seasons before injury and the two seasons post-recovery. If a player did not play two seasons before or after the injury, only one was used. Scores of -10 in this column indicate that the player did not return to play in the NBA after the injury. This usually meant that the player was never able to fully recover to compete at the NBA level again.
- *BPM diff* is the difference in the average on the player's BPM score of the two seasons before injury and the two seasons post-recovery. Again, if a player did not play two seasons before or after the injury, only one was used, and scores of -10 in this column indicate that the player did not return to play in the NBA after the injury.
- *O-LEBRON diff* is the difference in the average on the player's O-LEBRON score of the two seasons before injury and the two seasons post-recovery. If a player did not play two seasons before or after the injury, only one was used. Scores of -5 in this column indicate that the player did not return to play in the NBA after the injury.
- *D-LEBRON diff* is the difference in the average on the player's D-LEBRON score of the two seasons before injury and the two seasons post-recovery. Again, if a player did not play two seasons before or after the injury, only one was used, and scores of -5 in this column indicate that the player did not return to play in the

Vol. 2023 (2) 106 – 114
ISSN 2688-3651 [online]

NBA after the injury.

- *Change* describes the impact of the injury on the player's performance with one of four integers, determined by looking at the LEBRON diff and BPM diff columns: 1 if their performance significantly decreased, 2 if their performance dropped a medium amount, 3 if there was minimal to no change, and 4 if they returned and improved.
- *O-change* describes the impact of the injury on the player's offensive performance with one of four integers, determined by looking at the O-LEBRON diff column: 1 if their performance significantly decreased, 2 if their performance dropped a medium amount, 3 if there was minimal to no change, and 4 if they returned and improved
- *D-change* describes the impact of the injury on the player's defensive performance with one of four integers, determined by looking at the D-LEBRON diff column: 1 if their performance significantly decreased, 2 if their performance dropped a medium amount, 3 if there was minimal to no change, and 4 if they returned and improved.

In total, data was collected on 113 players, with 42 ACL injuries, 27 Achilles injuries, and 44 meniscus injuries. A total of twelve players never returned to the NBA, with six, one, and five suffering ACL, Achilles, and meniscus injuries, respectively.

One thing to note is that current defensive metrics for the NBA are generally poor, as defensive impact is much harder to measure than offensive. Most analytics teams do not trust them, and thus the values and results stemming from this column are less significant and much more subject to chance or human error. All continuous data was scaled and normalized before being run through the regression models.

3. Methods

Two statistical metrics were used in evaluating player performances. The first was a Luck-adjusted player Estimate using a Box prior Regularized On-off (LEBRON), which evaluates a player's total impact per 100 possessions. Each player's estimated impact is stabilized by adjusting for their offensive role. For example, three-point shots will not be weighted the same for sharpshooters who specialize in these shots, versus centers who set screens for their teammates to create open shots (Medvedovsky, 2022). It is then adjusted for luck, as team performances are not completely dependent on a single player. For example, a star athlete cannot control their opponent's three-point percentage or whether or not a teammate misses a free throw. Both of these can help or hurt their overall measured impacts when they are on the floor (Goldstein, 2018). LEBRON is one of the most accurate statistical metrics currently available due to the wide variety of factors it considers for each player. Its holistic evaluation of each athlete was very relevant to the research question because of its ability to precisely capture athletes' post-recovery performances within a single number.

LEBRON is also split into two categories: O-LEBRON, which measures offensive impact per 100 possessions, and D-LEBRON, which measures defensive impact per 100 possessions. The LEBRON score of a player is the sum of their O-LEBRON and D-LEBRON scores. Each player is given a single value in these three metrics for each season. (Narsu & McBasketball, 2022). The dataset can be found at https://www.bball-index.com/lebron-database/.

The other metric used was Box Plus-Minus (BPM), which measures a player's quality and contribution to the team from play-by-play regression with a value from -10 to 10. It uses a player's box score information, position, and the team's overall performance to estimate the player's contribution in points above the league average per 100 possessions played. It starts by assuming that every player on the team has contributed equally, then factors in all of the box score data relative to the other players on the team. Players are given a single value for each season (Myers, 2020). The dataset can be found at https://www.basketball-reference.com/leagues/NBA_2023_advanced.html. Although BPM may not be as accurate as LEBRON, it is still considered a strong statistical metric in evaluating athlete performances and is useful to this research because it provides a second value to corroborate with the LEBRON values. Furthermore, it compares a player to the league average, so it is easier to see the impact of an injury to a player's post-recovery performance.

Vol. 2023 (2) 106 – 114
ISSN 2688-3651 [online]

The regression models used were multiple regression (MR) and decision tree classification (DTC). These methods were used to find a correlation between the player characteristics and their post-recovery performances, indicated in the change, o-change, and d-change columns. The two-sample t-test was also used to corroborate the results. Finally, simpler calculations included averaging the values in the LEBRON diff, BPM diff, O-LEBRON diff, and D-LEBRON diff columns and comparing them based on the type of injury and injury history.

3.1 Multiple Regression

One limitation of linear regression, a common method used in searching for correlations in data, is that it can only look at one independent variable at a time to predict the dependent variable. In this case, many independent variables may work together to determine the dependent variable, so multiple regression was used. Consider the formula of MR:

$$y = a_1 n_1 + a_2 n_2 + a_3 n_3 + \cdots + a_i n_i + b$$

Where y is the calculated dependent variable, $n_1, n_2, . \ n_i$ are the i independent variables, and $a_1, a_2, . . a_i$ are the calculated coefficients for the different independent variables. Given specific values for each of the independent variables, plugging them into the equation should give the most likely change in performance. A percentage of the dataset was used to train the model, and the rest was used to test it. The R^2 score measures the model's validity using the testing data points.

3.2 Decision Tree Classification

The decision tree model is based on a tree-like structure and aims to predict the class of a target variable, in this case, the value in the change, o-change, or d-change columns. The root node, or the uppermost node, represents the entire sample of data. As the model progresses, nodes are split into two or more sub-nodes as the data is categorized and decisions are made. Nodes that do not split are called leaf or terminal nodes, representing the classes of the target variable. Decision trees are advantageous because they provide a non-linear model, much like multiple regression. However, it forces consideration of all possible outcomes of a decision and traces each path to a conclusion, allowing for much more thorough evaluation. A visualization of a decision tree in XGBoost, a gradient-boosted decision tree machine learning library, is shown in Figure 1.

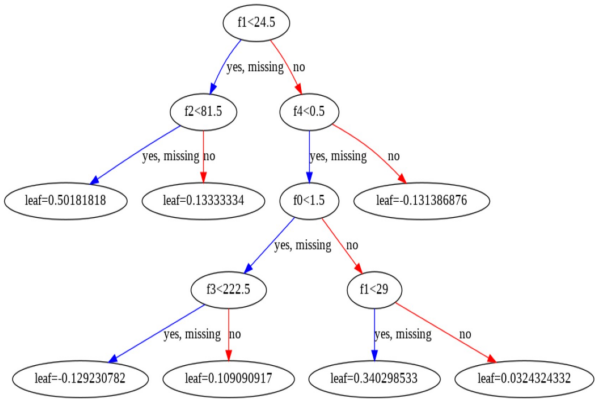


Figure 1. XGBoost decision Tree Structure

3.3 Student's T-Test

The two-sample t-test is a method used to test how significant the difference between two population means is. This is useful when attempting to determine whether the difference between two groups of data is statistically significant or simply due to random chance. It always starts with the null hypothesis that the two population means are equal and looks to reject it. The equation for the t-value is:

$$\frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Vol. 2023 (2) 106 – 114
ISSN 2688-3651 [online]

Where $\overline{x_1}$ and $\overline{x_2}$ are the corresponding means of the populations, n_1 and n_2 are their population sizes, and s_1 and s_2 are their standard deviations. Note that this calculation is only used when the variances of the two populations are not equal. If the variances are equal, t would be calculated as:

$$\frac{\overline{x_1}-\overline{x_2}}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad \text{where } s_p = \frac{\left((n_1-1)s_1^2\right)+\left((n_2-1)s_2^2\right)}{n_1+n_2-2}$$

Along with the student's t-test, a corresponding *p*-value is calculated. This value represents the probability that one can reject the null hypothesis. The smaller the *p*-value, the greater the probability that the difference between the two datasets is statistically significant. A *p*-value below 0.05 indicates that one can be 95% confident that the difference seen in the two groups is due to correlation and not due random chance, and is often the threshold to determining statistical significance. Thus, the calculated *p*-values that fit this criterion were recorded in a table.

4. Results

The return rates for ACL, Achilles, and meniscus injuries were 85.7%, 96.3%, and 88.6%, respectively. The average ages of a player at the time of injury were 25 years 10 months, 28 years 7 months, and 26 years 9 months for ACL, Achilles, and meniscus injuries, respectively. On average, a player's LEBRON score decreased by 0.412, 0.969, and 0.458 for ACL, Achilles, and meniscus injuries, respectively, and a player's BPM score decreased by 0.865, 1.573, and 0.477, respectively (Figure 2). Figure 3 shows the average decrease in O-LEBRON and D-LEBRON scores for each injury. Players were also split based on whether or not they had been previously injured, and the average decrease in all four metrics was calculated (Figure 4).

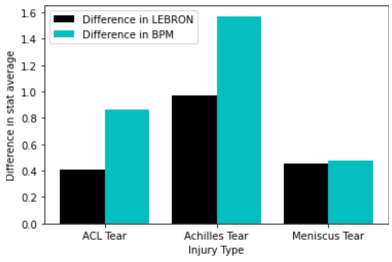


Figure 2. Average Decrease in Performance by Injury

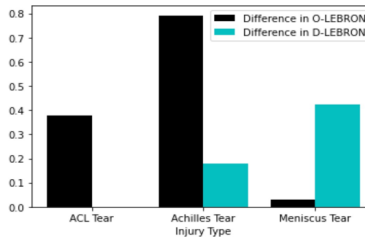


Figure 3. Average Decrease in Offensive and Defensive Performance by Injury

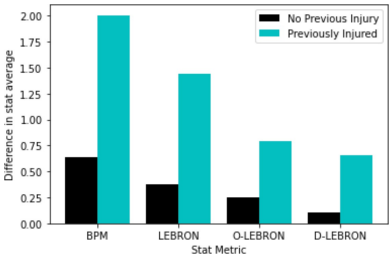


Figure 4. Average Decrease in Performance based on Injury History

Note that for all of the above results, if a player did not return to the NBA after the injury, they were not included in the calculations.

Next, each independent variable was graphed with each metric. The only graphs in which a general downward trend was noticeable were the ones that graphed athlete's ages. The older an athlete at the time of injury, the worse their performance post-recovery (Figure 5). Still, there were players of all ages that never returned to play, as indicated by the range of data points at the bottom of the graph. Furthermore, this trend was only noticed in three of the metrics. Age did not seem to have an impact on an athlete's defensive ability post-injury. (Figure 6). Finally, the majority of players evaluated averaged over 25 minutes per game. However, all 12 players who never returned to play averaged below 25 minutes per game in the season prior to injury (Figure 7).

Vol. 2023 (2) 106 – 114
ISSN 2688-3651 [online]

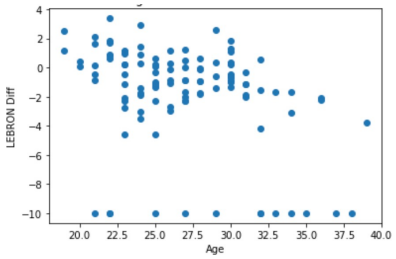


Figure 5. Age vs. Difference in LEBRON

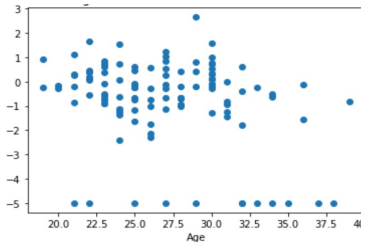


Figure 6. Age vs. Difference in D-LEBRON

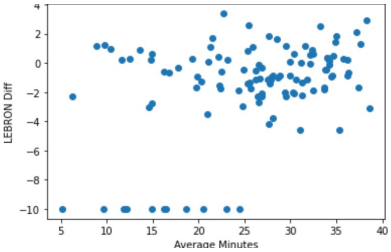


Figure 7. Average Minutes vs. Difference in LEBRON

4.1 Student's T-Test

Any player who did not return to play in the NBA was again removed before performing any calculations in this section. A knee injury combines results from ACL and meniscus injuries. Table 1 shows the most significant *p*-values resulting from two-sample t-tests. For example, in row 5, players who had suffered previous injuries performed much worse after recovering when compared to those who had not been previously injured (-1.44 compared to -0.37). The low corresponding *p*-value indicates that one can be confident this difference is in fact due to their injury history and not random chance.

Table 1. Arrays of Data Compared and Corresponding *p*-Values

Independent Variable 1	**Independent Variable 2**	**Dependent Variable**	**Independent Variable 1 Avg ± SD**	**Independent Variable 2 Avg ± SD**	***p*-value**
ACL Injury	Achilles Injury	Age	25.857 ± 4.45	28.593 ± 3.22	0.00026
Meniscus Injury	Achilles Injury	Age	26.773 ± 4.58	28.593 ± 3.22	0.02226
Meniscus Injury	Achilles Injury	Offensive LEBRON Difference	-0.029 ± 1.26	-0.792 ± 0.98	0.01127
Knee Injury	Achilles Injury	Offensive LEBRON Difference	-0.197 ± 1.30	-0.792 ± 0.98	0.03554
Previous Injury	No Previous Injury	LEBRON Difference	-1.444 ± 1.57	-0.371 ± 1.56	0.00831
Previous Injury	No Previous Injury	BPM Difference	-2.005 ± 2.60	-0.641 ± 2.57	0.04160
Previous Injury	No Previous Injury	Defensive LEBRON Difference	-0.661 ± 0.90	-0.106 ± 0.86	0.01400

4.2 Multiple Regression

Figure 8 shows the predicted values of BPM differences for the data used to test the model graphed with their actual values. The R^2 score for this model was 0.349.

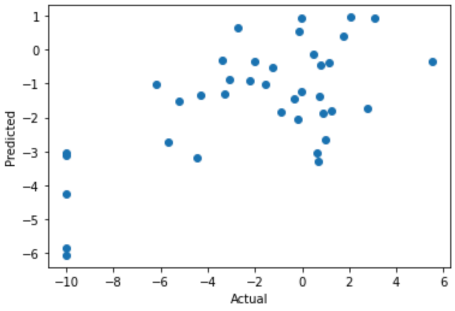


Figure 8. Multiple Regression Model for BPM Differences: Predicted vs. Actual Values

4.3 Decision Tree Classification

A decision tree model was run three separate times to predict the values in the change, o-change, and d-change columns, receiving accuracy scores of 0.462, 0.447, and 0.308, respectively. The graphs indicating the relative importance of each independent factor for each model are shown in figures 9, 10, and 11.

Vol. 2023 (2) 106 – 114
ISSN 2688-3651 [online]

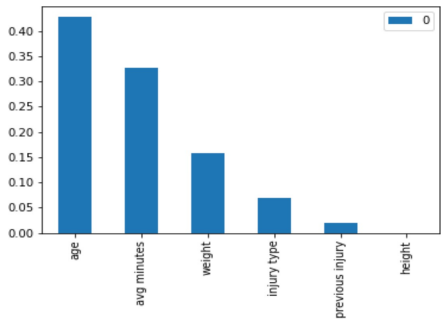


Figure 9. Relative importance of each independent factor in determining overall impact post-recovery

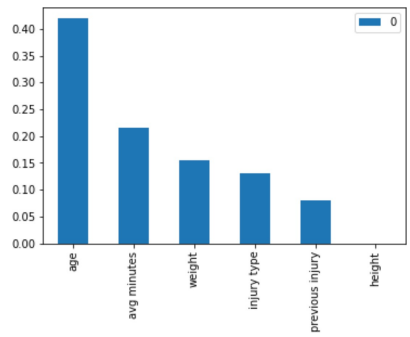


Figure 10. Relative importance of each independent factor in determining offensive impact post-recovery

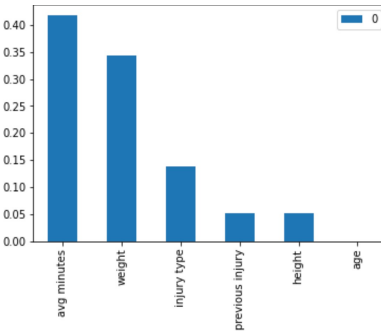


Figure 11. Relative importance of each independent factor in determining defensive impact post-recovery

Here is the confusion matrix for the overall change decision tree model:

$$\begin{bmatrix} 5 & 0 & 0 & 2 \\ 1 & 9 & 1 & 3 \\ 1 & 8 & 0 & 0 \\ 0 & 4 & 1 & 4 \end{bmatrix}$$

The sum of the numbers in the *i*th row is the total number of data points classified by *i*, while the number in the *j*th column of the *i*th row is the number of data points the model predicted to be *j* but was actually *i*. For example, there were 14 players classified by 2, meaning their performance dropped a medium amount, and the model predicted 9 of these players correctly. Here is the confusion matrix for the offensive change decision tree model:

$$\begin{bmatrix} 4 & 4 & 0 & 4 \\ 1 & 9 & 0 & 3 \\ 1 & 2 & 0 & 1 \\ 1 & 4 & 0 & 4 \end{bmatrix}$$

5. Discussion

5.1 Interpretations

Some of the results from this study are supported by common intuition. For example, age was a significant factor in determining a player's post-recovery performance. The older the player, the less of an impact they had on the floor, which makes sense because as a player ages, they can no longer play with the same intensity and aggressiveness as they could have if they were five or ten years younger (Figure 5). An injury only accelerates this decline. Furthermore, players who had previously suffered a similar injury performed significantly worse in almost every metric than those who did not (Table 1 & Figure 4), as second major surgery can take a serious toll on an athlete's body, especially if they are a couple of years older at the time of the repeat injury.

All three injuries caused decreases in overall, offensive, and defensive impact (Figures 2 & 3). Furthermore, an Achilles tear caused the worst decrease in offensive LEBRON, as indicated by Table 1, while the differences among the injuries in overall and defensive performances were statistically insignificant. This result corroborates with the findings of Benjamin Kester (2016) and Lafi Khalil (2020) who both concluded that ACL and Achilles injuries significantly harm player performances post-recovery. Taken together, these three studies support the conclusion that the Achilles and ACL tissues are very crucial to a player's impact on the court, as speed, agility, and explosiveness are all crucial to the sport. On the other hand, the data was swapped for the meniscus (Figure 3), with more damage being done to a player's defensive impact but little change to offensive impact. One possible explanation is that more jumping occurs during defense as players go up to contest shots or layups, and the meniscus acts as a shock absorber during

Vol. 2023 (2) 106 – 114
ISSN 2688-3651 [online]

hard landings.

One interesting result was that all twelve players who did not return to play averaged 25 minutes per game or less, while the majority of the other players averaged more (Figure 7). This result could be because players who average more time on the court are the ones who are capable of playing at a higher level and thus can take a setback to their careers, while bench players who already have limited roles will be cut without a second thought. Furthermore, six of these players suffered an ACL injury, five suffered a meniscus injury, but only one suffered a torn Achilles. These numbers suggest that although a torn Achilles has a larger impact on post-recovery performance, it is unlikely to end a career. Using the above logic, it is possible that an Achilles injury only occurs when a player reaches a decently high level of play and averages a significant number of minutes per night, meaning that the injury might result from wear and tear rather than a single freak accident. This conclusion is further supported by the extremely low *p*-values in the t-test results, which showed that the difference between the average ages of players who suffered each of the three injuries is statistically significant. Those who suffered Achilles injuries were on average three years older than those who suffered ACL injuries and two years older than those who suffered meniscus injuries, meaning that an Achilles tear usually occurs later into an athlete's career and thus is more likely the result of overuse. On the other hand, ACL and meniscus injuries occur earlier, which could be because these types of injuries occur much more suddenly.

The R^2 score for the multiple regression model of prediction was relatively low (Figure 8), indicating a poor and unreliable model for this dataset. The decision tree models scored a little bit better, however. It was trained to place each player into four classes, so a randomly guessing model would have had an accuracy score of around 25%. The models predicting change and offensive change were closer to 45%, meaning that although not ideal, they were much better than a random guess. Furthermore, many of the predictions were close, as seen in the decision matrix for the decision tree that predicted overall change. Since classes 2 and 3 were the most similar, with the former being medium change and the latter being little change, it makes sense that the model may have mixed them up. In both the overall and offensive change models, the order of the relative importance of each independent factor was the same. Age and average minutes were the two most important, which aligns with the results above (Figures 9, 10, & 11).

5.2 Limitations

The major limitation of this study was the paucity of data. ACL, Achilles, and meniscus injuries are not the most common types of injuries in the NBA, leaving the dataset used in this study to only include 113 players. With a smaller dataset, the models did not have as much information to create the ideal predictions, thus leading to the lower accuracy scores. Furthermore, LEBRON scores have only been calculated for players since 2010, meaning that any injuries prior to the 2009-10 season could not have been used in this study.

It is also possible that players' performances correlate more with their rehabilitation process rather than their physical traits like height or weight. Numbers such as hours spent in physical therapy or time between different stages of recovery may have been able to produce more accurate predictions in the decision tree models. Unfortunately, data on an individual athlete's path to recovery is not available for public use. Hopefully, future studies will be able to incorporate such data.

5.3 Next Steps

As stated above, there is data on NBA injuries that goes farther back than the 2009-10 season, but a different statistic than LEBRON would have to be used. Most of the current advanced statistics that evaluate a player's overall impact on the court are relatively new and likely don't go much further into the past either. Still, this study could be applied to more simple statistics like field goal percentage or points/rebounds/assists per game.

Other paths could include extending this analysis to the National Collegiate Athletic Association (NCAA), the college sports league that hosts the famous March Madness tournament every year. The same investigation could even be extended to other significant injuries in other sports, such as ulnar collateral ligament (UCL) injuries in the elbow joint for baseball pitchers.

Vol. 2023 (2) 106 – 114
ISSN 2688-3651 [online]

6 Conclusion

This study offered insight into the relative severities of ACL, Achilles, and meniscus tears and how they may impact an NBA athlete's post-recovery performance in spite of high return rates. Older age and injury history played definite roles in decreasing the LEBRON and BPM scores of an athlete, while averaging greater minutes per game may prevent one's career from coming to a brutal end. One crucial hypothesis derived from this study is that Achilles injuries are more harmful but more likely to be the consequence of over-usage than ACL ones, which can result from a single misplaced pivot foot. Additional research should be completed in the future as data accumulates, allowing for more generalized results and, therefore, a greater chance to predict and prevent these injuries.

Acknowledgment

I would like to thank Dr. Ganesh Mani and Dr. Michael Powell for their guidance in this study. Without their help, this project would not have been complete.

References

Goldstein, J. (2018, January 8). Nylon calculus: Defining and calculating luck-adjusted ratings for the NBA. *FanSided*. https://fansided.com/2018/01/08/nylon-calculus-calculating-luck-adjusted-ratings/

Hopkins, R. (2021). NBA Injuries from 2010-2020 (Version 3). *Kaggle*. https://www.kaggle.com/datasets/gho pkins/nba-injuries-2010-2018

Kester, B.S., *et al.* (2016). Athletic performance and career longevity following anterior cruciate ligament reconstruction in the National Basketball. Association. *Knee Surg Sports Traumatol Arthrosc* **25**, 3031–3037 (2017). https://doi.org/10.1007/s00167-016-4060-y

Khalil, L. S., *et al*. (2020). Effect of Achilles Tendon Rupture on Player Performance and Longevity in National Basketball Association Players. *Orthopaedic journal of sports medicine*, *8*(11), 2325967120966041. https://doi.org/10.1177/2325967120966041

Mayo Foundation for Medical Education and Research. (2022, November 30). ACL reconstruction. *Mayo Clinic*. https://www.mayoclinic.org/tests-procedures/acl-reconstruction/about/pac-20384598

Medvedovsky, K. (2022, August 19). Offensive Roles. *Basketball Index*. https://www.bball-index.com/offensive- archetypes/

Myers, D. (2020, February). About Box Plus/Minus (BPM). *Basketball Reference*. https://www.basketball-reference.com/about/bpm2.html

Narsu, K., & McBasketball, C. (2022, May 24). Lebron Introduction. *Basketball Index*. https://www.bball- index.com/lebron-introduction/

Sachdev, P. (2021, October 18). Achilles tendon injuries (tear, rupture) causes, symptoms, treatments. *WebMD*. https://www.webmd.com/fitness-exercise/guide/achilles-tendon-injury

Wheeler, T. (2022, January 26). Meniscus tear in knee: Symptoms, causes*, and treatments. WebMD*. https://www.webmd.com/pain-management/knee-pain/meniscus-tear-injury

Journal of Research High School

Vol. 2023 (2) 115 – 121
ISSN 2688-3651 [online]

Enhancing American Football Analytics: Classifying Play Videos as Run or Pass using Deep Learning

Kyle Zhou[1] *

[1]Sunset High School, Portland, OR, USA
*Corresponding Author: kyle.yiheng.zhou@gmail.com

Advisor: Jason Galbraith, Jason_Galbraith@beaverton.k12.or.us

Received March 14, 2023; Revised July 12, 2023; Accepted, August 7, 2023

Abstract

High school football coaches often rely on reviewing past season game footage to develop winning strategies. Identifying recurring patterns in the opposing team's offensive plays through video analysis helps coaches plan effective defensive tactics. However, accurately classifying play types from video clips using automated methods remains challenging and obtaining a sizable dataset of labeled plays for training models is difficult. This paper proposes a machine learning-based approach that utilizes the MoViNets model for action recognition. To overcome the challenge of limited labeled video clips, this paper utilized transfer learning to fine-tune MoViNets models that are pre-trained using a large dataset Kinetics-600. Extensive experiments were conducted to determine the optimal sampling scheme for the videos and compared the performance of two variants of MoViNets, the smaller A0 model and the scaled-up A3 model. The results show that using 24 frames sampled at 2 frames per second yields the best classification performance, and the A3 model achieves 81% accuracy on the test dataset, outperforming the A0 model's 74%. Ongoing data collection is expected to further improve the model's accuracy, potentially enabling automated play type classification.

Keywords: Machine learning, Deep learning, Convolutional Neural Network (CNN), Video classification, Transfer learning

1. Introduction

Applying machine learning in sports analytics has gained significant momentum in recent years. However, due to the vast diversity of game plays, extracting insights from American football remains a challenging task. In American football, the game is played on a 120-yard-long field with a width of 53 ⅓ yards. It involves two teams of 11 players each, one on offense and the other on defense. The offense executes plays, either by running or passing the ball, aiming to gain yards and score. Running plays involve the quarterback handing off the ball to a running back, while passing plays involve the quarterback throwing the ball to a receiver. The objective is to move the ball towards the end zone and score points by either kicking a field goal or advancing into the end zone. The offense has four downs (chances) to move the ball at least 10 yards, and if unsuccessful, the possession of the ball changes. Game film analysis is a common practice among high school football coaches, as it allows them to gain valuable insights into their opponents' strengths and weaknesses and develop effective strategic plans. A crucial aspect of this analysis is studying the offensive play types of the opposing team, including their tendency to pass or rush the ball. Manual annotation of past play video clips is often tedious and prone to errors. Recent advancements in computer vision and video understanding have made it possible to automatically classify play types from recorded video clips.

In previous studies, such as the one conducted by (Chen et al., 2014), authors utilized several processing techniques such as noisy detector, Hidden Markov Model, and KLT tracking to identify five different play types

Vol. 2023 (2) 115 – 121
ISSN 2688-3651 [online]

(Offense, Defense, Kick off, Punting, and Field goal) with a 77% accuracy rate. Additionally, another paper (Fernandes et al., 2019) predicted whether the play was a pass, or a run based on extracted play features such as the current down, yards to go for the first down, and point differential. They compared four different models (CART, KNN, Random Forest, and neural network) and found that the neural network achieved the highest accuracy of 75.3%. More recently, (Liu et al., 2022) proposed a deep learning-based pipeline for player tracking in videos. Their end-to-end system integrates an object detection network, detection transformer, and convolutional neural networks (CNNs) for player identification and time tracking.

Deep learning models have proven to be more effective than traditional methods for recognizing intricate and nuanced patterns in video data. Among these techniques, 3D CNNs have demonstrated exceptional performance and are considered state-of-the-art for video classification tasks (Tran et al., 2018). However, the memory and computational demands of 3D CNNs can be prohibitively high, which poses challenges for their deployment. To address this issue, a family of mobile computer vision models called MoViNets has been developed (Kondratyuk et al., 2021). These models are specifically designed to overcome the limitations of 3D CNNs in terms of memory and computation, while still maintaining high accuracy.

The objective of this research is to explore the effectiveness of multiple MoViNets models in automating the classification of football play types. The study focuses on analyzing video clips obtained from 10 matches played by Sunset High School's football team during the 2022 season, specifically from September 2022 to November 2022. These video clips consist of 1039 play clips, each labeled as either a run or a pass. Given the limited amount of available data, the research employs transfer learning techniques as described by Hosna et al. (2022). Transfer learning involves leveraging pre-trained deep learning models that have been trained on extensive datasets. By utilizing these pre-existing models, the performance of the MoViNets models on smaller datasets can be enhanced.

The primary goal of the research is to achieve a reasonably high classification accuracy of at least 80%. This level of accuracy would enable the automation of the manual process involved in tagging play clips with their respective run or pass labels. This system holds enormous potential to aid coaching staff of all levels by saving them long hours spent tagging and allowing them to focus on more crucial tasks.

Initially, MoViNets base model A0 was used to conduct experiments to analyze the classification impact of video clip duration and sampling rate. The scheme that sampled a total of 24 frames at a rate of 2 frames per second outperformed other schemes, achieving a classification accuracy of 74% on the test dataset. As anticipated, the test accuracy increased to 81% by simply employing a scaled-up MoViNets A3 model. Future work includes using game clips from previous seasons, which is expected to further enhance the model's performance.

2. Data and Preprocessing

During the past football season, the Sunset High School football team played 10 games, and for each match, separate video clips were recorded and analyzed using the Hudl Assist Service (Hudl Service, n.d.). In total, there were 1226 plays, with most of them being labeled as either Run (534) or Pass (505). The footage was filmed from the press box of the stadium with a singular position camera. Each play clip was then cut from the start of the play to the end before being posted to Hudl. This study will only focus on distinguishing between Run and Pass plays by training video classification models.

Since the videos' length ranges from 12 seconds to 28 seconds and are recorded at 30 frames per second, the video pre-processing step involves selecting frames from the clip to be used as input to the model. This preparation pipeline is depicted in Figure 1. The pipeline includes frame sampling, image resizing, normalization, optional augmentation, and dataset formation and train/validation/test split. Frame sampling involves selecting a subset of frames from the video based on a specific sampling rate. Image resizing transforms the selected frames to a specific size to ensure consistent input for the model. Optional augmentation can be applied to further increase the diversity of the dataset and improve the model's performance. Finally, the dataset is formed and split into training, validation, and test sets to evaluate the model's performance.

The accuracy of the classification is significantly impacted by two key hyperparameters: the number of frames chosen, and the sampling rate used. Figure 1 also shows four sampling schemes evaluated in this study.

Vol. 2023 (2) 115 – 121
ISSN 2688-3651 [online]

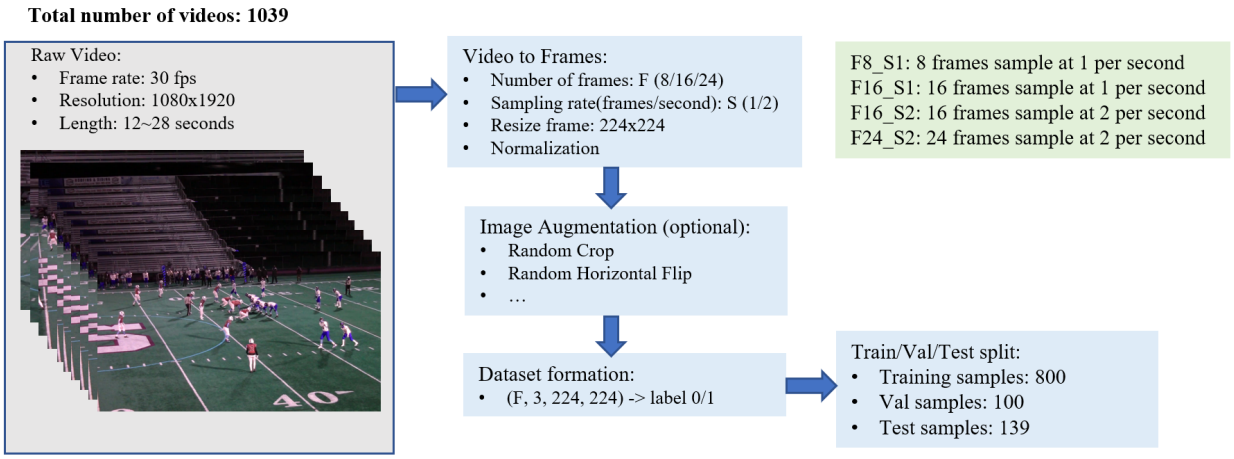


Figure 1. Video preprocessing pipeline and four frame sampling schemes.

3. Method

3.1 Video Classification Model

There are several types of video classification models, which can be broadly categorized as follows:

1. 2D CNN: These models process each frame of the video as an image using 2D CNNs and then combine the predictions from each frame to classify the video.
2. 3D CNN: These models process the video frames as a sequence and use 3D CNNs to learn spatiotemporal features which helps to capture the motion information in the video.
3. RNN: Use recurrent neural networks (RNNs) to model the temporal dependencies between frames.
4. Transformer-based models: These models use attention mechanisms to selectively focus on certain parts of the video frames, enabling them to better capture salient information.

Each type of video classification model has its own advantages and disadvantages, and the choice of model will depend on the specific task and data at hand.

This paper explored one specific video classification model called Mobile Video Networks (MoViNets), which have been designed to be efficient in terms of computation and memory usage based on 3D CNN architecture. As a result, MoViNets are suitable for deployment on mobile devices and support online inference. The MoViNets model family has been optimized using the Neural Architecture Search (NAS) technique to find the most efficient architecture. Furthermore, the introduction of Stream Buffers allows for the processing of videos in small, consecutive sub-clips, ensuring a constant memory requirement without compromising long temporal dependencies.

The MoViNets model architecture includes six different versions labeled A0 through A5, balancing model efficiency and classification accuracy. The A0, A1, and A2 variants are optimized for fast performance and are well-suited for deployment on mobile devices. In contrast, the A3 and other variants have more parameters and can achieve higher levels of accuracy. This study used the A0 model to assess the impact of video sampling rate and number of frames on model performance, as it has a shorter training time. Later, the classification accuracy was compared between the A0 and A3 models to evaluate the benefits of the higher model capacity of the A3 variant.

3.2 Transfer Learning with MoViNets

Transfer learning is a machine learning technique that involves repurposing a pre-existing model trained on a specific task as a foundation for a new, related task. This strategy enables a model to capitalize on the insights acquired while solving a problem, and apply them to a distinct but related problem, even when there is limited labeled data available for the new task. Typically, a pre-trained model is employed as a feature extractor to gather pertinent

Vol. 2023 (2) 115 – 121
ISSN 2688-3651 [online]

information from the input data. This information is then passed on to a new model to solve the new task. The new model can be fine-tuned, retrained, or modified using the extracted features and some labeled data from the new task. Employing this method can save considerable time and resources when compared to training a new model from scratch. Transfer learning has been applied successfully across various domains, including natural language processing, computer vision, and speech recognition.

TensorFlow Hub (Tensorflow Hub, n.d.) provides a range of pre-trained machine learning models that can be used as is or fine-tuned for specific tasks. One of these models is MoViNets, a 3D CNN architecture which has several versions pre-trained on Kinetics-600, a vast dataset of around 495,000 videos across 600 action categories (Carreira et al., 2018). To cater to the specific task of run and pass detection in football videos, a new model was created by taking a pre-trained MoViNets model and freezing its convolutional base. The original classifier head was then replaced with a new one that is tailored to the run and pass labels. Finally, transfer learning was performed by training the new classifier head using the pre-processed football videos.

4 Results

Although MoViNets models are designed to be more computationally and memory-efficient than other architectures, training the model can still be resource-intensive. To run all the experiments in the study, Google Colab (Google Colab, n.d.) was used with backend GPU support. The same training hyperparameters were set for all scenarios to ensure fair comparison, including using the Adam optimizer with a learning rate of 0.0005 and default values for all other parameters. A batch size of eight was used and trained the model for five epochs.

4.1 Sampling rate and frame counts

The MoViNets A0 pre-trained model was used, with its classification head replaced, as the baseline to examine the effect of frame count and sampling rate. Table 1 illustrates four distinct scenarios that were evaluated, varying in terms of total frame count and sampling rate. Covered video duration is the length of the video used in the analysis and is calculated by dividing the number of frames by the sampling rate.

Table 1. Four sampling schemes.

	Number of frames	Sampling rate (frames per second)	Covered Video duration (seconds)
F8_S1	8	1	8
F16_S1	16	1	16
F16_S2	16	2	8
F24_S2	24	2	12

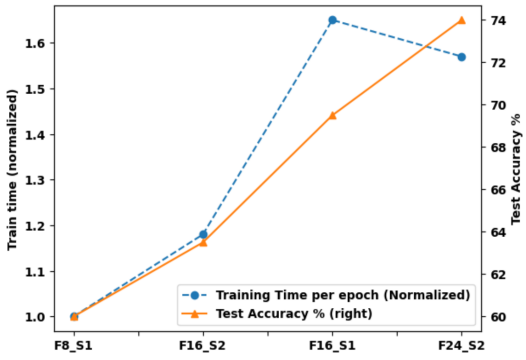


Figure 2. Training time and test accuracy for four sampling schemes.

Figure 2 shows the normalized training time and test accuracy for four different scenarios. Processing more frames leads to longer training time, as demonstrated by the F16_S1 scheme, which takes 65% more time to train each epoch than F8_S1. Higher sampling rates, on the other hand, result in shorter video length for the same number of frames. As a result, the epoch training time for F16_S2 is considerably shorter than that for F16_S1.

Capturing more temporal information by increasing the number of frames and sampling rate leads to higher test accuracy. The model F24_S2 attains the highest accuracy of 74.0%.

The confusion matrix for F16_S1 and F16_S2 is depicted in Figure 3. Both schemes extract 16 frames from the video clip, but with different sampling rates: F16_S1 samples one frame per second, while F16_S2 samples two frames per second.

Vol. 2023 (2) 115 – 121
ISSN 2688-3651 [online]

Two popular metrics used to assess the quality of a classifier's outcomes are precision and recall. Precision is calculated by dividing the true positive by the sum of true positive and false positive, while recall is determined by dividing the true positive by the sum of true positive and false negative. In simple terms, precision measures a classifier's ability to minimize the false positive rate and recall measures how well the classifier identifies all positive examples. In Figure 4, a comparison is presented between the precision and recall of the F16_S1 and F16_S2 sampling schemes. The results show that when sampling two frames per second (F16_S2), the precision for Pass is improved, but the precision for Run is negatively impacted compared to F16_S1 which samples 1 frame per second. Conversely, the recall shows the opposite pattern.

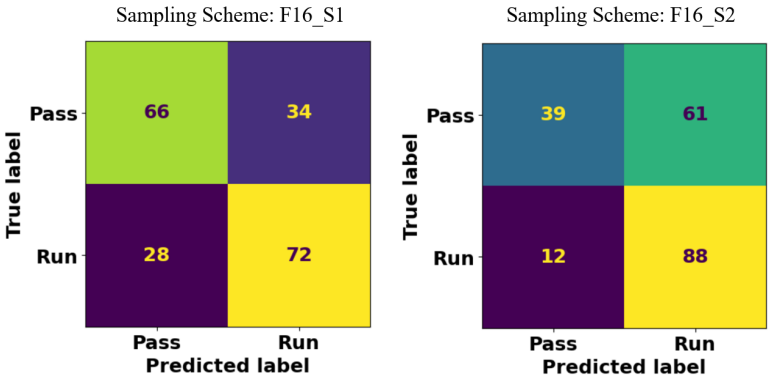


Figure 3. Confusion matrix for scheme F16_S1 and F16_S2.

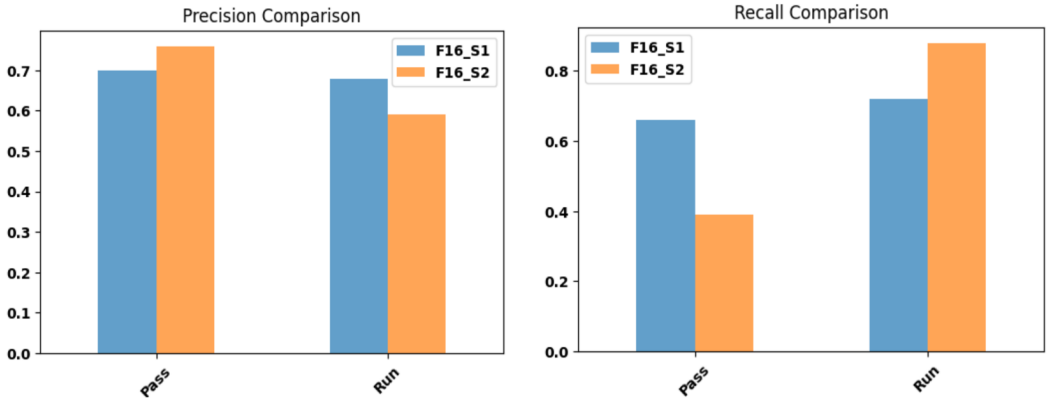


Figure 4. Precision/Recall comparison for schemes F16_S1 and F16_S2.

4.2 MoViNets A0 vs. A3

MoViNets offers a variety of models with ascending capacities, with A0 being the most efficient and A5 being the most powerful. Table 2 illustrates that MoViNets A3 has over three times more parameters than the baseline A0. Typically, the pre-trained feature extractor weights remain fixed, and only the classification head is trained. In this case, A0 has 911,583 pre-trained weights, while A3 has 4,688,129.

Table 2. MoViNets A0 vs. A3 parameter comparison.

	MoViNets A0	MoViNets A3
Total Params	1,900,769	6,217,987
Trainable Params	989,186	1,529,858
Non-trainable params	911,583	4,688,129

Generally, increasing the model capacity can enhance performance, provided there is sufficient training data. Utilizing the F24_S2 sampling scheme, the test accuracy of A3 model surged to 81.0%, compared to 74.0% achieved by the A0 model as shown in Table 3.

Table 3. Test accuracy comparison between A0 and A3 both using F24_S2 scheme.

	A0	A3
Test Accuracy (%)	74.0	81.0

Vol. 2023 (2) 115 – 121
ISSN 2688-3651 [online]

5 Discussion

This study conducted an analysis of 1039 plays taken from 10 official games played by Sunset High School's football team against other schools in the state during the 2022 season from September to November. This data set formed the basis of the analysis, and it was used to develop and refine the models. To increase the size of the data set, play clips from previous seasons between teams other than Sunset High School should be added, if possible. Incorporating additional data will improve the coverage of the data distribution and increase the accuracy of the classification algorithms. Furthermore, utilizing larger data sets will enable us to employ more advanced machine learning models, such as the scaled-up MoViNets A4 and A5, which can handle more complex and varied data sets. This should significantly improve the performance of Pass/Run classification.

As part of ongoing research, the aim is to investigate the impact of image resolution on classification accuracy. Currently, during video pre-processing, all sampled images are resized from the original 1080x1920 resolution to 224x224. However, some studies suggest that using higher-resolution images could enhance classification accuracy, albeit at the cost of higher computing resource overhead.

Additionally, the field of video classification is constantly evolving with the emergence of new algorithms. Among these, transformer-based models have gained significant attention due to their promising performance. Specifically, TimeSformer has achieved state-of-the-art results on various benchmarks (Bertasius et al., 2021). This model extends the Transformer architecture to handle spatiotemporal data and has been shown to effectively capture both spatial and temporal features in videos.

6 Conclusion

Machine learning techniques are becoming increasingly important in sports analytics to enhance team performance and gain valuable insights. By analyzing vast amounts of data, such as video footage, machine learning can identify patterns, predict outcomes, and make strategic decisions. In the context of American football, for example, analyzing the play types used by opposing teams during previous games is crucial for developing an effective defense strategy. However, manually labeling play types in video clips can be a tedious and error-prone task.

In recent years, developing effective machine learning models for action recognition in videos has become increasingly important in the field of sports analysis. However, one of the major challenges for American football has been the lack of a large dataset of labeled plays for training the models. This paper proposes a machine learning-based approach that employs the state-of-the-art MoViNets model for action recognition in videos. To overcome the issue of limited data, transfer learning is utilized, where a pre-trained MoViNets model is fine-tuned using around 1000 labeled football video clips. To determine the optimal sampling scheme for the videos, extensive experiments are conducted. The results of these experiments indicate that using 24 frames sampled at a rate of 2 frames per second produces the best classification performance.

Furthermore, the study compares the performance of two different MoViNets models, namely the efficient A0 model and the scaled-up A3 model. The results of this comparison demonstrate that the A3 model achieves higher accuracy, with a score of 81% accuracy on the test dataset. As data collection continues, the accuracy of the model is expected to improve further, potentially enabling automated play type classification.

In summary, the study proposes a hopeful method for classifying Pass/Run in football videos using machine learning. This approach could have substantial implications for game analysis and decision-making. The results of this research also demonstrate the potential for use of transfer learning and the state-of-the-art models such as MoViNets to overcome challenges related to limited data and achieve accurate classification of sports videos.

Acknowledgment

This research was mentored and supported by Sunset High School’s computer science teacher Mr. Jason Galbraith as part of the Artificial Intelligence class. I also want to thank the Sunset High School football program for providing me with labeled game film from the past season.

Vol. 2023 (2) 115 – 121
ISSN 2688-3651 [online]

References

Bertasius, G., Wang, H., & Torresani, L. (2021). Is Space-Time Attention All You Need for Video Understanding? https://doi.org/10.48550/ARXIV.2102.05095

Carreira, J., et al. (2018). A Short Note about Kinetics-600.

Chen, S., et al. (2014). Play type recognition in real-world football video. 2014 IEEE Winter Conference on Applications of Computer Vision, WACV 2014. 652-659. 10.1109/WACV.2014.6836040.

Fernandes, C., et al. (2019). Predicting plays in the National Football League. Journal of Sports Analytics. 6. 1-9. 10.3233/JSA-190348.

Google Colab. (n.d.). https://colab.research.google.com/

Hosna, A., et al. (2022). Transfer learning: a friendly introduction. Journal of Big Data. 9. 10.1186/s40537-022-00652-w.

Hudl Service. (n.d.). https://www.hudl.com/

Kondratyuk, D., et al. (2021). MoViNets: Mobile Video Networks for Efficient Video Recognition. 16015-16025. 10.1109/CVPR46437.2021.01576.

Liu, H., et al. (2022). Deep Learning-based Automatic Player Identification and Logging in American Football Videos.

Tensorflow Hub. (n.d.). Collection of all MoViNet models. https://tfhub.dev/google/collections/movinet

Tran, D., et al. (2018). A closer look at spatiotemporal convolutions for action recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 6450–6459, 2018.

Journal of Research High School

Vol. 2023 (2) 122 – 130
ISSN 2688-3651 [online]

Breakthroughs in Honey Bee Health: Local Summer Weather Humidity Conditions Influence Winter Colony Survival (Part III)

Kaitlyn N. Culbert[1] *

[1]Toms River High School North, Toms River, NJ, USA
*Corresponding Author: culbertkaitlyn@gmail.com

Advisor: Christine Girtain, cgirtain@trschools.com

Received March 20, 2023; Revised July 19, 2023; Accepted, August 23, 2023

Abstract

Honey bees (*Apis mellifera*) are a vital species in our ecosystem. They pollinate 73% of the world's cultivated crops and have an economic impact of $20 billion in the United States and $183 billion dollars globally. Unfortunately, in 2006, beekeepers reported colony losses of 60-90%. Commercial beehives in the US have decreased over 50% in the last 70 years. The greatest single contributor to the decline of honey bee health is the *Varroa destructor* mite. Currently, all commercially available thymol-centered systems are gel-based and work by direct contact with the mite. These systems are also highly dependent on temperature and humidity for effectiveness. An earlier laboratory investigation (Part I) and field study (Part II) examined the use of thymol-based essential oil for miticide efficacy and the effective use of mist diffusers to eliminate the dependence of the essential oils on temperature and humidity conditions. Recognizing the significance and impact of hive humidity to *Varroa* mite survival and honey bee *Varroa* mortality, a linear regression model analysis was developed to investigate the significance of weather-related variables, particularly temperature and humidity, to colony loss. Weather data was collected from the National Oceanic and Atmospheric Administration (NOAA) and the National Centers for Environmental Information. Percent colony loss data was provided by the Bee Informed Partnership from 2009-2021 for the state of New Jersey. To date, this research represents the first and only available study examining local weather-related humidity variables to colony losses. The summer humidity-related weather variables of relative humidity, dewpoint, vapor pressure deficit, and temperature were found to be statistically significant to the percent of winter colony losses.

Keywords: Varroa destructor; Apis mellifera, Mites, Honey bees, Weather, Humidity, Colony Collapse Disorder, Regression analysis

1. Introduction

Honey bees, *Apis mellifera,* provide vital ecosystem services. They serve as crucial pollinators for agriculture, responsible for 73% of all cultivated crops (Randall, 2020). As much as one-third of our food depends on pollinators and bees are fundamental to the world's food security (Ramsey et al., 2017). In fact, bee pollination accounts for approximately US$20 billion in added crop value (USDA, 2021). Additionally, honey bees also produce honey, pollen, royal jelly, beeswax, propolis, and venom for nutritional and medicinal uses for an additional US$300 million annually (Calovi et al., 2021). In the United States, there are approximately 2.5 million commercially farmed honey bee hives and around 500,000 colonies kept by hobbyists and semi-professional bee keepers (Penn State Extension, 2013). Unfortunately, these numbers are declining at a rapid rate. Losses are attributed to Colony Collapse Disorder which occurs when there is a sudden loss of a colony's worker bee population yet the queen, brood and a relatively abundant amount of honey and pollen reserves remain. Various reports have suggested losses between 30% to 50% of winter bee colonies in the US, its lowest point in the past 50 years (EPA, 2021). Annual fluctuations in winter colony losses

Vol. 2023 (2) 122 – 130
ISSN 2688-3651 [online]

have been linked to pests and diseases, bee management, including bee keeping practices and breeding, agricultural practices, pesticide use and the change in climatic conditions. Regional factors including honey flow dynamics and the interplay of weather conditions, topographic variables affecting temperature and moisture, and the composition of the surrounding landscape have also been examined (Johannesen et al., 2022; Calovi et al., 2021)

During winter months, colonies rely on existing honey stores, cease foraging for nectar and pollen, and halt brood rearing. Winter bees live several months while summer worker bees live for only a few weeks. Honey bee colonies; however, do not remain dormant during the winter and remain active to maintain the hive temperature between 24-34 degrees Celsius (75.2-93.2 degrees Fahrenheit) by forming a thermoregulating cluster (Calvoli et al., 2021). The ability to form a thermoregulating cluster enables them to survive sustained periods of cold temperature. A critical component to increasing the lifespan of winter bees is the control of *Varroa destructor* mites. A single *Varroa* mite can shorten the lifespan of a bee by one-third, and two mites can shorten it by one-half (Bryant, 2006). *Varroa* is an ectoparasitic mite that feeds on the fat bodies of developing honey bee larvae and adult bees and aggressively reproduces within an infected bee colony. Recent research brings to light the *Varroa* mite's focus on the fat body tissue (Ramsey et al., 2019). Because *Varroa* weakens and ultimately kills colonies by out-reproducing their host, bee keepers initially used acaricides, pyrethroids, and organophosphates pesticides (Bahreni et al., 2020). The frequent use of these synthetic miticides resulted in the development of resistance to many of the chemical components of these miticides (Bahreni et al., 2020; Traynor et al., 2016).

Essential oils are an alternative to chemical pesticides. They are cheaper, environmental-friendly, and pose fewer risks to the health of bees and consumers. Most importantly, *Varroa* have not developed resistance to essential oils for honey bee mite control (Ghasemi et al., 2011; Damiani et al., 2009). Currently, numerous essential oil compounds have been evaluated for miticidal activity. One of the proven successful essential oils is thymol. It works by disorienting the mite and blocking its pores (Tennessee's Honey Bees, 2021). Thymol is the only compound of essential oils widely used in beekeeping with 70%-90% efficacy against *Varroa* (Garrido, 2018). The most widely used beekeeping products with thymol are Apiguard®, ApiLifeVar® and Thymovar® (Garrido, 2018). None of these systems utilize thymol to reach reproducing *Varroa* mites in the brood cell. These systems only kill the mites on the adult bees (Garrido, 2018). All commercially available thymol-based systems are gel-based and its effectiveness is highly dependent upon the ambient temperature and relative humidity within the hive. Temperature and humidity affect the rate of essential oil evaporation (Sabahi, 2017) and since these systems are gel-based, they are only effective when there is direct contact with the mite (Garrido, 2018).

Following a laboratory investigation (Part I, Culbert, 2022) demonstrating the potential utility of battery-operated mist diffusion of thymol-based essential oils, a field study (Part II, Culbert, 2023) was conducted in *A. mellifera* colonies with bee hives naturally infested by *Varroa destructor* mites. The field study demonstrated that thymol-based essential oils delivered with battery-operated mist diffusers can achieve a high level of *Varroa* mite control. The battery-operated mist diffusion system effectively eliminated fluctuations in temperature and relative humidity and enabled a continuous-release mist diffusion of thymol-based essential oils throughout the hive.

Recognizing the significance and impact of hive humidity and temperature to *Varroa* mite survival, a linear regression model analysis (Part III) was conducted to investigate the significance of local, atmospheric weather-related variables to colony loss. To date, few studies have evaluated the effects of weather-related factors on honey bee colony winter survival. A study of honey bee winter survival in Germany showed high rates of loss in the winter which were then followed by low rates the following autumn based on foraging activity (Johannesen et al., 2022). The authors concluded that colony loss rates in winter are influenced by the honey flow dynamics of the preceding one and a half years. In a recent study conducted in Michigan utilizing a random forest regression model to determine the importance of climate, weather, and land cover on honey bee colony productivity, the investigators concluded broad climate conditions constrained regional floral communities while land use and weather act to further modify the quantity and quality of pollinator nutritional resources (Quinlan et al., 2022). Investigators from Penn State also examined the importance of weather, topography, land use, and management factors on overwintering mortality (Calovi et al., 2021). The authors concluded growing degree days and precipitation from the proceeding summer were the strongest predictors of overwintering survival. Landscape quality factors were found not to be significant. No measurements involving local humidity-related weather variables were included in any of the aforementioned studies. In a search of

Vol. 2023 (2) 122 – 130
ISSN 2688-3651 [online]

the National Institutes of Health PubMed® database, no studies incorporating local-humidity related factors and colony collapse are available.

Acknowledging the impact of hive humidity and temperature to *Varroa* mite survival, a linear regression model analysis (Part III) was conducted to investigate the significance of atmospheric weather-related variables to colony loss. Recognizing that a complex suite of factors would enable beekeepers to make more informed decisions when treating their colonies in the fall, an investigation into the weather-related humidity variables was undertaken. Percent colony loss data provided by the Bee Informed Partnership (BIP) from 2009-2021 for the state of New Jersey and local weather-related variables, including humidity variables, were analyzed. The BIP data is the most comprehensive (capturing summer colony loss and winter colony loss from 2009) and provides the most complete information for the hundreds of apiaries located in New Jersey. Using this unique dataset, the humidity-related variables of relative humidity, dewpoint, vapor pressure deficit, and temperature of the preceding summer were found to be statistically significant to the percent of winter colony losses. Humidity-related summer local meteorological variables may be utilized to predict and alert beekeepers to actively treat for *Varroa* in the fall to prepare for upcoming winters when high colony losses are anticipated. It is important, this research is studied as it may contribute towards the development of a model to understand the contribution of humidity-related variables and its relationship to *Varroa* and honey bee overwintering survival will help to support honey bee management in New Jersey.

2. Materials and Methods

The following section is structured into two subsections. First, the empirical dataset includes a 12-year dataset of honey bee colony wintering survival in addition to the observational weather dataset. Seven meteorological variables were defined to detect and evaluate specific weather variable against the percent of colony loss. Second, the methods of data processing and statistical analysis are described.

2.1 Empirical Data Source: Overwintering Colony Survival

The percent of winter colony lost dataset is derived from the BIP Winter Loss Survey from the years 2009-2021. BIPs flagship service, The BIP National Loss and Management Survey is the longest national effort to monitor honey bee mortality rates in the US. Sent to more than 22,000 beekeepers, the survey functions as a pivotal indicator of honey bee health in the US. (Bee Informed Partnership, 2022). Although winter loss surveys are also available from the USDA/NASS (United States Department of Agriculture/National Agricultural Statistics Service), those reports are only available from 2016-2021 and specific colony locality data is not available. Due to the time and geography limitations of the USDA/NASS dataset, the BIP dataset was utilized for this investigation.

Based upon the BIP surveys, the average number of NJ colonies participating in the winter loss survey varied from 674 to 34,273 colonies with an average of 6,806 within the state. In total, the survey represents 88,478 colonies and 1,455 beekeepers in NJ over a twelve-year period from 2009-2021.

2.2 Empirical Data Source: Weather Variables

Daily and monthly weather variables were obtained from the National Weather Service (NWS), an agency of the US federal government tasked with providing weather forecasts, warnings of hazardous weather, and other weather-related data to organizations and the public for the purposes of protection, safety, and general information. The NWS is a part of the NOAA branch of the Department of Commerce. Sixty weather stations are located throughout New Jersey. To ensure locality of data, all weather variables were cross-referenced to the closest weather station of the colonies participating in the BIP National Loss and Management Survey. For each apiary location, weather variables, more specifically, temperature, precipitation, humidity, dewpoint, wind, and wind gust data was derived from NJ weather (NJ weather, 2023) and data for vapor pressure deficit was sourced from the Iowa State University, Automated Data Plotter (Iowa State University, 2023).

The NOAA provides each station's data in an MS Excel sheet which is broken down to 5-minute intervals. Since the BIP honey bee dataset is reported in summer and winter intervals only, the NOAA data was aligned to the same

Vol. 2023 (2) 122 – 130
ISSN 2688-3651 [online]

reporting intervals. The data was split into four quarters, each containing three months. After doing so, the averages were calculated and used to represent each quarter. All variables that were not reported to the weather station as well as columns with null values were removed. Topographical, forage resource index, and insecticide toxic loads were also not included as these variables had minimal impact in comparison to the significance of climatic variables (Calovi et al., 2021). The dataset of all measured variables is defined in Table 1.

Table 1: Dataset of Measured Variables

% winter colony loss	Percent of colonies lost as reported by beekeepers in winter (December-February)
Quarters	Grouped by 3 months: March-May, June- August, September-November, December-February)
Monthly Mean Temperature	Average monthly average temperature (Fahrenheit) during the quarter
Monthly Total Precipitation	Average monthly total precipitation in the quarter
Monthly Relative Humidity	Average monthly relative humidity in the quarter
Monthly Vapor Pressure Deficit	Average monthly vapor pressure deficit in the quarter
Monthly Dewpoint	Average monthly dewpoint in the quarter
Monthly Wind Speed	Average monthly wind speed in the quarter
Monthly Wind Gust	Average monthly wind gust in the quarter

2.3 Statistical Analyses of Datasets

The analysis of the datasets was accomplished with linear regression. Linear regression was selected for two key reasons 1) enabled forecasting/predictive modeling based on a previously observed numerical data set of values 2) the goal of this research was to explain variation in the response variable based on variation in the explanatory variables. In particular, to determine whether specific explanatory variables may have or not have a linear relationship with the response of percent of winter colony losses.

All datasets were combined into one Excel spreadsheet. The data set was uploaded and run on dataclassroom.com to produce the resulting graphs.

2.4 Statistical Analyses Variables

When building the models, it was pivotal to select variables which were most representative of the entire quarter, therefore the use of monthly minimum, and maximum temperatures were not utilized to build the models as they only utilize one value to represent the entire month. Average monthly snowfall was also not considered in the analyses as not all stations reported these values. Colony stressors were also not used in the models as they are not available in the BIP dataset; however, colony stressors are recorded in the USDA dataset, with *Varroa* mites recorded to be the most common of the stressors. Since winter is the season where NJ beekeepers experience the most drastic loss, the analyses focused on winter loss as the other quarters do not experience nearly as severe losses. The BIP survey data specifically tracks winter and summer losses. Summer loss differs from winter loss as the major factor in summer loss is the health of the queens.

3. Results

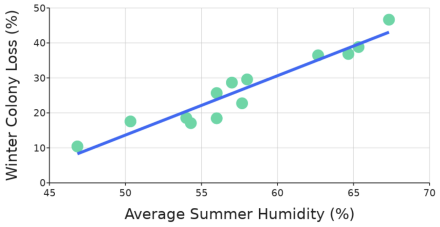
The p-value in regression analysis is used to determine whether the regression model fits the data better than the model with no predictor variables. A predictor that has a low p-value is likely to be a meaningful addition to the model because changes in the predictor's value are related to changes in the response variable. A p-value of ≤ 0.05 is statistically significant. The linear regression p-value for each independent variable tests the null hypothesis that the variable has no correlation with the dependent variable.

Vol. 2023 (2) 122 – 130
ISSN 2688-3651 [online]

The five most important variables for the percentage of winter colony loss included: relative humidity, dewpoint, vapor pressure deficit, precipitation, and temperature from the proceeding summer/warmest quarter (June, July, August). The variables of wind speed and wind gusts were not found to be statistically significant to winter colony loss.

3.1 Measured Variables

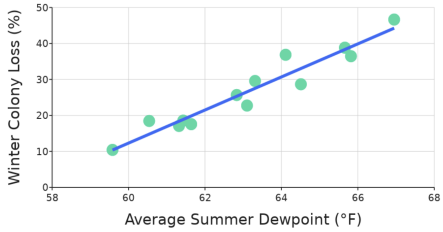
Relative Humidity (RH) refers to the moisture content (i.e. water vapor) of the atmosphere, expressed as a percentage of the amount of moisture that can be retained by the atmosphere at a given temperature and pressure without condensation. Based on the linear regression analysis, as depicted in Figure 1, the average summer RH was found to be statistically significant (p-value ≤ .05) to winter colony loss. As RH increases in summer, the percent of winter colony loss increases.



	r^2	Degrees of Freedom (df)	Slope	Std. Error (SE)	T-Score (Slope / SE)	P-value
Average Summer Humidity (%)	0.914	11	1.7	0.157	11	$3.282e^{-07}$

Figure 1: % Winter Colony Loss versus % of Summer Relative Humidity 2009-2021

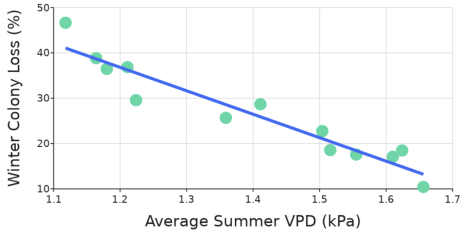
Dewpoint is the temperature the air needs to be cooled to (at constant pressure) in order to achieve a relative humidity (RH) of 100%. It is only dependent on the amount of moisture in the air. Based on the linear regression analysis, as shown in Figure 2, the average dewpoint was found to be statistically significant (p-value ≤ .05) to winter colony loss. As dewpoint increases in the summer, the percent of winter colony loss increases.



	r^2	Degrees of Freedom (df)	Slope	Std. Error (SE)	T-Score (Slope / SE)	P-value
Average Summer Dewpoint (°F)	0.923	11	4.6	0.399	12	$1.773e^{-07}$

Figure 2: % Winter Colony Loss versus % of Summer Dewpoint 2009-2021

Vapor Pressure Deficit (VPD) is the difference between the amount of moisture in the air and how much moisture the air can hold when it is saturated. VPD is used to measure dryness, or aridity, near the Earth's surface. Based on the linear regression analysis, as demonstrated in Figure 3, summer VPD was found to be statistically significant (p-value ≤ .05) to winter colony loss. As VPD increases in the summer, the percent of colony collapse decreases.



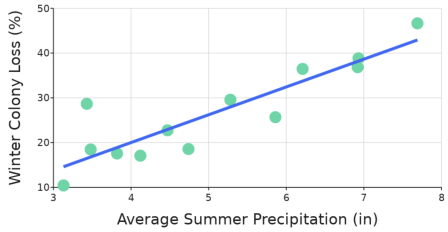
	r^2	Degrees of Freedom (df)	Slope	Std. Error (SE)	T-Score (Slope / SE)	P-value
Average Summer VPD (kPa)	0.914	11	-51.8	4.79	-11	$3.357e^{-07}$

Figure 3: % Winter Colony Loss versus % of Summer Vapor Pressure Deficit 2009-2021

Precipitation is any product of the condensation of atmospheric water vapor that falls under gravitational pull from clouds. Precipitation occurs when a portion of the atmosphere becomes saturated with water vapor (reaching 100% relative humidity), so that the water condenses and "precipitates" or falls. Based on the linear regression analysis, as depicted in Figure 4, summer precipitation was found to be statistically significant (p-value ≤ .05) to winter colony loss. As precipitation increases, the percent of colony collapse increases.

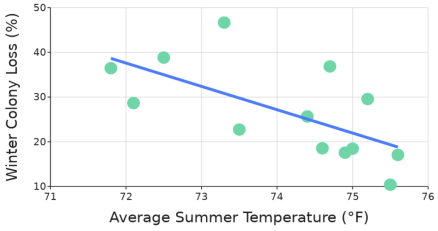
Based on the linear regression analysis, as depicted in Figure 5, summer temperature was found to be statistically significant (p-value > .05) to winter colony loss. As temperatures in the summer increase, the percent of colony collapse decreases.

Vol. 2023 (2) 122 – 130
ISSN 2688-3651 [online]



	r²	Degrees of Freedom (df)	Slope	Std. Error (SE)	T-Score (Slope / SE)	P-value
Average Summer Precipitation (in)	0.791	11	6.21	0.963	6.5	$4.725e^{-05}$

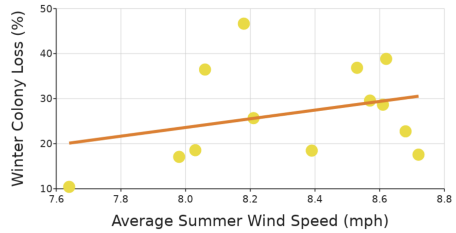
Figure 4: % Winter Colony Loss versus % of Summer Precipitation 2009-2021



	r²	Degrees of Freedom (df)	Slope	Std. Error (SE)	T-Score (Slope / SE)	P-value
Average Summer Temperature (°F)	0.412	11	-5.21	1.88	-2.8	0.01806

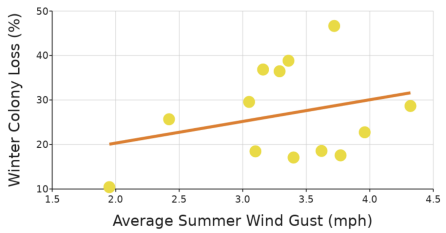
Figure 5: % Winter Colony Loss versus Average Summer Temperature 2009-2021

The variable of average monthly summer wind speed in Figure 6 was not found to be statistically significant ($p>.05$) to winter colony loss.



	r²	Degrees of Freedom (df)	Slope	Std. Error (SE)	T-Score (Slope / SE)	P-value
Average Summer Wind Speed (mph)	0.0923	11	9.65	9.12	1.1	0.313

Figure 6: % Winter Colony Loss versus Summer Wind Speed 2009-2021



	r²	Degrees of Freedom (df)	Slope	Std. Error (SE)	T-Score (Slope / SE)	P-value
Average Summer Wind Gust (mph)	0.0829	11	4.87	4.89	1	0.3402

Figure 7: % Winter Colony Loss versus Summer Wind Gusts 2009-2021

The variable of average monthly summer wind gust in Figure 7 was not found to be statistically significant (p-value > .05) to winter colony loss.

3.2 Predictive Regression Analysis Model

The linear regression analysis revealed the summer conditions of vapor pressure deficit, humidity, dewpoint, precipitation, and temperature were found to be most significant to colony losses of the subsequent winter season. Utilizing the most significant winter colony losses as reported by the BIP NJ beekeepers, a predictive analysis was undertaken. In the 2015/2016, NJ beekeepers reported an average of 46.65% in winter colony losses. Utilizing the statistically significant summer models to run a predictive analysis of the 2015/2016 winter colony losses, the 2015 summer values of 1.118 k/Pa, 67.33%, 66.95 °F, 7.69, and 89°F were inputted. After applying these values into the linear regression models for VPD, humidity, dewpoint, precipitation, and temperature respectively, the model predicted 2015/2016 winter colony losses of 41.05%, 43.05%, 44.27%, 42.93% and 51.04% respectively with an average 2015/2016 winter colony loss prediction of 44.47% which was well within the ±5% margin of error to actual winter colony losses of 46.65% as reported by BIP NJ beekeepers. BIP beekeepers reported nationwide winter colony losses at 44.1% in 2015/2016.

3.3 2022-2023 Winter Colony Loss Prediction

By inputting the statistically significant variables of VPD, humidity, dewpoint, precipitation, and temperature into the linear regression models, at the time of the submission of this manuscript on February 2023, the models predicted an average 2022/2023 colony loss of 31.6% (±5%) for NJ beekeepers. On June 22, 2023, the BIP published their preliminary survey analysis of 2022-2023 (1 April 2022 – 1 October 2022) losses at an estimated 37.4% for managed colonies in the United States as a whole. While this number is not specific to NJ beekeepers and is only a preliminary

Vol. 2023 (2) 122 – 130
ISSN 2688-3651 [online]

nationwide estimate, our regression analysis appears to be within a ±6% margin of error. The final 2022/2023 BIP Loss and Management Survey report of NJ losses and national colony losses may be released up to two years following the availability of the preliminary survey results.

4. Discussion

Beekeepers who use management practices to control *Varroa* mite levels have an overall higher winter survival. However, even with management of *Varroa*, beekeepers still experienced high losses (25-60% mortality) (Calovi, et al. 2021). This analysis examined the influence of local weather variables on colony survival. Summer humidity-related factors were found to be strong predictors of overwintering survival. Of the analyzed weather variables, relative humidity, dewpoint, vapor pressure deficit, precipitation, and temperature of the proceeding summer (June-August) were found to be statistically significant to the subsequent winter colony losses (December-February). A dependence pattern of survival and humidity variables were correlated and the graphic visualization provides a powerful interpretation of how the humidity variables may affect the probability of winter colony survival.

One of the reasons for the proposed impact of weather is the limitation of honey bee foraging time. Honey bees forage for nectar, pollen, water, and propolis, to provide the resources which support colony health. For this reason, the average monthly temperature, wind speed, and wind gust were also investigated. In fact, past studies have demonstrated winter mortality of colonies are positively correlated with uncontrolled *Varroa* mite populations (Genersch et al., 2010; van Dooremalen et al., 2012). Once again, the specific timing of treatment was not available on the BIP dataset. Furthermore, supplemental feedings in the winter were also not tracked in the BIP survey data and may mitigate some of the colony loss reported. Lastly, the BIP survey data does not capture information on the colony size. Smaller and less dense colonies are less likely to overwinter than "high density" colonies.

This is the first study examining the influence of local weather-related humidity variables to colony losses. The humidity-related weather values of relative humidity, dewpoint, vapor pressure deficit, and precipitation, and temperature were found to be significant. The significance of the humidity-related variables may be linked to the bees' ability to thermoregulate the hive effectively during the winter which ultimately can reduce the lifespan of winter bees and contribute to colony mortality. Furthermore, because *Varroa destructor* is known to lose fecundity at absolute humidity of 4.3 kPa, the control of *Varroa* based on humidity-related variables may be linked to honey bee overwintering survival. These findings suggest that honey bees have a "goldilocks" preferred range of summer humidity conditions. Falling outside of this range decreases the probability of surviving the winter. Based on the linear regression model using the most statistically significant summer weather variables of relative humidity, dewpoint, vapor pressure deficit, precipitation, and temperature linear regression analysis predicts NJ winter colony losses for 2022/2023 to be 31.6 % (±5%). The results of this investigation and its subsequent predictive analysis may allow for the development of a comprehensive beekeeping system which includes thymol-based mist diffusion of essential oils and a predictive tool for forecasting honey bee winter survival to support beekeepers' management decisions. The incorporation of humidity-related, local summer weather variables may be utilized to predict and alert beekeepers to actively treat for *Varroa* in the fall to prepare for upcoming winters when high colony losses are anticipated.

Acknowledgements

I would like to acknowledge the following individuals who provided advice, guidance and access to their apiary throughout the phases of my research: Dr. Aaron Reedy (DataClassroom), Dr. Steven Decker (Rutgers), Professor Michael Haberland (Rutgers), Dr. Ron Hutchison (Stockton), Dr. Jack Cameron (University of Florida), Randy Oliver (commercial beekeeper), Geff Vitale (CJBA), Cliff Moore (CJBA), Peter Van Mater (CJBA), my beekeeping mentor Angela Juffey (CJBA), and Mann Lake Ltd for their donation of beehives and equipment.

References

Bahreini, R., Nasr, M., Docherty, C. et al. (2020). Evaluation of potential miticide toxicity to *Varroa destructor* and honey bees, *Apis mellifera*, under laboratory conditions. *Sci Rep* 10, 21529.

Vol. 2023 (2) 122 – 130
ISSN 2688-3651 [online]

Beekeeping - Honey Bees. (2013, June). Penn State Extension. June 2023, from https://extension.psu.edu/beekeeping-honey-bees.

Bryant, D. (2006, February 16). *USDA Bee Lab Teams Pursue Varroa Mite Controls*. Farm Progress. January 22, 2021 from https://www.farmprogress.com/usda-bee-lab-teams-pursue-varroa-mite-controls.

Colony Collapse Disorder (2021, January).US Environmental Protection Agency, from https://www.epa.gov/pollinator-protection/colony-collapse-disorder.

Calovi, M., et al. (2021). Summer weather conditions influence winter survival of honey bees (*Apis mellifera*) in the northeastern United States. *Sci Rep* 11, 1553 (2021). https://doi.org/10.1038/s41598-021-81051-8.

Culbert, K. (2022). "Breakthroughs in Honey Bee Health: Continuous-Release Mist Diffusion of Thymol-Based Essential Oils." *Journal of Research High School*. 2022 (1) 59 – 65. https:// ebf144_0d75aae6c19b43ffa4067efbba37afbe.pdf (journalresearchhs.org.).

Culbert, K. (2023). "Breakthroughs in Honey Bee Health: Continuous-Release Mist Diffusion of Thymol-Based Essential Oils, The Field Study". *Journal of Research High School*. 2023 (1) 46-52. https:// ebf144_d500600c3427454898d548b0627aa88b.pdf (journalresearchhs.org.).

Damiani N, et al. (2009). "Acaricidal and insecticidal activity of essential oils on *Varroa destructor* (Acari: Varroidae) and *Apis mellifera* (Hymenoptera: Apidae)." *Parasitol Res.*106(1):145-52.

Garrido, C. (2018, February 27). Thymol *Varroa* Control. *Bee Culture,* from https://www.beeculture.com/thymol-varroa-control/.

Ghasemi V, Moharramipour S, Tahmasbi G. (2011). Biological activity of some plant essential oils against *Varroa destructor* (Acari: Varroidae), an ectoparasitic mite of *Apis mellifera* (Hymenoptera: Apidae). *Exp Appl Acarol.* 55(2):147-54.

Helping Agriculture's Helpful Honey bees. (2021, January). US Food and Drug Administration, from https://www.fda.gov/animal-veterinary/animal-health-literacy/helping-agricultures-helpful-honey-bees.

Iowa State University, Automated Data Plotter for Vapor Pressure Deficit. August 23, 2022, from https://mesonet.agron.iastate.edu/plotting/auto/?_wait=no&q=126&network=NJ_ASOS&zstation=EWR&year=2017&var=vpd&_r=t&dpi=100&_fmt=png.

Johannessen J, et al. (2022). *Annual Fluctuations in Winter Colony Losses of Apis mellifera Are Predicted by Honey Flow Dynamics of the Preceding Year*. Insects 13 (829). https:// (PDF) Annual Fluctuations in Winter Colony Losses of Apis mellifera L. Are Predicted by Honey Flow Dynamics of the Preceding Year (researchgate.net).

Loss and Management Surveys. Bee Informed Partnership. August 2022, from https:// Loss & Management Survey – Bee Informed Partnership.

National Centers for Environmental Information (NCEI). August 31, 2022, from National Centers for Environmental Information (NCEI) (noaa.gov).

National Weather Service: NOAA. August 31, 2022, from National Oceanic and Atmospheric Administration from https://www.weather.gov/wrh/climate.

New Jersey Weather. August 2022, from *https://njweather.org/*

Quinlan, GM, et al. (2022) Grassy-herbaceous land moderates regional climate effects on honey bee colonies in the North Central *US. Environ Res Lett 17* from https://(PDF) Grassy–herbaceous land moderates regional climate effects on honey bee colonies in the Northcentral US (researchgate.net).

Vol. 2023 (2) 122 – 130
ISSN 2688-3651 [online]

Ramsey SD, Ochoa R, Bauchan G, et al. (2019). *Varroa destructor* feeds primarily on honey bee fat body tissue and not hemolymph. *Proceedings of the National Academy of Sciences*. 116 (5) 1792-1801.

Randall, B. (2020, June 22). The Value of Birds and Bees from USDA. From https://www.farmers.gov/blog/conservation/value-birds-and-bees.

Sabahi Q, et al. (2017). Continuous release of oregano oil effectively and safely controls *Varroa destructor* infestations in honey bee colonies in a northern climate. *Exp Appl Acarol*.72:263-275.

Traynor, K., et al. (2016). In-hive Pesticide Exposome: Assessing risks to migratory honey bees from in-hive pesticide contamination in the Eastern United States. *Sci Rep* 6, 33207.

Three Primary Essential Oils are Useful in Beekeeping. (2021). Tennessee's Honey Bees from https://www.tennesseeshoneybees.com/index.php?main_page=page&id=27&chapter=0.

Honey Bee Surveys and Reports. (2022, August 23). U.S. Department of Agriculture, National Statistics Service. from USDA - National Agricultural Statistics Service - Surveys - Honey Bee Surveys and Reports.

van Dooremalen C., et al. (2012). Winter survival of individual honey bees and honey bee colonies depends on level of *Varroa destructor* infestation. April 27. PLoS ONE 7:e362 from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0036285.

Journal of Research High School

Vol. 2023 (2) 131 – 136
ISSN 2688-3651 [online]

JRHS Outstanding Research Paper Award

Statistical Evaluation of the Correlations between Socioeconomic Factors and the Amount of Trihalomethane in Drinking Water in the State of NY

Stephanie Lee[1] *

[1]Horace Mann Upper School, Bronx, NY, USA
*Corresponding Author: stephanieyl30@gmail.com

Advisor: Sang Park (Ph.D.), spark1@ggc.edu

Received June 14, 2023; Revised July 13, 2023; Accepted, August 11, 2023

Abstract

Trihalomethanes are common byproducts of disinfection processes in public water systems. The relationships between the total amount of trihalomethanes in public water systems in New York and the corresponding socioeconomic variables were studied. A significant amount of chemical and demographic information representing 81% of the NY state population, was obtained from the NY State Department of Health, the US Environmental Protection Agency, and the US Census Bureau. Statistical tools such as Pearson Coefficients with P-values, Grubbs test, and Mean Comparison using Student's t-test were adopted to evaluate the correlations of total trihalomethanes concentration and various socioeconomic factors. Data analysis found negative correlations between the total amount of trihalomethanes and socioeconomic factors, such as mean household income, Asian percentage, and Hispanic percentage. In addition, the water source and the size of the public water system are considered critical factors. The lowest concentrations of total trihalomethanes were detected in communities served by groundwater with higher median household income and higher Asian populations.

Keywords: Trihalomethanes, Socioeconomic, Drinking Water, Statistical Evaluation, New York

1. Introduction

The recent water contamination crises in Flint, Michigan (because of lead in 2014) and the San Joaquin Valley (caused by nitrates and arsenic in 2007) have highlighted the unequal impact of contaminant exposure and poor water system management on populations of color living in poverty. Numerous studies have linked socioeconomic factors, such as income as well as racial and ethnic diversity, to a decline in water quality (Schaider et al., 2019; Switzer & Teodoro, 2017). According to a countrywide examination of drinking-water quality violations between 1982 and 2015, 8.0% of public water systems (PWS) had at least one health-related violation. Overall, there were over 95,000 breaches throughout the 34, with disinfection byproducts (DBPs) accounting for approximately 25% (Allaire et al., 2018). Although disinfection is generally recognized as a significant public health victory for its potential to inhibit the growth of pathogens in drinking water, trihalomethanes are the most common class of decontamination byproducts (DBPs) that form when natural organic matter and antiseptics like chlorine interact during the treatment of drinking water supplies (DeMarini, 2020). Several epidemiological studies showed associations between rectal, colon, and bladder cancers and chlorinated drinking water. These findings support the proposition that trihalomethanes potentially harbor carcinogenic properties in the human body, as substantiated by the outcomes derived from laboratory trials on animal subjects (Costet et al., 2011; Hildesheim et al., 1998). Since 1979, trihalomethanes have been subject to regulations by the United States Environmental Protection Agency (EPA). Currently, the maximum contamination level (MCL) for total trihalomethanes (TTHM) is set at 80 micrograms per liter (80 ppb). The Environmental Working Group (EWG) is the only group whose TTHM standards are stricter than the ones required by federal and state laws, which

Vol. 2023 (2) 131 – 136
ISSN 2688-3651 [online]

the EWG and others consider inadequate. Their maximum allowed concentration is 0.15 ppb (Total Trihalomethanes, n.d.).

In the United States, only a few studies have examined the correlation between trihalomethane levels and sociocultural factors (Harris, 2009; Christman et al., 1983). According to these studies, trihalomethane concentrations in New York were positively correlated with the median family income, racial composition of localities, and community size. However, the previous findings were state-wide normalized information, and could not provide insight into what might be causing the correlations or what other physical factors might be at play.

Considering the previous studies, this research hypothesized that PWSs providing service to communities that are more affluent would have lower trihalomethane levels because of differences in system characteristics and treatment technologies. This study examined the disparities in trihalomethane concentrations among different New York public water systems by identifying the factors that affect these concentrations. Particularly, the association between trihalomethane levels and socioeconomic indicators was analyzed at the city (or town or village) level in the state of New York to determine how the features of water treatment systems may influence these relationships. Multiple statistical approaches were used to accomplish these goals. This is what makes this study unique among others of similar scope.

2. Data Analysis Methods

All communities investigated in this study were identified using the US EPA Drinking Water Information System (SDWIS) to determine the correct PWS IDs before linking the TTHM concentration data with the US Census Bureau data, such as median household income, race percentage, and population. The communities that were not identified on the Census Bureau website were excluded from this study. Most of the excluded towns were small villages with less than 5,000 community members. The total amount of PWS data retrieved was from over 210 systems providing services for approximately 80% of the residents of the state of New York. For towns receiving services from multiple PWS, all TTHM datasets were incorporated.

The correlations between the TTHM concentrations and socioeconomic factors were obtained using the Pearson correlation coefficient (ρ) method with corresponding p-values (Equation 1).

$$\rho = \frac{\sum_{i=1}^{n}(x_i-\underline{x})(y_i-\underline{y})}{(n-1)s_x s_y} \quad (1)$$

where $\underline{x}$, $\underline{y}$, s_x, and s_y indicate the averages and the standard deviations of groups x and y, respectively.

The significance level of the p-value (alpha) for either accepting or rejecting the null hypothesis, "no correlation between TTHM and socioeconomic factors," was set at 5% (0.05). The linear regression tool in the MS Excel data analysis function was used to determine the correlation coefficients and p-values. Additionally, samples were obtained from the NY Department of Health 2019 data for trihalomethanes concentrations and the 2016-2020 Census Bureau data for demographic information.

The outliers of the collected data were identified from each group of data using the Grubbs test, as shown in Equation 2 (Harris, 2009).

$$G = \frac{|questionable\ value - \underline{x}|}{s} \quad (2)$$

where $\underline{x}$ and s indicate the average and standard deviation of the data set, respectively.

A sample was excluded from the calculation when its computed g-value exceeded the critical g-value at the 95% confidence level. If the critical g-value for the specific sample number was unavailable, two pairs of the nearest (one above and one below) critical g-value sample numbers were used to determine the slope, which was then used to calculate a reasonable g-value for the specific sample number adopted for this research.

Vol. 2023 (2) 131 – 136
ISSN 2688-3651 [online]

Further analysis was performed to determine whether there are statistically significant differences between the mean values of TTHM of two different socioeconomic factors. The 95% confidence level student's t-test values were compared with the computed t-test based on the mean comparison between the two selected factors. Equation 3 was used to determine the computed t-test between the two groups (Harris, 2009).

$$t = \frac{|\underline{x_1} - \underline{x_2}|}{s_{pooled}} \sqrt{\frac{n_1 n_2}{n_1+n_2}} \qquad \text{where} \quad s_{pooled} = \sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1 + n_2 - 2}} \tag{3}$$

where s_{pooled} is the pooled standard deviation.

When the t-test value was greater than 95% confidence level student's t value, the compared groups were considered statistically different, rejecting the null hypothesis that there was no significant difference in mean values. Considering the sample sizes for the adopted socioeconomic factors (70 - 208), the tabulated student's t-test values were set between 1.960 and 1.994 at the 95% confidence level, indicating that the two samples were considered significantly different when the computed t-test values exceeded 1.994. All computed t-test values in this research were either higher than 1.994 or lower than 1.960, but none of them fell within that specific range.

The same mean comparison statistical data analysis was conducted after grouping PWS based on the water sources such as groundwater and surface water. The samples within the same socioeconomic variable were divided into two groups: high 50% vs. low 50%. When the total number of samples was odd, the median value was placed into the low 50% group. Based on Equation 3, the differences in specific socioeconomic factors between high and low 50s were reviewed to determine whether they were statistically significant.

3. Results

We examined the relationship between TTHM concentrations and various socioeconomic factors, including the mean household income, racial composition, and population of the serving area. We analyzed 285 PWS, covering 81% of the population of NY.

According to the statistical evaluation, certain socioeconomic factors, such as mean household income, percentage of Asian population, and percentage of Hispanic population, were significantly correlated with the TTHM concentration. Table 1 summarizes the Pearson correlation coefficients and the p-values between TTHM and various socioeconomic factors examined in this research.

Table 1. Pearson correlation coefficients between total trihalomethane (TTHM) concentrations and socioeconomic factors ($\alpha = 0.01$)

	Correlation Coefficient	P-value	Outliers (%)
Income vs. TTHM	-0.430	3.0×10^{-14}	0.96 %
% Asian vs. TTHM	-0.304	1.7×10^{-7}	1.44 %
% Hispanic vs. TTHM	-0.195	9.1×10^{-4}	0.96 %
% White vs. TTHM	0.105	0.076	0.96 %
% Black vs. TTHM	0.011	0.860	1.44 %
Serving size vs. TTHM	0.001	0.991	1.91 %

Table 1 indicates that a relatively small number of samples were identified as outliers and excluded from the statistical analysis. The percentages of the extracted samples for each factor can be found in Table 1. Figures 1 and 2 show the distribution of TTHM concentrations as a function of mean household income and the percentage of Asian population, respectively. According to the data analysis, the TTHM concentration showed a statistically significant relationship with medium income, indicating that an increase in income results in the decrease in the TTHM concentration. Asian and Hispanic percentages illustrated a similar trend, with an inverse relationship with the concentration of TTHM. Particularly, the median household income and the percentage of Asian population revealed a higher negative correlation constant, rejecting the null hypothesis that there is no relationship between TTHM concentration and these variables. However, the percentage of White and Black populations and the total serving size of the water system showed statistically unbiased TTHM concentrations. Figures

Vol. 2023 (2) 131 – 136
ISSN 2688-3651 [online]

1 and 2 illustrate the sample distributions for medium household income and the percentage of Asian population adopted for this research.

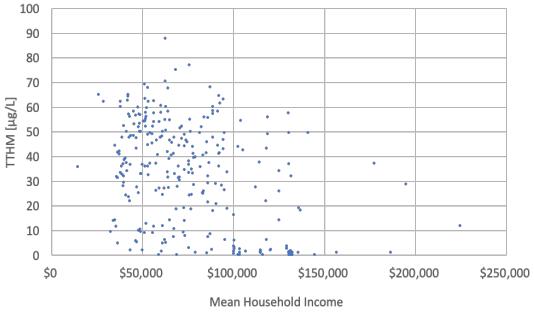


Figure 1. Mean household income plotted against total trihalomethane concentrations at city levels in New York State. Note: TTHM denotes total trihalomethane concentrations

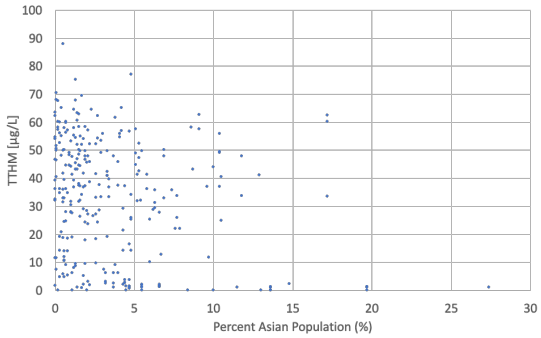


Figure 2. Percentage of Asian population plotted against total trihalomethane concentration at city levels in New York State. Note: TTHM denotes total trihalomethane concentrations.

(mean household income, percentage of Asian population, and community population) were statistically different depending on the water source. As shown in Table 4, no significant difference was observed between the mean TTHM concentrations based on the percentage of Hispanic populations in the two water sources.

Depending on water quality and availability, the State of New York has utilized both surface water and groundwater for drinking water resources. Tables 2 and 3 present the fundamental statistical data calculated separately for the two water resources, focusing on the three statistically significant variables. The average concentration of TTHM in groundwater is about 10.5 µg/L, which is significantly lower than that in surface water (44.0 µg/L). However, the mean values of the three main variables showed higher values in groundwater than those in surface water. According to Table 1, when the correlation coefficients were calculated for each water resource, the groundwater coefficients decreased more than the general values, indicating that the surface water coefficients were neither noticeable nor significant.

Several studies have been conducted to compare the statistical difference between the two mean TTHM concentrations of socioeconomic factors, using the mean student's t-test value comparison method. The computed t-test value of the two sets of TTHM concentrations between the two water sources, including all available socioeconomic factors with the application of Equation 2, was 16.6 (Table 4). This value was significantly greater than the higher cut-off student's t-test value of 1.994, indicating that the two mean TTHM concentrations were statistically different. Table 4summarizes the other results of the comparison between socioeconomic factors. Based on the cut-off range of student's t-test values (1.960-1.994), the mean TTHM values for the three socioeconomic factors

Table 2. Statistical data for TTHM concentration and mean household income with different water sources.

	TTHM (µg/L)			Mean household Income (US $)		
	Mean	STD*	Median	Mean	STD*	Median
Groundwater	10.5	11.9	5.2	99,836	76,903	89,685
Surface water	44.0	14.1	43.6	62,414	49,925	52,493

* Standard deviation

Table 3. Statistical data for Asian and Hispanic Populations with different water sources.

	Asian population (%)			Hispanic population (%)		
	Mean	STD*	Median	Mean	STD*	Median
Groundwater	5.8	5.9	3.9	12.3	9.1	11.6
Surface water	3.0	3.5	1.8	9.1	10.1	7.1

* Standard deviation

Vol. 2023 (2) 131 – 136
ISSN 2688-3651 [online]

4. Discussions

While previous studies have indicated a relationship between the THM concentration and various natural variables (12, 13), a detailed systematic evaluation of socioeconomic factors has not yet been performed. Therefore, we examined the amount of THM in drinking water resources as a function of various socioeconomic variables, including median household income, ethnicity, and water source. According to the Pearson coefficients and p-values calculated, the mean household income has a robust reverse relationship with the concentration of TTHM, as shown in Table 1. In addition, the percentages of Asian and Hispanic populations showed negative correlations with the amount of TTHM in the water system. In contrast, the percentages of White and Black populations showed no statistical significance in relation to the amount of TTHM concentration. Since the data analysis indicated that the TTHM concentrations in groundwater were significantly lower than the ones in surface water sources (Table 2), further evaluations were performed to determine whether the water sources influenced the amount of TTHM in the water system. The mean comparison student's t-test method was used to compare the t-test values to the computed t-test values in order to investigate whether the socioeconomic factors in each water source were statistically different. As shown in Table 4, the mean comparison evaluation method confirmed that the difference in the amounts of TTHM between groundwater and surface water was statistically significant. These results suggested that the type of source water plays an essential role in the amount of TTHM in the water system. Additionally, this was confirmed by the computed t-test value of the mean TTHM concentration comparison, which is significantly higher than the student's t-test value at the 95% confidence level. Furthermore, the median household income and the percentage of Asian population were found to be different, supporting the observation that water source impacts the TTHM concentration. The percentages of Black and White populations were confirmed as unbiased variables in the two water sources and had no correlation with the TTHM amount in the water system (Tables 1 and 4). Moreover, the Hispanic population was the only socioeconomic variable that did not show a significant difference between the two water sources, but the amount of TTHM is negatively correlated with the percentage of population. Furthermore, although the average size of the population served by surface water is 43,000 people, which is significantly higher than that for groundwater cases (24,000 people/system), a correlation was not observed between the amount of TTHM and service size.

Table 4. Computed t-test values to determine the statistical difference between groundwater and surface water.

Socioeconomic Factor	Computed t-test value	S_{pooled} value
All socioeconomic samples	16.599	12.447
Mean Household Income	6.151	30,743
% Asian population	3.331	4.247
Population	2.495	56,067
% Black population	1.721	6.749
% White population	0.748	14.043
% Hispanic population	0.723	9.214

The social variables examined in this research were further statistically analyzed within each water source. For both groundwater and surface water, a mean comparison test of TTHM amount was conducted based on each socioeconomic variable. Each group was divided into two groups: upper 50% and lower 50%. The median value was included in the lower 50% when there were fewer samples. Figure 3 shows a summary of the data analysis. The approximate cut-off t-value for statistical significance was 1.99. When the TTHM amount was evaluated without considering the source of water, the differences in the amounts of TTHM were statistically significant between the upper 50% and lower 50% of the median household income and the percentages of Asian and Hispanic populations. These observations were consistent with the general correlation coefficient results, as shown in Table 1. When the same socioeconomic variables were evaluated for surface water systems, no correlation was observed between socioeconomic variables and TTHM. Furthermore, the median household income and the percentage of Asian populations were statistically significant between the upper and lower 50% of the values for communities served with groundwater.

Vol. 2023 (2) 131 – 136
ISSN 2688-3651 [online]

In summary, the water source is a critical factor in controlling the amount of TTHM in water. The amounts are significantly correlated with the median household income and percentage of Asian population when the communities are served with groundwater.

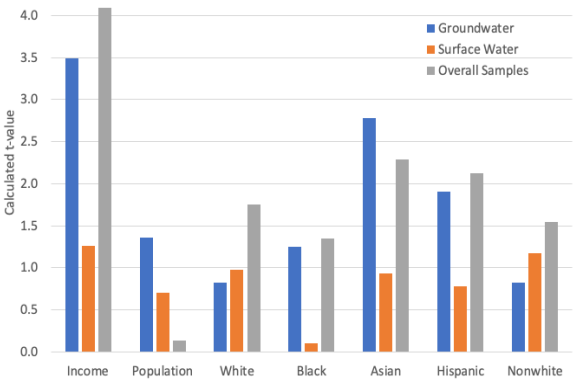


Figure 3. The t-test results for different water sources comparing the upper and lower 50% of each category. When the sample number is odd, the median value was included in the lower 50% of samples.

5. Conclusion

This study examined the relationship between the amount of trihalomethanes, a potential carcinogen, and various environmental and socioeconomic factors, such as income, water sources, and demographic factors. Generally, the amount of TTHM in water is related to the median household income, water source, and some demographic factors, such as composition of Asian and Hispanic populations. The correlation factor increased significantly when the public water system used groundwater. Moreover, high-income families and Asian residents living in communities served by groundwater sources benefited from low TTHM concentration than those served by surface water sources.

References

Allaire, M., et al. (2018). National Trends in Drinking Water Quality Violations. *Proceedings of National Academy of Sciences*, 115(9), 2078–2083. doi:10.1073/pnas.1719805115.

Christman, R., et al. (1983) Identity and yields of major halogenated products of aquatic fulvic acid chlorination. *Environ. Sci. Technol.* 17, 625 - 628.

Costet, N., et al. (2011) Water Disinfection By-products and Bladder Cancer: Is There a European Specificity? A Pooled and Meta-analysis of European Case–Control Studies. *Occupational and Environmental Medicine*, 68(5), 379−85. doi: 10.1136/oem.2010.062703.

DeMarini, D. (2020). A Review on the 40th Anniversary of the First Regulation of Drinking Water Disinfection By-products. *Environmental and Molecular Mutagenesis*, 61(6), 588–601. doi:10.1002/em.22378.

Harris, D. *Exploring Chemical Analysis*. 4th edition, W. H. Freeman and Company, New York, 2009.

Hildesheim, M., et al. (1998). Drinking Water Sources and Chlorination Byproducts II. Risk of Colon and Rectal Cancers. *Epidemiology*, 9(1), 29–35. www.jstor.org/stable/3702610.

Schaider, L., et al. (2019) Environmental Justice and Drinking Water Quality: Are There Socioeconomic Disparities in Nitrate Levels in U.S. Drinking Water? *Environmental Health*, 18(3). doi: 10.1186/s12940-018-0442-6.

Switzer, D., & Teodoro, M. (2017) The Color of Drinking Water: Class, Race, Ethnicity, and Safe Drinking Water Act Compliance. *Journal of American Water Works Association*, 109(9), 40–45. doi:10.5942/jawwa.2017.109.0128.

Total trihalomethanes (TTHMs), Retrieved May 20, 2023 from https://www.ewg.org/tapwater/contaminant.php?contamcode=2950

Journal of Research High School

Vol. 2023 (2) 137 – 145
ISSN 2688-3651 [online]

Traffic Sign Detection and Recognition with Deep Learning, CNNs, YOLOv3, and Keras

Aniket Gupta[1] *

[1]Silver Creek High School, San Jose, CA, USA
*Corresponding Author: anigupta102@gmail.com

Advisor: Prof. Susan Fox, fox@macalester.edu

Received February 25, 2023; Revised July 22, 2023; Accepted, August 28, 2023

Abstract

The growing industry of transportation being integrated with Artificial Intelligence has contributed to applications regarding driving assistance and autopilot software. With this application being relatively new, more research still needs to be done to improve these algorithms' recognition/classification of traffic signs. In order to improve modern driving softwares, I proposed designs for a basis for an object detection model and a functioning neural network classification model to be used as an end-to-end solution. The object detection model serves as a locator of a traffic sign in an image to be inputted into the classification model to determine the identity of the sign. The detection model was designed using Keras and YoloV3. During the development phase for the classification model, I used a previously made model as a basis and used image processing techniques of deep learning, OpenCV, and the multi-layers of convolutional neural networks to modify it. Subsequently, I used the German Traffic Sign Recognition Benchmark (GTSRB) for training which contains over 50,000 images of German traffic signs for training and validation. The classification model resulted in 99.71% training accuracy and 98.7% - 99.21% validation accuracy. Furthering, this study discussed each of the convolutional layers used for the classification model and how it was changed from the original model.

Keywords: Image Processing, Deep Learning, OpenCV, Convolutional Neural Networks, Classification, Object Detection, Keras, YoloV3

1. Introduction

The transportation industry has advanced immensely within the last few decades, especially in automobiles. Recently, car manufacturers have begun integrating AI for driving assistance and even autopilot. One aspect of this situation is figuring out how to make autopilot and driving assistance safer for drivers and passengers.

For autopilot, one skill the AI/car needs is to identify street signs to evaluate its surroundings and abide by traffic laws for the passenger's safety. Driving assistance must also properly detect, identify, and classify signs to account for the millions of car accidents due to driving errors. Much attention has been brought to Traffic Sign Recognition as there is much research and work on this problem. Traffic sign recognition involves two main parts first, detecting the sign from a video feed or still image, and second identifying what sign it is. Detecting a sign uses image processing and deep learning object detection architectures. Classifying the type of sign uses a deep learning CNN classification model.

In recent years, deep learning has been the “go-to” over machine learning for tackling this problem as it requires less human intervention during training. A sub section of deep learning, Convolutional Neural Networks (CNN), based on computer vision (or OpenCV) with other models, has been used to efficiently analyze street view images and pick out unique features that can be used to identify signs. However, this has been proven difficult as many unaccounted

Vol. 2023 (2) 137 – 145
ISSN 2688-3651 [online]

factors interfere with the identification process, such as shadows, light reflection, the sign being partially covered, the quality of the sign being poor from the camera, etc.

Thus to tackle this problem and to improve driving softwares, the objective of this research project is to produce a system that can efficiently detect and classify traffic signs, with both easy and rough environmental conditions, to be applied in real time applications. With sufficient identification and classification of traffic signs, the world of driving could become significantly safer as computers and driving algorithms can gain more information regarding the vehicle's current surroundings.

Furthermore, this two-part system involves object detection and classification programs which I created by improving a CNN model that has been proven effective on one dataset and use it for another dataset, the GTSRB dataset, which stands for German Traffic Sign Recognition Benchmark dataset. This famously known dataset contains 43 different classification classes and more than 50,000 images for training and testing, making it a big enough dataset for me to train a coherent classification model.

To produce this project, TensorFlow and Keras were used for the models. "TensorFlow is a library containing multiple machine learning tasks and APIs such as Keras" (Terra 2022). Keras is a deep learning API developed by Google and runs on TensorFlow and is highly used because it is built with Python.

1.1 Deep Learning

Deep learning is related to artificial intelligence. It utilizes big chunks of data with complex algorithms to create artificial neural networks to train computers and machines to make data-based decisions (happiest minds, 2022). Feature extraction is a crucial aspect of deep learning. Feature extraction is where features, patterns, and properties are extracted from inputted data given to the neural network model. This is also how image recognition is done. Kondamari, Itha (2021) defines image recognition as where the model extracts details and patterns from small sections of the image using the pixels, which depend on other surrounding pixels. These models have been proven to be great substitutes for complex codes/programs for extracting features.

1.2 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks are deep learning algorithms trained to take in an input image, assign importance by adding learnable weights and biases to various objects in the image, and be able to differentiate one from the other (Saha, 2018). The goal of CNNs is to shape or reduce the image into a form that is easier to process without losing the necessary features to extract for a prediction in image recognition or classification. These networks contain many layers/filters, each summing up to allow the network to detect spatial relationships to allow the classification of an object. Spatial relationships are where one thing depends on another, such as pixels in an image correlating/depending on its surrounding pixels and vice versa.

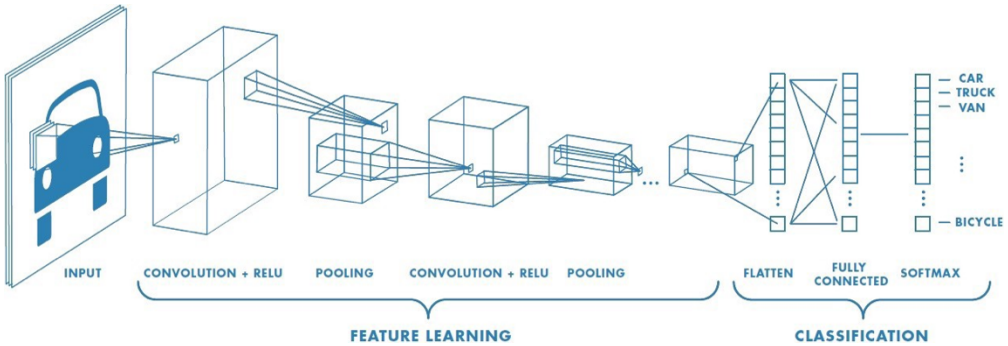


Figure 1. A diagram of a CNN with all usual its layers and classification portion (Saha, 2018).

Vol. 2023 (2) 137 – 145
ISSN 2688-3651 [online]

CNN models have an input layer that takes in an image and an output layer at the end, where the output is a reduced vector of probability scores. As seen in figure 1, in between are the many different layers in a CNN network: convolutional, max pooling, flatten, dense, fully connected layer, and softmax. The first couple of layers is for the model to learn features, while the second portion is to classify the features detected.

The first layer is the convolutional layer which captures low-level features (colors, gradient orientation, etc.) and uses the convolutional mathematical operation to extract high-level features (such as edges from the input image) and uses these to train its layers and adapt to allow the network to perceive the image as humans would (Saha 2018). The convolution operation is expressed as:

$$s(t) = (x * w)(t)$$

where x is the input data or the pixels from the input image, and w is the kernel applied to these pixels (Kondamari & Itha, 2021). A Kernel is a matrix with smaller dimensions than the image and filter. This operation slides this filter over the input until every location is covered by the kernel at least once and is multiplied by its respective values on the input and then summed up (Dertat, 2017). This sum is then placed in the feature map of the same size as the input (Figure 2), which becomes the output. This output is then sent as an input to the next layer and so on.

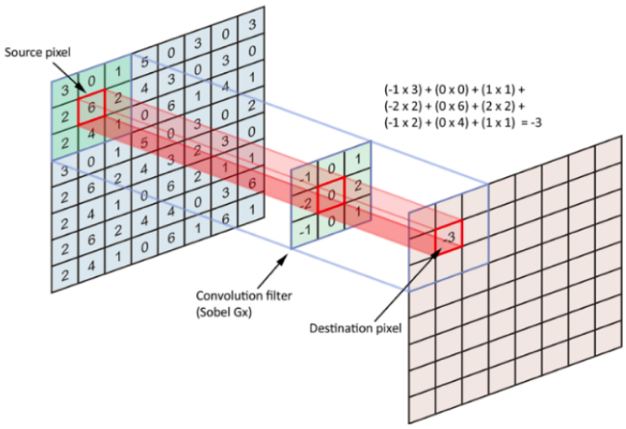


Figure 2. The process of how the operation is used with the kernel to produce a feature map (Cornelisse 2018).

The Pooling layer reduces the dimensions/size of the feature map summarizing the features presented in the feature map, so the amount of computation is decreased, making the data much more efficient to process (Saha 2018). There are two types of pooling, Max Pooling, and Average Pooling. Max Pooling returns the maximum values from the image region the kernel is covering as it slides over, while Average Pooling returns the average values.

Max Pooling Noise Suppressant where it reduces the noise (unnecessary features usually in the background that are picked up) as the dimensions are reduced. After the pooling, convolutional and pooling layers may be repeated as many times as necessary for the model.

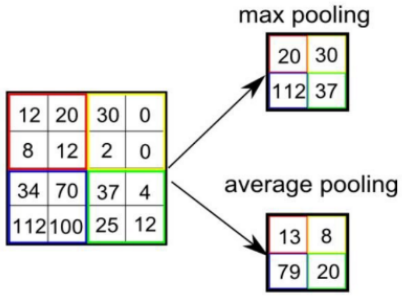


Figure 3. The process of how max pooling and average pooling works is shown. Max pooling takes the largest value in each kernel, while average pooling takes the average.

The next layers are the Fully-Connected layers which make up the last few layers of the network for classification. The output from the final layer before the Fully-Connected (Conv, MaxPool, AvgPool) is a three-dimensional matrix that is flattened into a one-dimension vector to be connected to the Fully-Connected layers (Arc, 2018). After the model runs through a series of epochs (the number of passes through the dataset the learning algorithm has completed), it can differentiate between dominating and low-level features with the Softmax Classification technique. The model outputs a vector with zeros and ones for each classification class from the dataset where a "1" is the identified class.

1.3 Object Detection with YOLOv3 and Keras

YOLO stands for "You Only Learn Once." YOLO is a real-time object detection algorithm or CNN that identifies specific objects in videos or images by providing class probabilities of the detected images (Meel, 2019 & Anirudh, 2020 & Karimi, 2021). This means it will provide the likelihood of each object detected being a certain object. The YOLO algorithm uses a DarkNet framework. The number of layers depends on the version but typically consists of 3

Vol. 2023 (2) 137 – 145
ISSN 2688-3651 [online]

x 3 or 1 x 1 filters, convolutional and residual (Genç, 2019 & Karovalia, 2021). A neural network has data flowing through each layer consecutively, where the output of one layer is the input for the next. A residual layer is when a residual connection provides an alternate path to reach the later parts of the neural network by skipping layers in between (Wong, 2020).

First, an image is split into grid cells with equal dimensions smaller than the image's dimensions to analyze the image piece by piece. Next, YOLO uses the concept of anchor boxes to detect multiple objects. In the model, the parameter for the number of anchor boxes and their respective shapes would be inputted, and each object lying in a close neighborhood will be detected and assigned to a box. More anchor boxes, more objects that can be detected and processed. The object detected in the image would be bounded with a bounding box as an outline consisting of its width, height, class (general categories like person, car, traffic light, building, etc.), and the center of the bounding box.

During this process, Non-max Suppression (NMS) ensures an object is only detected once rather than multiple times when the object lies across multiple grid pixels in the image. NMS works by discarding cells whose probability of the object being present calculated in the final softmax output layer is less than or equal to 60%, then will take the cell with the largest probability among the remaining possible candidates and select it as its prediction. Finally, NMS will discard any leftover cells with an intersection-over Union (IOU) value greater than 60% (Natashia, 2020). Karimi (2021) defines IOUs as a phenomenon in object detection which describes how these boxes and grid cells overlap. The YOLO model would then use this IOU to output a box surrounding the object and its label. This is the predicted box. Karimi also stated that YOLO would ensure the bounding box and the prediction box is the same as the IOU will equal 1.

1.4 Related Works

Latif et al. (2020) developed a recognition software for traffic signs (in this case, Arabic traffic signs) to aid drivers in reducing vehicle accidents on the streets and help with self-driving cars or some recognition system. Their project used Deep CNN to develop their system, which first detects the sign and then classifies or recognizes it. They collected 2718 images to form their Saudi Arabian Traffic and Road Signs database. Before Latif et al. (2020) published their paper, they had tested their initial CNN network, and their improved one showed 100% accuracy for all batch sizes tested on their developed dataset. With their CNN model already showing promising results, it would be useful to implement it in my research.

Saha, Islam et al. (2018) developed a unique traffic recognition software using not one but three CNN models. Two models were sub-models meant to classify signs, but the main model was trained to decide which of the two models to use to classify the signs using data augmentation. This is a unique approach resulting in accuracy results of close to 99%. However, it would be more efficient to use one or maybe two models in total than three or more. They also used the GTSDB dataset to train their models in this research project.

Pramod Sai Kondamari and Anudeep Itha (2021) discussed Traffic Sign Recognition (TSR) for drivers and self-driving cars. The author's techniques used Deep Learning CNN and OpenCV. The dataset they used was the GTSRB dataset. They discussed methods used during experimentation of how their model worked and layers used(four convolutional layers, two max-pooling layers, and two dense layers) (Kondamari & Itha, 2021). Their model gave results of 95% model accuracy with 30 epochs. Their paper was a great reference to me as they heavily explained each concept (deep learning, CNNs, CNN layers, etc.) involved in their project.

An object detection model was produced by Valentyn Sichkar & Sergey A. Kolyubin (2020). They discussed in their paper where the GTSDB dataset was used to train their first detection model, which would detect and separate the signs into four categories. The GTSDB dataset was split into training and validation sections with 85% and 15% proportions, respectively, with 630 and 111 images. Once detected, the located fragment of the sign would be cut and inputted into the second classification model to be classified and outputted with its coordinates, bounding boxes, and labels.

Vol. 2023 (2) 137 – 145
ISSN 2688-3651 [online]

2 Materials and Methods

My project successfully implemented the classification network, and I designed an object detection network for the same dataset with YOLOv3 to predict the location of the signs (localization) from video frames and images.

The overall structure would work like this. First, the image would go through preprocessing, resized to 30 by 30, and not converted to grayscale. Subsequently, the image would go through the object detection model to detect the sign contained in the image. The next part would be to lower the region of interest (ROI) to the area of the detected sign and then input that region into the classification model to identify which sign it is. Finally, the model would return information on the sign to the autopilot algorithm.

2.1 Classification Model

I used the GTSRB dataset to train the classification model, which contains over 50,000 RGB images. For preprocessing, I resized each image in the dataset to 30 x 30 and stored them in a NumPy array, and the labels for each sign in another array. Afterwards, I split the images with 80% for training and 20% for validation.

Table 1. Summary of the complete final Classification model

Layer	Output Shape
conv2d (Conv2D) (input shape (30, 30, 3)) RGB image (32 layers)	(None, 30, 30, 2)
max_pooling2d (MaxPooling2D) (2, 2)	(None, 15, 15, 32)
conv2d_1 (Conv2D) (64 layers)	(None, 15, 15, 64)
max_pooling2d_1 (MaxPooling2D) (2, 2)	(None, 7, 7, 64)
conv2d_2(Conv2D) (128 layers)	(None, 7, 7, 128)
max_pooling2d_2 (MaxPooling2D) (2, 2)	(None, 3, 3, 128)
flatten (Flatten)	(None, 1152)
dense (Dense)	(None, 128)
dense_1 (Dense) (Softmax)	(None, 43)

Furthering, I built the model as a sequential model. There are 3 convolutional layers, 3 max-pooling, 1 flattening, and 2 dense layers, as seen in Figure 4. All convolutional layers used the relu activation, and the kernel initializer was set to uniform, and the padding set to same. The first portion of layers serve to provide feature learning for the model. The first layer of the portion is a 2D convolutional layer with 32 filters and 3,3 kernel size, and an input shape of (30, 30, 3). The second one is a 2D max-pooling layer with a pool size (2, 2). Following is another convolutional layer with 64 filters and a 2D max pooling with (2, 2) pool size. Lastly, there is one convolutional layer and max-pooling with 128 filters and (2, 2) pool size. After feature learning, the next set of layers serve as classification. We first transform the pooled feature map into a one-dimensional vector with flattening. Then we have two dense layers with the first activation set to relu and the second dense is an output layer with its activation set to softmax. The Adam optimizer was used for training to handle sparse gradients on “noisy” problems. A final summary of the classification model can be seen in Table 1.

2.2 Object Detection Model

For sign detection, although more research needs to be completed, a pre-trained YOLO network was modeled. The model used the DarkNet code base and GTSDB dataset (German Traffic Sign Detection Benchmark). 106 2D convolutional block layers were loaded, and the model was defined and loaded with the provided model weights. Once the model was loaded with its weights, the model would be saved to use for a prediction. We would require an input

Vol. 2023 (2) 137 – 145
ISSN 2688-3651 [online]

image, in this case, a street-view image of a traffic sign (or a video frame from a video feed) and have the model predict the location of the sign.

3 Results

As stated before, the classification network produced in this research project was based on the neural network provided by Latif et al. in their research paper (Table 2).

Their model gave 97% accuracy from training their model with the Arabic dataset. Although when I was training the model with the GTSRB dataset, the accuracy results were minimal. After attempting to train over 50 epochs, I observed no improvements in training and validation accuracy and training and validation loss remained constant at 3.49% and 3.48% respectively.

For the final CNN model, I removed the two dropout layers from the original model and also followed a convolutional, max pooling layer pattern to efficiently process the input data. There are three conv2d layers, each has max-pooling right after. Following these changes, the model showed significant improvement as it stopped underfitting and could learn better without the dropout layers shown in Figure 5. The model showed about 99.71% training accuracy and 98.7% - 99.21% validation accuracy. However, the validation loss was minimally increasing over time with each epoch but stayed low enough for sufficient accuracy results.

Table 2. Summary of the original classification model (Latif et al. 2020).

Layer	Output Shape
conv2d (Conv2D) (input shape (30, 30)	(None, 28, 28, 64)
max_pooling2d_1 (MaxPooling2)	(None, 14, 14, 64)
dropout_1 (Dropout)	20%
max_pooling2d_1 (MaxPooling2D)	(None, 12, 12, 32)
max_pooling2d_2 (MaxPooling2D)	(None, 6, 6, 32)
flatten_1 (Flatten)	(None, 1152)
dense_1 (Dense)	(None, 128)
dropout_2 (Dropout)	(None, 128)
dense_2 (Dense)	(None, 32)

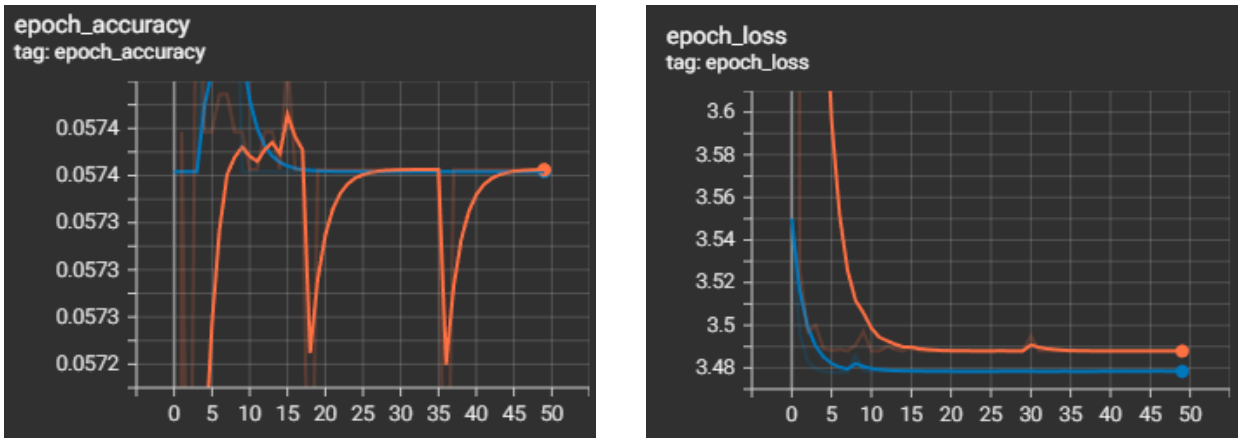


Figure 4. The overall accuracy and loss over 50 epochs of the original CNN model. Orange is the training data and blue is the validation data.

For object detection, I tested the model with other generic pictures that should work (a zebra, multiple zebras, planes, etc.), and the model was properly able to detect them. This confirmed the model was working. However, in the MSCOCO dataset, I discovered that it never included images of signs for the model to detect. As such, when inputted an image of a German street sign, the model could not recognize it. Further research will need to be done to see if this model can efficiently run and detect signs when trained on the GTSDB dataset.

Vol. 2023 (2) 137 – 145
ISSN 2688-3651 [online]

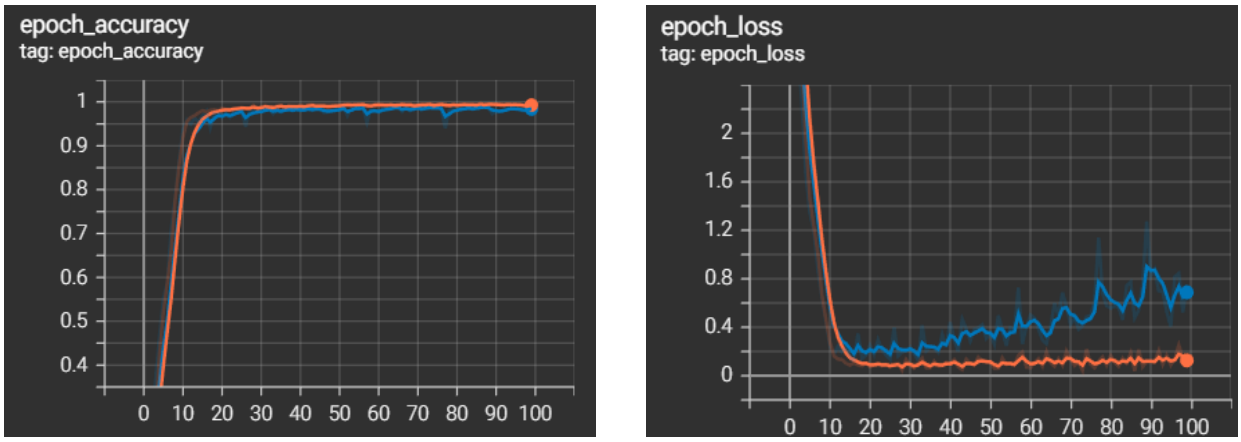


Figure 5. the accuracy and loss data for both training and accuracy across 100 epochs for the final CNN classification model. Orange is the training data, and blue is the validation data.

4 Discussion

4.1 Improving Original Model

When using the original classification model developed by Latif et al., the model showed no improvements in accuracy or loss. The reason for this was perhaps because this model was developed alongside a significantly smaller dataset. Another possibility was the two dropout layers (Table 2) took out too much data. Instead of causing "generalizing" to prevent the model from memorizing, too much data was dropped, and the model could not learn and train effectively.

To improve results, I removed the two dropout layers as too much data was being taken out to work with the current dataset. Another thing I had to account for was the lack of convolutional layers in the original model, as there is only one 2D convolutional layer. More CNN layers were needed so more elementwise multiplication could be conducted on the 2D input data to have better feature maps for the output/production. I also increased the output shape of the first layer as well as the max pooling layers to leave more information. After implementing these changes, the results significantly improved (Figure 5). Although, with each additional epoch, the training loss stayed relatively constant, the validation loss started to increase. The reason for this was perhaps because I didn't convert the images to grayscale (making the images gray) during preprocessing. This would've reduced some of the background noise in each image, preventing the validation loss from increasing and increasing accuracy results.

4.2 Future Work

This research project aims to integrate these detection and classification models into built-in vehicle systems to improve driving assistance or autopilot software. Too many deaths occur due to driving errors; if the car or vehicle could recognize signs, help manage the speed, and be aware of how dangerous the surroundings are, many lives could be saved. For future work, my first step would be to get the sign detection model working and add improvements based on performance. To start out with, the object detection model can be attempted to be trained by the GTSDB dataset and add improvements depending on results. Next would be to see where minimal improvements to the CNN model could be added. Although the model produces promising results, even a 1-2% increase in accuracy to close to 100% will be more beneficial for driving assistance and autopilot software, perhaps by converting the images to grayscale to reduce noise. After, I would then prepare a program to use both models to detect the sign, lower the ROI to the region of the sign(s), and input it into the classification model to classify. There, I would have it return information on the sign such as its shape, label of sign, the image of the sign, etc.

Vol. 2023 (2) 137 – 145
ISSN 2688-3651 [online]

5 Conclusion

In this paper, I presented a working classification CNN model as well as a proposed plan for an object detection CNN model (and presented methods that worked in the past) to detect and classify signs.

The CNN classification model proposed by Latif et al. didn't work for the GTSRB due to the dropout layers. Furthermore, the classification model was improved by adding more convolutional layers with max-pooling layers and removing the dropout layers. The model was modified to work with RGB images and trained on the GTSRB dataset to achieve increased accuracy in performance.

Object detection proved to be a little tricky in the given time as detecting signs can be difficult as we can't solely rely on finding triangular, rectangular, or circular shapes to find these signs. Signs in images or videos can partly be covered by an object, or the lighting may interfere wildly if the light bounces back from the sign. Using Keras, YOLOv3, DarkNet, and a dataset (such as GTSDB) can be used to train a sign detection model.

From this research, it is well enough possible for computers to detect and classify signs in real-time with minimal errors successfully. AI for self-driving or driving assistance can majorly be improved with it being able to properly assess the potential dangers of the surroundings and follow traffic laws simply by correctly identifying signs.

Acknowledgment

I would like to thank my mentor, Professor Fox for the guidance of using OpenCV, deep learning, CNN's, and object detection. I would also like to thank her for her guidance throughout the research process when brainstorming, researching, and model building and for editing this paper.

References

Latif, G. et al. (2019). Autonomous traffic sign (ATSR) detection and recognition using Deep CNN. Procedia Computer Science, 163, 266–274.

Alpaydin, E. (2021). Machine learning. Amazon. Retrieved July 3, 2022.

Anirudh. (2020, July 9). German traffic signs detection using yolov3. Medium. Retrieved July 3, 2022.

Arc. (2018, December 25). Convolutional Neural Network. Medium. Retrieved July 7, 2022.

Bouguezzi, S. (2020, November 20). GTSDB - German Traffic Sign Detection Benchmark. Kaggle. Retrieved July 7, 2022.

Brownlee, J. (2017, July 3). Gentle introduction to the adam optimization algorithm for deep learning. Machine Learning Mastery. Retrieved July 7, 2022.

Brownlee, J. (2019, October 7). How to perform object detection with yolov3 in Keras. Machine Learning Mastery. Retrieved July 3, 2022.

Chollet, F. (2017). Save and Load Models : Tensorflow Core. TensorFlow. Retrieved July 3, 2022. Updated: 6/16/2022

Convolutional Neural Networks (CNN) with Deep Learning. HappiestMinds. (2020, January 28). Retrieved July 3, 2022.

Dertat, A. (2017, November 13). Applied deep learning - part 4: Convolutional Neural Networks. Medium. Retrieved July 3, 2022.

Doval, G. et al. (2019). Traffic sign detection and 3D localization via deep convolutional neural networks and Stereo Vision. 2019 IEEE Intelligent Transportation Systems Conference (ITSC

Vol. 2023 (2) 137 – 145
ISSN 2688-3651 [online]

Forson, E. (2022, March 13). Recognising traffic signs with 98% accuracy using Deep Learning. Medium. Retrieved July 3, 2022.

freeCodeCamp.org. (2018, April 24). An intuitive guide to Convolutional Neural Networks. freeCodeCamp.org. Retrieved July 3, 2022.

Genç, Ö. (2019, January 4). Hands on machine learning example. Medium. Retrieved July 3, 2022.

Jain, S. (2021, July 29). CNN: Introduction to pooling layer. GeeksforGeeks. Retrieved July 7, 2022.

Karimi, G. (2021, April 15). Introduction to yolo algorithm for object detection. Section. Retrieved July 3, 2022.

Karovalia, A. (2021, April 25). Traffic signs detection using tensorflow and YOLOV3/ yolov4. Medium. Retrieved July 3, 2022.

Khanvt. (2022, June 6). Traffic signs classification. Kaggle. Retrieved July 3, 2022.

Kondamari, P. S., & Itha, A. (2021). A Deep Learning Application for Traffic Sign Classification. Faculty of Computing at Blekinge Institute of Technology, i-40. Supervisor: Shahrooz Abghari, PhD

McCullum, N. (n.d.). The flattening and full connection steps of Convolutional Neural Networks. Nick McCullum Headshot. Retrieved July 3, 2022.

Meel, V. (2022, April 19). Yolov3: Real-time object detection algorithm (what's new?). viso.ai. Retrieved July 3, 2022.

Mykola. (2018, November 25). GTSRB - German Traffic Sign Recognition Benchmark. Kaggle. Retrieved July 3, 2022.

Natashia, N. (2020, June 15). Yolo V3 object detection with keras. Medium. Retrieved July 3, 2022.

Rivas, P. (2020). Deep learning for beginners: A beginner's guide to getting up and running with deep learning from scratch using Python. Packt Publishing Ltd.

Rosebrock, A. (2020, July 13). R-CNN object detection with Keras, tensorflow, and Deep Learning. PyImageSearch. Retrieved July 3, 2022.

Saha, S. (2018, December 17). A comprehensive guide to Convolutional Neural Networks-the eli5 way. Medium. Retrieved July 3, 2022.

Saha, S. et al. (2018). An efficient traffic sign recognition approach using a novel deep neural network selection architecture. Advances in Intelligent Systems and Computing, 849–862.

Sichkar, V. (2021). Traffic signs detection by Yolo V3, opencv, keras. Kaggle. Retrieved July 3, 2022.

Simplilearn. (2021, September 18). What is Keras and why it so popular in 2021: Simplilearn. Simplilearn.com. Retrieved July 3, 2022.

Stewart, M. (2020, July 29). Simple introduction to Convolutional Neural Networks. Medium. Retrieved July 3, 2022.

Świeżewski, J. (2020, May 22). Yolo algorithm and Yolo Object Detection. Machine Learning. Retrieved July 7, 2022

Team, K. (n.d.). Keras Documentation: MaxPooling2D layer. Keras. Retrieved July 3, 2022.

Terra, J. (2022, May 27). Keras vs tensorflow vs pytorch [updated]: Deep learning frameworks: Simplilearn. Simplilearn.com. Retrieved July 3, 2022.

Wong, W. (2021, December 18). What is residual connection? Medium. Retrieved July 7, 2022.

Journal of Research High School

Vol. 2023 (2) 146 – 153
ISSN 2688-3651 [online]

Crack and Powder Cocaine: Exploring Disparities in Cocaine-Related Laws and Their Enforcement

Alexandra Vassilev[1] *

[1]Westford Academy, Westford, MA, USA
*Corresponding Author: fps.phs@gmail.com

Advisor: Ivaylo Vassilev, I.I.Vassilev@soton.ac.uk

Received December 31, 2022; Revised July 28, 2023; Accepted, September 11, 2023

Abstract

The widespread use of cocaine in the US in the 1980s and the subsequent War on Drugs brought about many long-lasting effects on racial disparities in the justice system. An uneven sentencing ratio between the two forms of a widely used drug, crack and powder cocaine, was set by politicians to prevent further usage, but it appears that for decades Black users were put at an unfair disadvantage in the legal and law enforcement systems. While this extreme ratio has been brought down in recent years, the law still holds that users who possess crack cocaine are punished more severely. This paper draws on Donald Black's theory of law to try to offer an explanation of this disparity. The paper argues that Black's theory provides a useful framework for highlighting how social inequalities shape the legal and law enforcement systems and are then also reflected in the ongoing bias which affects the levels of punishment of different social groups. Specifically, the paper argues that a combination of factors related to inequalities of not just race but also socio-economic status needs to be considered in explaining, and trying to find solutions, to the persistent disparities in how the use of crack and powder cocaine is coded in law and how the law is enforced.

Keywords: Racial disparities, Crack cocaine, Powder cocaine, Donald Black's theory of law, Social inequalities

1. Introduction

1.1 Background

Cocaine originates in South America and was initially used by native Americans for medical purposes. The leaves of the coca plant were used to create powder, which was injected or inhaled. If injected cocaine gives the user a quicker and more intense high than when it is inhaled. Crack cocaine is based on powder cocaine, which is mixed with baking soda and water, creating a hard rock-like substance. Crack cocaine comes in small, easy to distribute doses (Coca, n.d.), it is typically smoked and, similar to injecting powder cocaine, gives the user an intense high that lasts for only a couple of minutes (Crack vs. Cocaine, 2022). Though produced and used differently, crack and powder cocaine are chemically the same drug, and the short-term effects (e.g. nausea, elevated body temperature, heightened blood pressure) and long-term effects (e.g. adverse impact on the liver, kidneys, the neurological and cardiovascular systems) are almost the same (Lisw, 2023). Despite the lack of meaningful differences between crack and powder cocaine there are enduring disparities in how their use is treated in law and in law enforcement. The dominant framing of this problem in academic debates and in the media tends to be around race and institutional racism. The paper reviews existing evidence and hypotheses that other factors, such as socio-economic status are at least as important.

1.2 Legal disparities and the changing meanings of using powder and crack cocaine

Powder cocaine became popular in Europe as an anesthetic for eye surgeries during the 19th century but later began to make its mark on North America in the late 20th century after it expanded its use from a painkiller to a recreational drug (Hutchinson, 2022). Cocaine was being significantly exploited without addressing the addictiveness of the drug and the mental and physical consequences of using it. The use of cocaine was primarily attributed to Blacks, and was associated with rumors about cocaine users being violent and having greater resistance to bullet wounds than other people (Hart, 2014).

Vol. 2023 (2) 146 – 153
ISSN 2688-3651 [online]

In the 1980s it was the use of crack cocaine, not cocaine as such, that came to be associated with Blacks (Beckett et al. 2005). This might in part be due to the easy distribution, smaller packaging, and lower price of crack cocaine (Hutchinson, 2022) and because a larger percentage of the Black population lives in poor areas or in poverty. For example, between 1980 to 1986, Blacks constituted 12.2% of the population and 59% lived in poverty (O'Hare, 1987), which for white Americans was 80% (Frey, 2020) and 10.2%, respectively (Census Bureau Reports 1980 Poverty Statistics, 1982). The rapid growth of the use of crack cocaine became a social and political concern in the mid-1980s (Crack Cocaine Sentencing Policy, 1997). The Anti-Drug Abuse Act of 1986 was aimed at addressing these concerns. The Act focused on creating a higher sentence for the possession and distribution of crack cocaine compared to that of powder cocaine. Specifically, possession of every 5 grams of crack cocaine would carry the same sentence as possession of 500 grams of powder cocaine, 1:100 ratio. There was no clear rationale supporting this decision as crack and powder cocaine are chemically the same drug. Before being passed into law, "[C]ongress simply considered various arbitrary ratios (including [1-to-20]) and settled on the [1-to-100] ratio, with no evidence to support that figure" (England, 2013). This law was widely discussed as leading to high arrest and conviction rates of Blacks as compared to whites. Substantial evidence has been developed since the period, which explores the impact of the passing and application of the 1986 Act. Before the 1986 Act was passed, Blacks had an 11% higher drug sentencing rate than whites, and after four years, the rate rose by 49% (Vagins & McCurdy, 2006). Figure 1 displays a line graph showing the difference in these rates of Black and white drug arrests, with Black arrests surpassing that of white arrests throughout all three decades. The greatest difference in arrests was recorded in 1987, where 1,346 (per 100,000) Blacks were arrested compared to 299 (per 100,000) whites for drug-related crimes (Human Rights Watch, 2009).

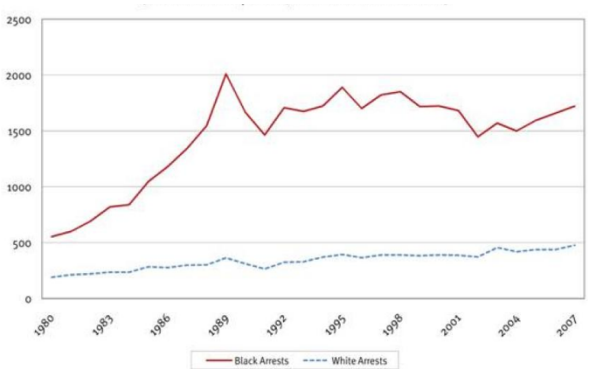


Figure 1. Comparing the Rates of Black and White Drug Arrests Between the Years 1980 and 2007 (Human Rights Watch, 2009).

Almost two decades after the Act was passed, 66% of crack users were either white or Hispanic (Vagins & McCurdy, 2006), which may indicate that the prevalence of crack cocaine use in this period was similar among different groups in the US. However, another 2006 sentencing report released in Massachusetts found that 80% of defendants who received mandatory sentences were Black or of another race, while of those arrested for state-level drug offenses, 45% were white (Greene et al., 2006). Another statement from 2006 highlights these issues with an ex-federal prosecutor of Massachusetts claiming that between drug users, Black defendants were more often charged with "a drug-free zone offense and face[d] the two-year mandatory prison sentence than whites" (Greene et al., 2006). The Act has also been seen as ineffective in targeting major drug felonies, which was one of its stated aims, while appearing to reinforce other evidence about the underlying racial disparities it created and reinforced. For example, it is estimated that around 13% of Black adult men may be disenfranchised due to being given felony convictions (Vagins & McCurdy, 2006).

Partly as a recognition of the lack of evidence about different impacts of powder and crack cocaine, and partly due to pressure related to the racial aspect of the disparity in sentencing, the Fair Sentencing Act of 2010 was passed during Obama's presidency. The Fair Sentencing Act aimed to address the racial bias in the system by reducing the sentencing for the use of powder and crack cocaine from 1:100 ratio to a 1:18 ratio (Carle, 2010). While it is still not a completely equal sentencing ratio, reports over the years have shown that this has led to a gradual decrease in the number of Blacks sentenced for the use of cocaine. However, a 2020 report found that the crack offenders arrested by police were still predominantly Blacks, 77.1% of all arrests, compared to 6.3% for whites and 15.9% for other ethnicities, respectively (Crack Cocaine Offenses, 2018).

There are two related questions that can still be asked. These are first, why has the 1:18 ratio not been made a 1:1 given the overwhelming evidence of the absence of any substantive differences between powder and crack cocaine? And second, why are Blacks disproportionately affected in terms of arrests and convictions? The former, relating to how the use of cocaine is coded in law, can be explained through reference to deterrence theory and cost-benefit analysis. One of the explanations that can account for the latter question is around racial bias on the level of individuals and institutional racism, especially in relation to police discretion in enforcing the law.

Vol. 2023 (2) 146 – 153
ISSN 2688-3651 [online]

2. Deterrence theory: justifying disparity in the legal framing of the use of crack and powder cocaine

The basis for many acts passed on criminal legislation can be attributed to, and justified through a reference to, deterrence theory, which states that "human behavior is driven by cost-benefit ratios" (Walker & Mezuk, 2018) and originates from philosopher Jeremy Bentham and Cesare Beccaria's utilitarian perspective on human behavior (Tomlinson, 2016). Drawing on deterrence theory, it may be argued that crack cocaine was inexpensive and therefore more attainable than powder cocaine. The Anti-Drug Abuse Act of 1986 created an arrest ratio for which the cheaper alternative of the cocaine drug received greater punishment, meaning that citizens were more likely to be deterred from its usage.

A similar logic is consistent with the 2010 Fair Sentencing Act. Today, on average, crack cocaine on the street sells at roughly $60 (per gram), while powder cocaine can range from $93 to $163 (per gram) (Addiction Resource Editorial Team, 2021). The current ratio in cost between the two types of cocaine could be seen, from the perspective of deterrence theory, as encouraging avoidance of punishment and violation of the law. This commonly used rationale for crime reduction could provide further reasoning for the remaining disparity in sentencing ratios and drug arrests. However, there is evidence that such rationale is not applied consistently. White collar crimes, for instance, are typically given lighter sentences and therefore, recidivism rates are higher (Fredericks et al., 2016). Such crimes are more likely to be committed by people who are white and of higher socio-economic status, which may indicate an underlying race and class bias (Sohoni and Rorie, 2021).

3. Police discretion and racial bias in enforcing the law

Discretion, or self-judgment, is valued by police officers as they associate it with having a positive professional self-identification, higher financial rewards, and an expression of their achievement and the value of their contribution to society (Grawitch et al., 2009). There are several aspects of police work which require a degree of discretion. These can include prioritizing certain areas to focus on patrolling, the ways in which an agency operates, and the extent to which strict regulation is exerted over officers.

Police discretion is often put to the test where the officers more freely patrol and need to make decisions based on their assessment of the surrounding context and the situation. This prompts the question of how each officer carries their work with reasonable discernment, and how decisions may be shaped by pre-existing bias, individual and shared experiences from the past, and awareness of trends in data collected from varied arrests.

On the organizational level, higher levels of police discretion appear to be associated with, what is seen as, a high level of complexity of each agency and the context in which it operates. Organizational content (the environment of the agency, e.g. crime rates, social status and racial composition in the surrounding area) and an agency's control (how the varied police departments officers work to administer and formalize the organization through written policies and regulations) are some of the criteria according to which police agencies are regulated (Nowacki & Spencer, 2019). Results from a recent study on how police organization influences arrests found that "police officers in more structurally complex agencies tend to have more discretion. This is because as organizations grow more complex, direct supervision becomes less feasible" (Nowacki & Spencer, 2019).

Police discretion is also justified through defining some areas as 'hot spots' and 'broken windows' areas. These are small geographic, usually urban, areas where crime levels are high and as a result, policing tends to occur (Hot Spots Policing, n.d.). These areas also make police more likely to engage in police-citizen interactions and perform drug-related arrests while patrolling (Wheeler, 2019). Although adopting such methods could reduce crime rates, many of these police stops can lead to a racial disparity in arrests, i.e. high number of arrests for minor offenses compared to small number of arrests for people committing felonies. As Blacks are more likely to live in areas defined as 'hot spots', 'broken windows', and 'complex to police' they are more likely to be associated with higher levels of police discretion, and thus also of inherent police bias and institutional racism (Winship, 2021). This may help explain the policing applied to cocaine-related drug arrests as well. However, as Blacks represent 13.2% of the United States' total population but 23.8% of the population living in poverty (Creamer, 2020), it may be worth considering whether police discretion may also be related to bias against people living in deprived areas or are disadvantaged in other ways. A study on police demands and stress in smaller suburban areas found that officers working in lower socio-economic status communities were more likely to view these areas as high-crime areas (Grawitch et al., 2009). Furthermore, in an additional study on police discretion, it was found that when controlling for socioeconomic status, there was no difference in the rate of arrests between Blacks and whites (Campbell et al., 2021).

Racial bias and institutional racism understood through the prism of police discretion, and deterrence theory may be insufficient to fully address the two questions asked in the beginning of the paper. This is because there is evidence that there may be other processes shaping the legal framing of drug offenses and the application of law in practice.

Vol. 2023 (2) 146 – 153
ISSN 2688-3651 [online]

Specifically, the above findings indicate that while racism and discrimination within the legal and enforcement systems hold influence on arrests and convictions, socio-economic disadvantages through poverty and social standing should also be considered as important contributing factors. Donald Black's Theory of Law (1976) may be a useful framework to help account for the disparity in the arrests and conviction rates related to the use of crack and powder cocaine.

4. Donald Black's Theory of Law

Donald Black uses a social-structural framework to define law as "the number and scope of prohibitions, obligations, and other standards to which people are subject" (Black, 1976). Using the factors of stratification, morphology, culture, organization, and social control, Donald Black argues that the quantity of law could vary in situations involving certain obligations, and prohibitions can have differing effects on each individual (Gottfredson and Hindelang, 1979). Donald Black argues that social stratification can be measured on a vertical scale- the way an individual is positioned into society, and wealth, education, and race can have a large impact on how a crime is weighed in the criminal justice system (Campbell et al., 2021). Morphology, the way an individual is integrated in society, which can be measured on a horizontal scale, refers to the "aspect of social life, the distribution of people in relation to one another, including their division of labor, networks of interaction… whether societies, communities, neighborhoods, or organizations" (Gottfredson and Hindelang, 1979). For example, Black argued that education and culture, described as the aspect of social life, were directly related to changes in law (Kuo et al., 2011). In relation to drug offenses, what Black calls the social stratification, and the morphology and culture of society can offer a way of understanding how such offenses are defined in law and how the law is enforced.

This may suggest that crimes that are more often committed by individuals who are less well positioned in a society (lower in social stratification and morphology) are likely to be defined as more serious crimes within the legal system. Donald Black would refer to this as a deviance in law. It may also be more likely that the law is more strictly enforced for people in such a position. Additionally, those who are socioeconomically advantaged may be more likely to use their greater connections to "economic, social, political, and legal resources" with police to "ensure they work in their favor" (Hernández & Heimark, 2021).

5. Applying Black's theory: Understanding the disparity in the context of inequalities of socio-economic power and social status

In a recent study on the behavior of police and whether race plays a factor in their arrests, the results showed that "arrestee race did not independently influence type of police discretion" and "that arrestee race played a role when interacting with neighborhood concentrated disadvantage" (Campbell et al., 2021). The existing evidence indicates that Black users of cocaine who are arrested for cocaine offenses are typically of lower socioeconomic status (Hernández & Heimark, 2021), and wider data on arrests continue to question methods of law enforcement as policymakers realize how "crime control strategies… intensified law enforcement in low-income communities of color" (Hinton & Henderson, 2018). Figure 2 indicates that people who were arrested and booked multiple times in 2017 were more likely to be poor (e.g. have an annual income below $10,000), be unemployed, and have less than a high school education (Jones & Sawyer, 2019).

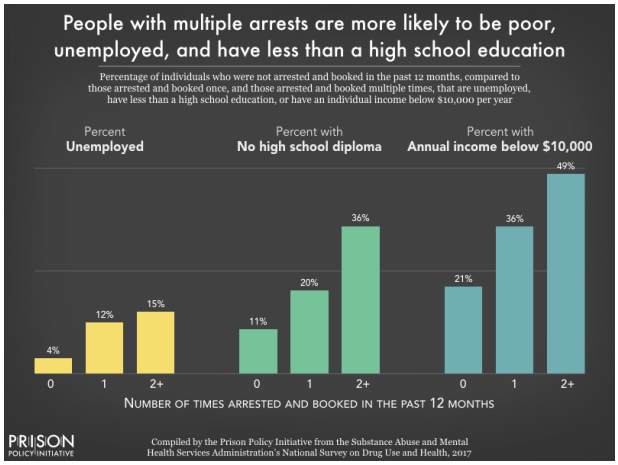


Figure 2. Number of Times Arrested and Booked in the Past 12 Months (Jones & Sawyer, 2019).

These findings are consistent with Black's framework and indicate that race on its own cannot sufficiently explain disparities in cocaine related arrests as well as the current 1:18 sentencing ratio. Race must instead be paired with other factors, such as those on the vertical and horizontal domains in Black's theory, in order to better understand police discretion and the decisions officers make in police-citizen interactions. If Black's theory is correct, we may expect that white people living on low incomes and in poorer neighborhoods, and who are also not well integrated into professional, local, and cultural networks, may experience high levels of arrests if using crack cocaine. Because poverty is associated with

Vol. 2023 (2) 146 – 153
ISSN 2688-3651 [online]

race (Parolin & Lee, 2022), the experience of the disparities in law and its application are more likely to be observable in the treatment of Blacks. However, while Blacks, compared to poor whites, are likely to experience additional disadvantages related to institutional racism, it should be noted that key factors driving unequal treatment of powder and crack cocaine appear to be related to differences in income, wealth and education. As for the sentencing ratio, Black's theory may also explain why the punishment for cocaine has greater leniency towards powder cocaine users. These are likely of higher social standing and able to afford the more expensive version of the drug; this may be compared to the easily accessible and more appealing option for groups with lower incomes and thus stricter tendencies for punishment directed at them. Developing a complementary focus on the role of socio-economic status and social integration in shaping the law and how it is enforced helps us in avoiding explanations and solutions that are focused on a single factor (e.g. race and racism), as well as in identifying other groups that are likely to also be affected by these disparities (e.g. people living in poverty).

6. Conclusion

This paper aimed to answer two questions: why the 1:18 crack to powder cocaine ratio has not been lowered to 1:1, and which factors may contribute to the disproportionate arrests of Black users. The existing evidence indicates that deterrence and police discretion theories only offer a partial explanation of the legal framing of cocaine offenses and the enforcement of the law.

Donald Black's theory offers a way of looking at drug related arrests and sentencing ratio for crack and powder cocaine in relation to a wider set of factors (summarized in vertical and horizontal domains) that include, but are not reducible to, institutional racism and disadvantage related to race. Existing evidence, consistent with Black's theory, indicates that socio-economic disadvantage, for example, is also an important predictor of disparities in powder and crack cocaine arrests. Black's theory may suggest that addressing the disparity in crack and powder cocaine arrests may require developing a better understanding as to how the legal and police systems are prejudiced against Blacks, but also against people of lower socio-economic status, people living in poor neighborhoods, as well as people who are less socially integrated into professional, local, and cultural networks. Focusing on socio-economic disparities and how these relate to other structural conditions in shaping drug laws, arrests, and imprisonment among different users offers a more realistic picture (compared to single factor explanations) of the underlying causes of the enduring disparities in how powder and crack cocaine use are framed. This research paper helps identify groups affected by these disparities that may be less visible and improves understanding of the experiences of cocaine users who are Black.

Identifying and lessening the disparities within the legal and law enforcement systems (such as in the treatment of powder and crack cocaine offenses) may require addressing problems of the enduring and wide-ranging impact of poverty and inequalities in income, wealth, and opportunities, in addition to problems of social exclusion, and everyday institutional racism. Black's theory offers a useful framework through which to explore such questions.

References

Black, D. J. (1976). The Behavior of Law. New York : Academic Press.

Crack cocaine sentencing policy: Unjustified and unreasonable. (1997). Sentencing Project.

Addiction Resource Editorial Team. (2021). Ho*w Much Does Cocaine/Crack Cocaine Cost?* (J. Smith, Ed.). Addiction Resource. https://www.addictionresource.net/cost-of-drugs/illicit/cocaine/

Census Bureau Reports 1980 Poverty Statistics. (1982, August 22). *The New York Times*, p. 32.

Beckett, K., et al. (2014). "Drug Use, Drug Possession Arrests, and the Question of Race: Lessons from Seattle." *Social Problems*, https://academic.oup.com/socpro/article/52/3/419/1631764.

Carle, D. (2010). *Leahy Praises Senate Passage Of Historic Bill To Address Crack/Powder Sentencing Disparity* [Press release]. https://www.leahy.senate.gov/press/leahy-praises-senate-passage-of-historic-bill-to-address-crack/powder-sentencing-disparity

Campbell, W., Griffiths, E., & Hinkle, J. (2021). The behavior of police: class, race, and discretion in drug enforcement. *Police Practice and Research*, *23*(3), 337–354. https://doi.org/10.1080/15614263.2021.2022482

Vol. 2023 (2) 146 – 153
ISSN 2688-3651 [online]

Coca. DEA Museum. (n.d.). https://museum.dea.gov/exhibits/online-exhibits/cannabis-coca-and-poppy-natures-addictive-plants/coca

Communications, N. W. (2015). *Powder vs. Crack: NYU Study Identifies Arrest Risk Disparity for Cocaine Use*. https://www.nyu.edu/about/news-publications/news/2015/february/-powder-vs-crack-nyu-study-identifies-arrest-risk-disparity-for-cocaine-use.html

Crack vs. Cocaine: What's the Difference? (2022, July 21). The Recovery Village at Palmer Lake. https://www.palmerlakerecovery.com/cocaine-addiction/difference-crack-cocaine/

Creamer, J. (2020, September 15). *Inequalities Persist Despite Decline in Poverty For All Major Race and Hispanic Origin Groups*. United States Census Bureau. https://www.census.gov/library/stories/2020/09/poverty-rates-for-blacks-and-hispanics-reached-historic-lows-in-2019.html

England, D. C. (2013). *Crack vs. Powder Cocaine: One Drug, Two Penalties*. CriminalDefenseLawyer. https://www.criminaldefenselawyer.com/resources/crack-vrs-powder-cocaine-one-drug-two-penalties.htm

Fellner, J., Walsh, S., & Roth, K. (n.d.). (rep.). (M. Smart, Ed.) *Punishment and Prejudice: Racial Disparities in the War on Drugs* (Vol. 12, Ser. Racially Disproportionate Incarceration of Drug Offenders).

Fredericks, K.A., McComas, R.E., & Weatherby, G. A. (2016). White Collar Crime: Recidivism, Deterrence, and Social Impact. *Forensic Research &Amp; Criminology International Journal*, *2*(1). https://doi.org/10.15406/frcij.2016.02.00039

Frey, W. H. (2020). The nation is diversifying even faster than predicted, according to new census data. *Brookings*. https://www.brookings.edu/research/new-census-data-shows-the-nation-is-diversifying-even-faster-than-predicted/

Friedersdorf, C. (2020). Apply broken-windows theory to the police. *The Atlantic*. https://www.theatlantic.com/ideas/archive/2020/06/apply-broken-windows-theory-police/613186/

Gottfredson, M. R., and Hindelang, M. J. (1979). A Study of the Behavior of Law. *American Sociological Review*, *44*(1), 3–18. https://doi.org/10.2307/2094813

Grawitch, M. J., Barber, L. K., & Kruger, M. H. (2009). Role identification, community socio-economic status demands, and stress outcomes in police officers. *Anxiety, Stress & Coping*, *23*(2), 165–180. https://doi.org/10.1080/10615800902935542

Greene, J., Pranis, K., & Ziedenberg, J. (2006). Disparity by Design: How Drug-Free Zone Laws Impact Racial Disparity – and Fail to Protect Youth. *Justice Policy Institute*. https://justicepolicy.org/wp-content/uploads/justicepolicy/documents/06-03_rep_disparitybydesign_dp-jj-rd.pdf

Hart, C. L. (2015). "How the Myth of the 'Negro Cocaine Fiend' Helped Shape American Drug Policy." *The Nation*. https://www.thenation.com/article/archive/how-myth-negro-cocaine-fiend-helped-shape-american-drug-policy/.

Hendricks, L., & Wilson, A. (2013). The Impact of Crack Cocaine on Black America. *National Forum Journal of Counseling and Addiction*, *2*(1). http://www.nationalforum.com/Electronic%20Journal%20Volumes/Hendricks,%20LaVelle%20The%20Impact%20of%20Crack%20Cocaine%20on%20Black%20America%20NFJCA%20V2%20N1%202013.pdf

Hernández, W., & Heimark, K. R. (2021). Does context matter? Examining robbery reporting in a high crime country. *Criminology & Criminal Justice*, *0*(0). https://doi.org/10.1177/17488958211031344

Human Rights Watch. (2009). *Drug Arrests and Race in the United States*. Decades of Disparity. https://www.hrw.org/report/2009/03/02/decades-disparity/drug-arrests-and-race-united-states

Vol. 2023 (2) 146 – 153
ISSN 2688-3651 [online]

Hinton, E., & Henderson, L. (2018). An Unjust Burden: The Disparate Treatment of Black Americans in the Criminal Justice System. *Vera Institute of Justice*. https://www.vera.org/downloads/publications/for-the-record-unjust-burden-racial-disparities.pdf

History.com Editors. (2018, August 21). *Cocaine*. HISTORY. https://www.history.com/topics/crime/history-of-cocaine

Hutchinson, S. (2022, November 1). *History of Cocaine in America: When Did it Become Illegal?* River Oaks. https://riveroakstreatment.com/cocaine-treatment/illegal-history-in-america/

Iceland, J., & Wilkes, R. (2006). Does Socioeconomic Status Matter? Race, Class, and Residential Segregation. *Social Problems*, *53*(2), 248–273. https://doi.org/10.1525/sp.2006.53.2.248

Jemal, A., et al. (2008). Widening of Socioeconomic Inequalities in U.S. Death Rates, 1993–2001. *PLoS ONE*, *3*(5), e2181. https://doi.org/10.1371/journal.pone.0002181

Jones, A., & Sawyer, W. (2019). *Arrest, release, repeat: How police and jails are misused to respond to social problems*. Prison Policy Initiative. https://www.prisonpolicy.org/reports/repeatarrests.html

Kuo, S., et al. (2011). Crime reporting behavior and Black's *Behavior of Law*. *International Sociology*, *27*(1), 51–71. https://doi.org/10.1177/0268580911423054

Lisw, M. H. M. (2023). *Crack vs Cocaine: What's The Difference Between Crack & Cocaine?* American Addiction Centers. https://americanaddictioncenters.org/cocaine-treatment/differences-with-crack

Nowacki, J. S., & Spencer, T. (2019). Police discretion, organizational characteristics, and traffic stops: An analysis of racial disparity in Illinois. International Journal of Police Science & Management, 21(1), 4–16. https://doi.org/10.1177/1461355719832617

O'Hare, W. P. (1987). Black Demographic Trends in the 1980s. *The Milbank Quarterly*, *65*, 35–55. https://doi.org/10.2307/3349950

Palamar, J. J., et al. (2015). Powder cocaine and crack use in the United States: An examination of risk for arrest and socioeconomic disparities in use. *Drug and Alcohol Dependence*, *149*, 108–116. https://doi.org/10.1016/j.drugalcdep.2015.01.029

Parolin, Z. and Lee, E.K., 2022. The role of poverty and racial discrimination in exacerbating the health consequences of COVID-19. The Lancet Regional Health–Americas, 7.
Peters, Mark G., & Philip K. Eure. NYPD, New York City, New York. (2016). *An Analysis of Quality-of-Life Summonses, Quality-of-Life Misdemeanor Arrests, and Felony Crime in New York City, 2010-2015* .

Practice Profile: Hot Spots Policing. (n.d.). CrimeSolutions, National Institute of Justice. https://crimesolutions.ojp.gov/ratedpractices/8

Quick Facts: Crack Cocaine Offenses. (2018). *United States Sentencing Commission*. https://www.ussc.gov/sites/default/files/pdf/research-and-publications/quick-facts/Crack_Cocaine_FY20.pdf

Sohoni, T., & Rorie, M. (2021). The whiteness of white-collar crime in the United States: Examining the role of race in a culture of elite white-collar offending. *Theoretical Criminology*, *25*(1), 66–87. https://doi.org/10.1177/1362480619864312

Tomlinson, K. D. (2016). An Examination of Deterrence Theory: Where Do We Stand? *Federal Probation*, *80*(3), 33.

Turner, D. S. (2016). *crack epidemic | United States history [1980s]*. Encyclopedia Britannica. https://www.britannica.com/topic/crack-epidemic

Vol. 2023 (2) 146 – 153
ISSN 2688-3651 [online]

Vagins, D., & McCurdy, J. (2006). Cracks in the System: Twenty Years of the Unjust Federal Crack Cocaine Law. In *American Civil Liberties Union*. https://www.aclu.org/sites/default/files/field_document/cracksinsystem_20061025.pdf

Walker, L. S., & Mezuk, B. (2018). Mandatory minimum sentencing policies and cocaine use in the U.S., 1985–2013. *BMC International Health and Human Rights*, *18*(1). https://doi.org/10.1186/s12914-018-0182-2

Wheeler, A. P. (2020) Allocating Police Resources While Limiting Racial Inequality, Justice Quarterly, 37:5, 842-868, DOI: 10.1080/07418825.2019.1630471

Winship, S., et al. (2021). Long Shadows: The Black-White Gap in Multigenerational Poverty. In *Brookings*. https://www.brookings.edu/wp-content/uploads/2021/06/Long-Shadows_Final.pdf

Journal of Research High School

Vol. 2023 (2) 154 – 160
ISSN 2688-3651 [online]

How the Financial Status of a Person Can Affect Their Ability to Take Unpaid Maternity Leave and as a Result, Lead to Gender, Health, and Financial Inequalities.

Sanam Mehta[1] *

[1]Syosset High School, Syosset, NY, USA
*Corresponding Author: sanammehta713@icloud.com

Advisor: Tyler Vitone, tyler@polygence.org

Received December 10, 2022; Revised July 27, 2023; Accepted, September 5, 2023

Abstract

Introduction: Maternity leave allows a mother and child to bond while healing emotionally and physically from childbirth. In the United States, however, not all employers have a paid maternity leave policy, and women often shorten their leave after childbirth for fear of losing their job. The Family Medical Leave Act (FMLA) is a US federal policy created with the aim to allow women to take maternity leave while securing their employment when they return. However, this is taken as unpaid leave. Additionally, access to FMLA unpaid leave is not universally available. Two case studies of mothers describing their difficulties with FMLA while recovering from childbirth were analyzed for themes and commonalities to help identify factors that may result in inequities across race, financial status, health and employment in a qualitative study. The case studies are examples that can be applied to many in the US. Financial status and job insecurity made taking unpaid maternity leave more difficult, and what influenced the amount of time spent on unpaid leave were found. Paid maternity leave is not always available for women, and they sometimes have to accept unpaid maternity leave via FMLA. This leads to inequities and additional financial, emotional and health burdens.

Keywords: Identity, Gender, Socioeconomic class, Education, Politics

1. Introduction

Maternal leave is the time a mother spends away from work after giving birth to a child. It allows the mother to heal physically and emotionally, and adjust to the needs of a new baby (Masters, 2022). A review of the literature shows that maternal leave is beneficial for the health of the mother and baby (Booth, 2023). Staehelin et. al. reported a review of 13 original studies which together found a positive association between the length of maternity leave and the mother's health. More specifically, increasing maternity leave from 6 or fewer weeks to 12 or greater weeks was associated with a reduction in depressive symptoms (Chatterji, 2004 and Hyde, 1995). McGovern et. al. reported that the length of maternity leave and maternal health has a direct association. With regards to the direct health of the child, it was found that the length of parental leave was inversely related to infant mortality rates (Ruhm, 2000).

As discussed above, there is a clear benefit to maternity leave. This benefit is even stronger when the maternity leave is paid leave. The transition to parenthood is difficult and takes a large toll on every new parent; these changes include weight changes, mood swings, hormonal shifts, and mood disorder risk (Abrams, 2022). A significant benefit of providing paid maternity leave is the overall improved health of women, both physical and mental. According to Abrams, paid leave after giving birth results in a 51% decrease in the risk of rehospitalization. This was demonstrated in Norway when the 12-week unpaid leave policy was changed to a 16-week paid leave policy, leading to a dramatic increase in the health of mothers in terms of blood pressure, pain levels, exercise, and smoking behaviors (Abrams,

Vol. 2023 (2) 154 – 160
ISSN 2688-3651 [online]

2022). Paid parental leave policies not only bolster health, but they also improve household security, relationship security, and reduce the levels of intimate partner violence. After California passed its paid family leave law, poverty risk among mothers of infants decreased by more than 10% (Abrams, 2022). Aside from physical benefits, paid maternity leave improves the mental health of parents.

Providing paid leave also results in greater health for the newborn children. Paid leave is associated with lower infant mortality rates and higher rates of vaccination, especially for lower income families (Abrams, 2022). The likelihood of infants being re-hospitalized in the first year is reduced by 47% (Coombs, 2021). A child's development and immunity is boosted as well due to the fact that women with access to paid leave breastfeed their children more often (Abrams, 2022). There is also an association between paid maternal leave and brain activity in infants that causes more mature early cognitive functioning (Abrams, 2022). Toddlers who had parents with paid parental leave had better language skills and fewer behavioral issues. The incidence of head trauma in children younger than 2 years old has been reduced when parental paid leave is provided due to lower levels of stress and abusive behavior of parents (Coombs, 2021).

Many countries outside of the United States (US) have recognized the need to support their working mothers and fathers in the early stages of parenthood and have actively taken measures to ensure that this support is provided. At least 178 countries have guaranteed paid maternity leave and 50 countries have guaranteed paid paternity leave (Human Rights Watch 2011). More than 100 countries offer a minimum of 14 weeks of paid maternity leave, including some of the most developed nations in the world such as Canada, Australia, and the United Kingdom.

Furthermore, for members of the Organization for Economic Co-Operation and Development (OECD), the average country provides 18 weeks of paid maternity leave, with 13 of these weeks at full wages (Human Rights Watch, 2011). The US is the only member of OECD that does not offer a national paid leave policy. Currently, there are only eight countries that do not guarantee paid leave, and of these eight countries, the United States is the only country that is classified as a wealthy or high-income country (Sholar, 2016). The US falls short of its international counterparts with about one-fifth of U.S. employees having little or no access to maternity leave (Ray, Gornick, & Schmitt, 2008). Paid leave also strengthens a country's generosity and performance by giving parents who work for low wages the financial security they need to take leave and care for their new children (Ray, Gornick, & Schmitt, 2008). The US as a country, however, has not implemented a universal paid leave policy. There are a few states that have implemented paid leave policies, such as New York. The benefits are evidenced by a 20% reduction in the number of female workers who left their jobs in the first year after giving birth and a 50% reduction in the first five years (Miller, 2020). For women without access to family paid leave, 30% quit their job within a year after giving birth and one in five did not return to employment over 10 years (Miller, 2020).

In 1993 the US enacted the Family and Medical Leave Act (FMLA). This is a federal labor law that was established to help lessen the detriment that workers experience when they have family emergencies and have to take time away from work. Maternity leave is one of the conditions that meet this criteria. Eligible employees are able to take up to 12 weeks of unpaid leave within a 12-month period for certain circumstances such as the birth of a child, childcare within the first year of birth, placement or subsequent care of an adopted or foster child, care for a relative with a serious health condition, or a serious health condition that can prevent the employee from performing his or her job's essential functions (French, 2019). FMLA guarantees that if a parent takes unpaid leave from work to take care of a newborn, their employed position will be protected, and their health benefits will continue while on leave. After the twelve weeks are over, they will be able to resume their job at an equal level of opportunity as was offered before the leave.

Although FMLA was passed with the good intentions of protecting employees needing time away from work for the reasons listed above, it is an imperfect solution. As noted above, even though FMLA ensures job security, the FMLA protected time away from work is unpaid. Masters explains that there are other restrictions to qualifying for FMLA. There are many stipulations for eligibility such as a required year of working at least 1,250 hours with the employer, and that the employee needs to continue making payments to the company plan for health insurance.

Additionally, FMLA only covers one person in a couple per company, you cannot be one of the highest earners within your company, and your employer may contest leave in certain circumstances (Masters, 2022). She explains that only about 60% of American workers qualify for FMLA benefits.

Vol. 2023 (2) 154 – 160
ISSN 2688-3651 [online]

2. Research Objectives

Two maternal case studies are examined to identify factors that would make taking maternity leave difficult. In the first case, the mother takes FMLA maternity leave, but as an unpaid benefit. This will allow us to find stressors under this situation. In the second case we find a mother who was not eligible for FMLA maternity leave, and instead used limited vacation time for a short leave. These cases will help us to identify gender, health, and socioeconomic stressors related to these two case studies that can be broadly applicable to the US population.

3. Methods

This qualitative study analyzes the experience of 2 different women who faced challenges with accessing paid leave. By doing a qualitative study, first-hand experience that discusses why one may or may not take maternity leave is able to be obtained which allows for factors in decision making to be narrowed down. Using two experiences pulled from the internet, these women not only talk about their decision to take maternity leave, but also how their decision affected their day to day lives and financial situations. Marie, a high school teacher in New York, took unpaid leave after giving birth but had to return to work before she was ready and resulted in a difficult and unstable childcare situation for her and her family. Jessica Shortall, a social entrepreneur, discusses how unpaid maternity leave leads to gender inequality and economic issues due to the working mothers not being able to work at their full potential, and as a result, sometimes choosing to leave the workforce. She mentions how society hides the struggles of working mothers behind a perfect facade and puts immense pressure on mothers to be able to handle both working and taking care of a baby. She tells the story of Tara, the sole earner for her family who struggles with postpartum depression and caring for a toddler and a newborn during her short, allotted maternity leave from work. The experiences of these two women, Marie and Tara, will help to analyze the struggles of maternity leave.

4. Results

4.1 Case Study 1: Marie

Marie is a 33-year-old high school teacher for 9 years, living in New York. She has two children: a 4-year-old son, Max, and newborn daughter, Olive. Marie shared her story with the International Labour Organization to make people aware of the difficulties that come with maternity leave. After giving birth to her second child Marie wanted to stay home and care for her newborn child, but she was not able to take paid maternity leave or sick leave because she had used up her days while caring for her first child. Fortunately, the job that Marie held offered unpaid maternity leave, securing her job even though she would not be paid for the time away (similar to FMLA leave).

Taking unpaid maternity leave meant that her growing family would be living off only a single income for a few months, putting her family in a difficult financial situation. It also placed her partner under additional pressure because he had to work overtime to generate additional income to provide for his family, instead of spending time with his newborn.

Unfortunately, Marie was only given 12 weeks to return before her health insurance would be put at risk. If she didn't feel ready to go back, she would be putting herself at risk for additional health problems such as postpartum depression. If she decides to take more time to recover, she risks compromising her health insurance and job position. Going back to work also meant Marie needed assistance with her family, however, childcare was an obstacle for Marie's family. It was expensive and caused her to be late to work daily due to having to drop her son to daycare every morning. Her work productivity was severely less than her peers due to this stress and she consistently would come in late or not at all.

Marie was lucky that her coworkers supported her by "providing (her) with space to pump breast milk" and covered her if she had conflicts with childcare, but even she felt that she received "no help or support from the government and social systems." For Marie, "breastfeeding is a job" and having to return to work dampened her ability to feed her child. Pumping can be very irregular and unexpected, so if it occurred while she was teaching, it could have disrupted her job, students, and it was also emotionally exhausting for her.

Vol. 2023 (2) 154 – 160
ISSN 2688-3651 [online]

What Marie experienced isn't unique and is something that many women deal with after giving birth. The lack of paid maternity leave has a direct correlation to gender, wage, and experience gaps. In traditional gender roles, childcare is often delegated to the mother. When women take this time off, they are at risk of earning lower wages and being passed up for promotions. The downstream effect of this was the risk of losing her job due to having limited time to recover from giving birth. Since Marie is a teacher, this affected the education that she was able to provide to her students, causing all of them to be indirectly affected by her limited maternity and healthcare benefits.

4.2 Case Study 2: Tara

Tara is a working mother with two children, a 2-year-old son and a newborn daughter. Tara was a manager for a small business, one that offered no paid maternity leave and is exempt from FMLA. On January 29, 2016 Tara gave birth to her second child and began her maternity leave. Tara's husband had a medical condition that prevented him from working therefore their family relied on Tara's income. Money was tight from medical expenses and the cost of providing for a full family on a single income. Tara's maternity leave consisted of 20 days, all of them from rolled over vacation days. Tara's company offered no paid leave, and she is ineligible for FMLA due to the low employee count at her company. Even if she were eligible, Tara would not have been able to take the unpaid leave because her "family can't afford the loss of even one paycheck." The doctors and nurses that attended to Tara during her C-section birth were shocked to hear about her situation and were worried about how returning to work so fast would affect Tara's health. Experts recommend 2 to 3 weeks of complete rest following a C-section surgery. Returning to work too quickly can be extremely dangerous considering that it is a major surgery. Tara's doctors also suggested that mothers should stay at home for 3 to 6 months following the birth of a child.

Ten days before returning to work, Tara emailed her bosses to let them know her expected date of return and mentioned "doing that gave me anxiety". Tara was "pretty depressed" and "feeling stressed about catching up at work" upon her return. She had a follow up appointment with her doctor to check for issues, both physically and mentally, but she said that regardless of the result of the appointment, "even if the docs checked it out and didn't think I was ok I'd still have to go back for the paycheck." This was a major health risk, causing her to later have to take time off from work and putting her in a worse situation. Tara was diagnosed with postpartum depression, a common problem for women after giving birth. Tara had less than 3 weeks leave. Increasing maternity leave from less than 6 weeks to at least 8 weeks is associated with a decline in depressive symptoms of 11% and a decline in depressive symptoms of 15% for at least 12 weeks. (Staehelin et al, 2007) In addition, Tara didn't feel "ready" to return to work, but she had no other option. She was exhausted from working all day and staying up all night to feed her kid. As a result, her work productivity was significantly lower, putting her job at risk. She was also feeling upset from the loss of meaningful time with her kids. With Tara at work and her husband busy caring for 2 young children, their household chores such as cleaning and cooking suffered.

5. Discussion

Maternity leave is an important time for a mother and a newborn child that can be beneficial for the health of the mother and early childhood development. Internationally maternity leave has been viewed as an essential component for parent and child and has been implemented as national policies abroad. In the US, however, there is not a universal policy for paid maternity leave. As an alternative, the US instituted FMLA which can be used in instances such as the birth of a child. However, FMLA is not a fool-proof solution as alluded to previously.

FMLA is unpaid leave, placing significant stress on families who might not be able to afford being without a paycheck or salary for that time. This especially hits hard for low-wage workers. As seen in Tara's case study, she was unable to take unpaid leave due to the fact that her income supported the family. Furthermore, her small business did not have a paid maternity leave policy. Tara suffered from postpartum depression and her efforts at work were not as efficient due to her exhaustion.We see here that if the US had a national policy for paid maternity leave, this would have afforded Tara better postpartum health benefits, improved bonding with her child, and improved health of the baby (Abrams 2022 and Coombs 2021). Employees who work part-time or have low-wage jobs are those for whom

Vol. 2023 (2) 154 – 160
ISSN 2688-3651 [online]

access to paid leave is most severely limited; since paid leave is limited, FMLA would provide at least some job security if they take maternity leave. Among the bottom 25% of low-paid workers, only 9% had access to paid family leave in 2020 (Coombs, 2021). People who get paid leave are much more likely to be affluent, well educated, and white, which only strengthens the racial inequities, rather than reducing this disparity (Abrams, 2022). People of color are much more likely to experience additional stressors during parenthood. The U.S. Bureau of Labor Statistics data indicates that about 47% of white parents, 41% of black parents, and just 23% of hispanic parents have access to paid leave (Abrams, 2022).

Even in cases where a mother qualifies for FMLA, this act only helps a small percentage of people. This is due in part to the fact that many mothers cannot afford to be on unpaid leave for 12 weeks since they have to continue paying their previous financial obligations as well as the new obligations that come with having a child. Women may choose to limit their maternity leave in order to provide a steady income, which can lead to long-term problems from the absence of safe and necessary recovery of the body and bonding with the newborn. This is particularly true when the mother has a Cesarean section or a complicated natural childbirth. In Marie's case we noted that she was able to take maternity leave, however this was taken as unpaid leave. Given her family situation, this creates a significant stress in caring for her 2 children and places a significant burden on her husband to work overtime. Additionally, she felt rushed to return to work after 12 weeks, even though she did not feel ready, out of fear of losing her health benefits and employment. While FMLA protects one's employment, it only does so for up to 12 weeks. This finite and generalized length of time may be short of what new families need (Abrams 2022).

Gender equity is diminished in the absence of paid parental leave policies. Traditional gender roles have created the divide between how women and men are expected to bear the load of caring after children. We saw this in the case of Marie when her husband worked overtime while she was on unpaid maternity leave. Often, women would reduce their employment or their wages due to the expectation that they would be the primary caretaker. This leads to the gender pay and opportunity gap in many fields of employment.

Congress has consistently been reluctant to expand FMLA or enact a new paid leave act, restricting this progress. One reason for this is that socially conservative lawmakers oppose such policies because they are advocates for limiting the government's role in family affairs. Another reason is the difficulties in organizing a campaign for FMLA expansion that will gain enough support and attention. The most compelling explanation for why FMLA is not being broadened in the United States, according to Sholar, is that businesses often oppose these mandates since they fear the loss of profit or the ability to stay competitive. However, research has suggested the contrary. The expansion of FMLA will not harm these businesses and will actually help them form a stronger work environment and relationship with their employees by increasing communication and preventing women from leaving the workforce. Since businesses have such powerful lobby groups in the United States, the legislators often protect the business' concerns, even if they are unfounded (Ray, Gornick, & Schmitt, 2008). Looking at the experiences other countries have had implementing flexible parental leave with pay and job protection, it appears to be feasible and not harmful to businesses. Furthermore, it promotes gender equality, makes the work environment stronger, and results in more profit.

The financial status of a person can affect their ability to take unpaid maternity leave and as a result, leads to gender, health, and financial inequities. As Marie and Tara described in their recollection of their experiences, the lack of support from government systems put their families at risk of financial instability, created an unhealthy environment, put their jobs at risk, and has affected the relationship they are able to have with their children. Marie reflected on her struggles with childcare, and the difficult position she was put in when she had to make the choice between good health and healthcare. Tara described her postpartum depression and having to return to work after a major surgery because her family could not afford for her to miss a day of work. The ability to take time to recover after an emotional and exhausting experience should not be treated as a privilege but more so as a basic right. These women are in dangerous states, both mentally and physically, yet they are being treated as machines who do not need to heal. Something has got to change.

6. Conclusion

The US has not enacted a paid maternity leave policy due to pressure from lobbyists and resistance from

Vol. 2023 (2) 154 – 160
ISSN 2688-3651 [online]

employers. Because of this, gender equity is diminished, and traditional gender roles after the birth of a child still contribute to the wage gap between men and women. The lack of a national paid maternity leave policy also creates further disparity between the lower economic class and the middle and upper classes. The health of a mother and her child and the bond they form in early childhood should not be determined by socioeconomic class, but it should be a protected privilege embraced by our national government. As a nation we could benefit from this policy as it would reduce stress, reduce mortality, improve perinatal health of the mother and child, and empower families to spend time with each other and not worry about job security. Paid leave policies have so many more benefits than those unique to the mother and baby, they can strengthen the country overall.

References

Abrams, Z. (2022). The urgent necessity for paid parental leave. *Monitor on Psychology*, 53(3), 50-56. https://www.apa.org/monitor/2022/04/feature-parental-leave

Booth, J. & Adams, R. (2023, January 30). *How Maternity Leave Affects Your HealthI.* Forbes Health. https://www.forbes.com/health/family/how-maternity-leave-affects-health

Chatterji, P. & Markowitz, S. (2004). *Does the Length of Maternity Leave Affect Mental Health?* National Bureau of Economic Research. http://www.nber.org/papers/w10206

Failing its Families: Lack of Paid Leave and Work-Family Supports in the US. (2011, February). Human Rights Watch. https://www.hrw.org/news/2011/02/23/us-lack-paid-leave-harms-workers-children

French, E. (n.d.) *What is FMLA? Demystifying Maternity Leave and Family Leave Laws.* Career Contessa. https://www.careercontessa.com/advice/what-is-fmla/

Hyde, J. S. et. al. (1995, June 1). *Maternity Leave and Women's Mental Health*. Psychology of Women Quarterly. 19(2), 257-285. https://doi.org/10.1111/j.1471-6402.1995.tb00291.x

Masters, M. (2018, January 22). *Everything New Mothers Need to Know About Maternity Leave.* Mommy Knows Best. https://mommyknowsbest.com/blogs/blogs/everything-new-mothers-need-to-know-about-maternity-leave

Masters, M. (2022, March 10). *What every mom needs to know about maternity leave.* What to Expect. https://www.whattoexpect.com/pregnancy/maternity-leave

McGovern, P., et. al. (1997, May). *Time Off Work and the Postpartum Health of Employed Women*. Medical Care. 35(5), 507-521.

Miltenburg, A. S., et. al. (2016). Maternity care and Human Rights: what do women think? *BMC International Health and Human Rights*, 16(17), 1-10. https://doi.org/10.1186/s12914-016-0091-1

Miller, S. (2020, January 15). *Paid family leave, on the rise, helps women stay in the workforce*. SHRM. https://www.shrm.org/resourcesandtools/hr-topics/benefits/pages/paid-family-leave-helps-women-stay-in-the-workforce.aspx

Coombs, S. (2021, May). *Paid Leave Is Essential for Healthy Moms and Babies*. National Partnership For Women And Families. Retrieved January 3, 2023, from https://www.nationalpartnership.org/our-work/health/moms-and-babies/paid-leave-is-essential-for.html

Ray, R., Gornick, G., & Schmitt, J. (2008, September). *Parental leave policies in 21 countries: Assessing Generosity and Gender Equality*. Center for Economic and Policy Research. http://www.cepr.net/documents/publications/parental_2008_09.pdf

Ruhm, C. J. (2000, November). *Parental Leave and Child Health.* Journal of Health Economics. 19(6), 931-960.

Vol. 2023 (2) 154 – 160
ISSN 2688-3651 [online]

Sholar, M. A. (2016, November). *The history of family leave policies in the United States.* The American Historian. https://www.oah.org/tah/issues/2016/november/the-history-of-family-leave-policies-in-the-united-states

Shortall, J. (2016, March 3) *Life in the Only Industrialized Country Without Paid Maternity Leave.* The Atlantic. https://www.theatlantic.com/business/archive/2016/03/no-paid-maternity-leave/470889

Shortall, J. (2015, November 4). *The American case for paid maternity leave.* YouTube Tedx Talks. https://www.youtube.com/watch?app=desktop&v=SJyE40koQyA

Staehelin, K., Coda Bertea, P., Zemp Stutz, E. (2007) Length of maternity leave and health of mother and child – a review. *International Journal of Public Health, 52(4), 202-209.*

US: Lack of paid leave harms workers, children. (2011, February 23) Human Rights Watch. Retrieved January 3, 2023, from https://www.hrw.org/news/2011/02/23/us-lack-paid-leave-harms-workers-children

USA: Unpaid maternity leave - the harsh reality for many working women. (2014, March 6). YouTube. Retrieved January 3, 2023, from https://www.youtube.com/watch?app=desktop&v=ME64x4YxlZQ

Journal of Research High School

Vol. 2023 (2) 161 – 165
ISSN 2688-3651 [online]

A Comparative Study on the Regulation of Inflammation by Lycium Barbarum Polysaccharide and Exocarpium Citri Grandis

Runmei Li[1] *

[1]Experimental High School Attached to Beijing Normal University, Beijing, China
*Corresponding Author: llcrym@163.com

Advisor: Ning Tao, Institute of Biophysics, tao@ibp.ac.cn

Received December 18, 2022; Revised July 20, 2023; Accepted, August 25, 2023

Abstract

Inhibition of inflammation has important clinical value, and the present study was proposed to compare the effects of *Lycium Barbarum* Polysaccharide (LBP) with those of *Exocarpium Citri Grandis* (ECG) in the regulation of inflammation. In this study, bacterial lipopolysaccharide (LPS) was used to induce the production of inflammatory factors by macrophage cell line MSC-2 cells to construct a cell model of inflammation. LBP and ECG extracts were used to act on this inflammatory cell model, and qPCR was used to detect the changes of inflammatory factors IFN-γ, TNF-α and IL-6. The results showed that LBP down-regulated IFN-γ and IL-6, while TNF-α increased in the inflammatory cell model; the extract of ECG down-regulated all the above three cytokines. In regulating the secretion of inflammatory factors, ECG may be more advantageous, while LBP is recommended to be used with caution for inflammatory diseases in the presence of high TNF-α level.

Keywords: Exocarpium Citri Grandis, inflammation, MSC-2 cells.

1. Introduction

Inflammatory diseases involve complex interactions between pathogens, cells and medications targeted to control inflammatory processes(Nathan, 2002). Infectious inflammation is one of the most common causes of inflammatory diseases, which could possibly result in worldwide epidemic such as COVID-19. The pathogenesis of infectious inflammation is the secretion of large amounts of inflammatory factors by immune cells. Cytokines play an important role in the immune response by regulating cells and transmitting signals. However, immune hyperactivation can lead to cytokine storms, which in turn triggers acute systemic inflammatory effects and secondary organ dysfunction that may ultimately result in death due to multiple organ failure(Fajgenbaum & June, 2020).Current treatment strategies for inflammatory conditions include utilization of antibiotics, adrenal glucocorticoids and monoclonal antibodies (e.g., TNF-a monoclonal antibody, Trastuzumab), but the use of antibiotics may lead to issues like reduced immunity or secondary infection, and monoclonal antibodies are expensive.

Among well-known herbs in traditional Chinese medications, *Lycium barbarum* and *Exocarpium Citri Grandis* (ECG) are often described as "antipyretic" and "anti-inflammatory and analgesic". *Lycium barbarum*, also called goji berry, is commonly used in healthcare products because of its multiple pharmacological efficacies. The main component of goji berry that accounts for its potential anti-inflammatory effects is Lycium Barbarum Polysaccharide (LBP)(Kwok et al., 2019; Tian et al., 2019). Meanwhile, the medicinal effects of ECG were recorded in the "Supplements to Compendim of Materia Medica" written by Zhao Xuemin, an herbalist of the Qing Dynasty, stating that it "treats phlegm very effectively." In the Qing dynasty, Li Yongcui in his work "evidence of the treatment of the Huijin" described the disease "phlegm" as "wet, fluid ..... what phlegm has, if the external wind, summer and dry heat

Vol. 2023 (2) 161 – 165
ISSN 2688-3651 [online]

invasion ... ...Qi and blood are disturbed, then phlegm is born." This corresponds to mucus and phlegm in inflammatory types such as bacterial pneumonia and Covid(Henig & Kaye, 2017; Ochani et al., 2021). Ancient Chinese medicine practitioners, through their continuous accumulation of experience, justified that the pharmacological efficacy of ECG could be effective in treating inflammation, which provided great inspiration for this study. Meanwhile, in the "COVID Treatment Protocol (Trial Version 6)" published by National Health Commission of the People's Republic of China in February 2020, ECG was included in the recommended prescription for "dampness and toxicity in the lung" used in the general clinical treatment period, suggesting the medicinal value of ECG in the treatment of inflammatory diseases. In this research, whether Lycium Barbarum and ECG can regulate the secretion of inflammatory factors by activated macrophages is examined through experiments.

2. Materials and Methods

Lycium Barbarum Polysaccharide: provided by Shandong Xinnuo Food Company (Zhongning County, Ningxia). Exocarpium Citri Grandis: provided by Guangdong zhanjiang Juhong hall company. Extraction method: Take 5g of dehydrated fruits of Citrus aurantium crushed, material-liquid ratio 1:10, in 80% ethanol solution in 55 ℃ water bath for 4h, to obtain Citrus aurantium extract, the main component is flavonoids. Diluted 160 times with culture medium when used(Hu, 2017). MSC-2 (immortalized MDSCs isolated from the spleens of immunosuppressed BALB/c mice, using a retrovirus encoding the v-myc and v-raf oncogenes) is provided by Francois Ghiringhelli, Department of Medical Oncology, Center GF Leclerc, 2100 Dijon, France. LPS: Sigma, USA.

Modeling procedure: 50,000 MSC-2 cells were inoculated in 6-well plates and cultured for 24 h. LPS solution was added at a final concentration of 0.25 μg/ml. LBP was dissolved in DMEM medium to make 100μg/mL LBP solution and filtered to remove bacteria. The LBP extract was filtered and diluted 160 times with DMEM medium. 10 μL and 20 μL of 100 μg/mL LBP solution and 10 μL and 20 μL of ECG extract were added to 4 wells of the six-well plate, respectively, where LBP solution has been added. The medium was then incubated for 24 hours. Real-time fluorescence quantitative PCR (qPCR) was used to detect cytokine expression. Cells in six-well plates were taken, RNA was extracted by Trizol method, transcribed into cDNA, and the relative amounts of TNF-α, INF-γ, IL-6, and the original mRNA template of the internal reference gene DAPDH were examined respectively.

3. Results

The results showed that the expression of three genes, TNF-α, INF-γ and IL-6, in the LPS control group was higher than that in the blank control group, confirming the success of the inflammation model construction. On this basis, the expression of TNF-α was reduced in the ECG extract groups compared with the LPS control group, but it did not show a significant dose-dependence; while the expression of INF-γ was all reduced to zero, and the expression of IL-6 was reduced, in which the reduction effect was significantly stronger in the high-dose group than in the low-dose group.

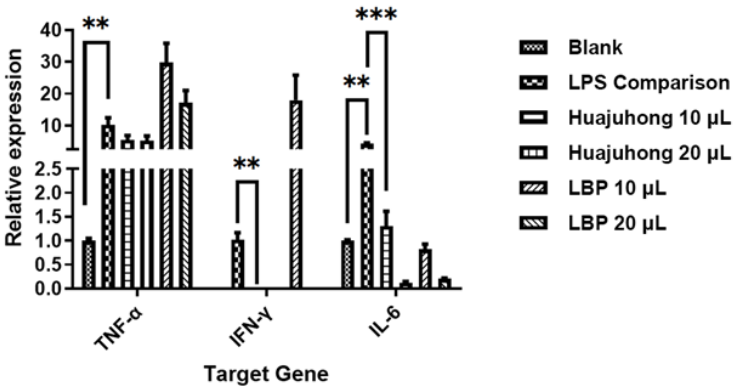


Figure 1. Relative expression of TNF-α, INF-γ and IL-6 in 6 groups. (Huajuhong indicates ECG extracts in this figure)

The expression of TNF-α in the LBP group was higher than that in the LPS control group, and the increase was greater in the low-dose group than in the high-dose group; the expression of INF-γ was higher in the low-dose group than in the control group; the expression of IL-6 was lowered, and the lowering

Vol. 2023 (2) 161 – 165
ISSN 2688-3651 [online]

effect was significantly stronger in the high-dose group than in the low-dose group, showing some similarity with that of the ECG extract group. Since three samples from each treatment group were tested using qPCR, error bars in the figure were determined using the three results from the according group. Stars between groups indicate the significant differences between data groups, which were determined by p value calculated using t test. P value that is smaller than 0.01 is marked by two stars, and p value smaller than 0.001 is marked by three stars.

4. Discussion

Modern medicine considers inflammation as a spontaneous defense response of the body to inflammatory factors such as bacteria and viruses, a process by which the immune system tries to suppress inflammatory pathogens through multiple pathways, usually manifesting as redness, swelling, heat, pain and dysfunction.

TNF-α, the predominant inflammatory factor, induces increase of multiple inflammatory factors expression and leads to ischemia and thrombosis(Idriss & Naismith, 2000). Inappropriate or excessive activation of TNF-α signaling is associated with chronic inflammation and may eventually lead to the development of pathological complications, such as autoimmune diseases(Jang et al., 2021). IFN-γ induces phagocytosis of blood cells by macrophages, leading to hemocytopenia in patients with inflammation. IL-6 is also is a pro-inflammatory cytokine that increases antibody production and induces acute phase reactants(Fajgenbaum & June, 2020).

How to reduce the expression of inflammatory factors is a key issue in clinical treatment. In severe pneumonia, TNF-α and IFN-γ acting together can lead to inflammatory cell death, requiring the use of antibodies to inhibit the secretion of inflammatory factors(Karki et al., 2021). Trastuzumab is effective in reducing the effects of IL-6(Tanaka, Narazaki, & Kishimoto, 2014), but it has the problem of high price. Therefore, drugs that can effectively reduce these major inflammatory factors are urgently needed to be discovered.

Present studies showed that 70% ethanol-extracted ECG components exhibited great anti-cough, anti-phlegm and anti-inflammatory effects when applied to ammonia-induced cough model in mice, phenol red excretion in mice and xylene-guided ear swelling model in mice(Jiang et al., 2014). In addition, in mice models of air pollution particle-induced pulmonary inflammation, ECG flavonoids significantly inhibited PM2.5 -stimulated overproduction of TNF-α, IL-1β, IL-6, and IL-18 and increased the numbers of white blood cell, neutrophils, lymphocytes, and monocytes in bronchoalveolar lavage fluid of the model mice(Zhu et al., 2019). The anti-inflammatory effects of ECG flavonoids were also studied using LPS-induced RAW264.7 cell model(Hu, 2017). The MSC-2 cells used in this study are characterized by a high degree of differentiation and some inflammatory properties(Apolloni et al., 2000). Studies using this cell type can provide in-depth validation of the anti-inflammatory potency of ethanolic extracts of ECG and provide experience for future studies.

In this study, the effects of LBP and ECG extracts on the regulation of TNF-α, IFN-γ and IL-6 were investigated. The results showed that ECG extract had a significant reduction effect on IL-6 and IFN-γ, but the reduction of TNF-α was relatively limited and did not meet the needs of clinical treatment well. LBP treatment group showed small increase in TNF-α, and the reduction of IFN-γ only existed in groups with high LBP concentration (20 μL). These variations of amount changes among different cytosines could be caused by the difference in structure of LBP and ECG extracts, since LBP is a type of polysaccharide while ECG extracts are mainly composed of flavonoids. Whereas their ways of acting might be similar because of the shared effect of reducing IL-6. Exact mechanisms of their regulatory effects still need further research.

Therefore, the combination of ECG and TNF-α monoclonal antibodies may have better anti-inflammatory effects and has application potential. In addition, some traditional Chinese medicine formulas also suggest the possibility of combined use of ECG and other herbal medicines. For example, in the "Pulmonary Disease System Pill" recorded in the "Beijing Selected Traditional Chinese Medicine Formulas", the combination of ECG, shiso, licorice and poria was used together. The licorice extract can reduce the expression of TNF-α(Yang et al., 2015), and platycoside E in poria can reduce the expression of TNF-α and IL-1β(Ji et al., 2020). The good combination of these herbs can reduce the expression of TNF-α, IFN-γ, IL-6 and other cytokines, which can well compensate for the deficiency of ECG in reducing TNF-α.

Vol. 2023 (2) 161 – 165
ISSN 2688-3651 [online]

5. Conclusion

Both LBP and ECG embodied some anti-inflammatory effects, and the ECG extract showed stronger effects in reducing TNF-α and IFN-γ. The experimental results confirm that the active ingredients of these two herbs can inhibit the secretion of inflammatory factors to a certain extent, but the different inhibitory effects on different types of inflammatory factors also indicate that a single type of herbal extract is not effective in treating inflammatory diseases. This suggests that future research could investigate the medical potential of different species of plant extracts used together.

Acknowledgment

I thank my research advisor, professor Ning Tao, for his expertise and assistance throughout all aspects of my study and for his help in writing the manuscript.

References

Apolloni, E. et al. (2000). Immortalized myeloid suppressor cells trigger apoptosis in antigen-activated T lymphocytes. *J Immunol, 165*(12), 6723-6730. doi:10.4049/jimmunol.165.12.6723

Fajgenbaum, D. and June, C. H. (2020). Cytokine Storm. *N Engl J Med, 383*(23), 2255-2273. doi:10.1056/NEJMra2026131
Henig et al. (2017).

Henig, O. and Kaye, K. S. (2017). Bacterial Pneumonia in Older Adults. *Infect Dis Clin North Am, 31*(4), 689-713. doi:10.1016/j.idc.2017.07.015

Hu, M. J. (2017). *The Anti-inflammatory Effect and Mechanism of Flavonoids in Exocarpium Citri Grandis on LPS-Induced RAW264.7 Cells.* (Master). Huazhong Agricultural University, Available from Cnki

Idriss, H. T. and Naismith, J. H. (2000). TNF alpha and the TNF receptor superfamily: structure-function relationship(s). *Microsc Res Tech, 50*(3), 184-195. doi:10.1002/1097-0029(20000801)50:3<184::AID-JEMT2>3.0.CO;2-H

Jang, D. et al. (2021). The Role of Tumor Necrosis Factor Alpha (TNF-alpha) in Autoimmune Disease and Current TNF-alpha Inhibitors in Therapeutics. *Int J Mol Sci, 22*(5). doi:10.3390/ijms22052719

Ji, M. et al. (2020). The Pharmacological Effects and Health Benefits of Platycodon grandiflorus-A Medicine Food Homology Species. *Foods, 9*(2). doi:10.3390/foods9020142

Jiang, K. et al. (2014). Antitussive, expectorant and anti-inflammatory activities of different extracts from Exocarpium Citri grandis. *J Ethnopharmacol, 156*, 97-101. doi:10.1016/j.jep.2014.08.030

Karki, R. et al. (2021). Synergism of TNF-alpha and IFN-gamma Triggers Inflammatory Cell Death, Tissue Damage, and Mortality in SARS-CoV-2 Infection and Cytokine Shock Syndromes. *Cell, 184*(1), 149-168 e117. doi:10.1016/j.cell.2020.11.025

Kwok, S. S. et al. (2019). A systematic review of potential therapeutic use of Lycium barbarum polysaccharides in disease. *BioMed research international, 2019*.

Nathan, C. (2002). Points of control in inflammation. *Nature, 420*(6917), 846-852. doi:10.1038/nature01320

Vol. 2023 (2) 161 – 165
ISSN 2688-3651 [online]

Ochani, R. et al. (2021). COVID-19 pandemic: from origins to outcomes. A comprehensive review of viral pathogenesis, clinical manifestations, diagnostic evaluation, and management. *Infez Med, 29*(1), 20-36. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/33664170.

Tanaka, T., Narazaki, M., & Kishimoto, T. (2014). IL-6 in inflammation, immunity, and disease. *Cold Spring Harb Perspect Biol, 6*(10), a016295. doi:10.1101/cshperspect.a016295

Tian, T. et al. (2019). Extraction, structural characterization, and biological functions of Lycium barbarum polysaccharides: A review. *Biomolecules, 9*(9), 389.

Yang, R. et al. (2015). The Pharmacological Activities of Licorice. *Planta Med, 81*(18), 1654-1669. doi:10.1055/s-0035-1557893

Zhu, Z. et al. (2019). Effects of Total Flavonoids from Exocarpium Citri Grandis on Air Pollution Particle-Induced Pulmonary Inflammation and Oxidative Stress in Mice. *J Food Sci, 84*(12), 3843-3849. doi:10.1111/1750-3841.14966

Journal of Research High School

Vol. 2023 (2) 166 – 177
ISSN 2688-3651 [online]

The Influence of Food Allergies on Health-Related Quality of Life

Isabel Louie ,[1] *

[1]Sacred Heart Academy, Hempstead, NY, USA
*Corresponding Author: ilouie514@gmail.com

Advisor: Dr. Stephen J. Sullivan, stephensullivan@sacredheartacademyli.org

Received February 15, 2023; Revised August 6, 2023; Accepted, September 13, 2023

Abstract

Since the 1990s, the prevalence of food allergies has grown. Individuals with food allergies must prepare for potentially life-threatening allergic reactions. Previous research has shown that food allergies contribute to increased anxiety because of avoidance of allergens, ultimately contributing to the health-related quality of life of both the patient with the allergy and the parents of a child with a food allergy. It is necessary to investigate the impact of health-related quality of life with an inclusive demographic of participants, a component past studies lack, to better create food allergy management strategies. The influence of food allergies on health-related quality of life was closely examined through the distribution of three surveys: a general health-related quality of life questionnaire based on a number of reported unhealthy days, and a food allergy-specific health-related quality of life adolescent and parental burden questionnaire. 113 adolescents and 274 parents completed these surveys after providing informed consent. Adolescents with a food allergy had fewer overall unhealthy days than non-allergic adolescents, but reported a greater number of unhealthy mental and physical health days. Adolescents with long- term health impairments reported a greater number of overall unhealthy days than adolescents with food allergies. Parents of a child with a food allergy reported experiencing a larger quantity of overall and physically unhealthy days than parents of a child without a food allergy, though no significant conclusion could be drawn regarding the connection between the parental burden associated with food allergies to health-related quality of life.

Keywords: Allergies, Adolescent, Health-Related Quality of Life, Parent, Stress

1. Introduction

Food allergy (FA) prevalence has risen significantly throughout the past three decades, an epidemic and "food safety and public health concern" that impacts 31.6 million people in the United States (Centers for Disease Control and Prevention (CDC), 2020, para. 1; Platts-Mills, 2015; Food Allergy Research & Education, n.d.-a). The variety of food allergens has increased along with these rising cases (Jones, 2020), but the most common, or the "Big 9", are "milk, eggs, fish, shellfish, tree nuts, peanuts, wheat, soybean, and sesame" (U.S. Food and Drug Administration (FDA), 2022, para. 4).

FA and food intolerance are often unclearly distinguished in research (Teufel *et al.*, 2007). It is therefore necessary to first differentiate the two. As a result of a FA, consuming a particular food protein, or allergen, causes the immune system to overreact (Food Allergy Research & Education, n.d.-b), whereas food intolerance is a digestive issue. Secondly, the term "health-related quality of life" (HRQOL) will be used in this study, referring to the effect quality of life has on one's "perceived physical or mental health over time" (CDC, 2018a, para. 7).

Vol. 2023 (2) 166 – 177
ISSN 2688-3651 [online]

2. Literature Review

The combination of increasing prevalence of FA and the lack of treatment options to completely resolve it leaves patients with FA to cope with a constant fear of allergic reaction, even to the point of experiencing anaphylaxis, a potentially deadly reaction. They must stay perpetually vigilant. In addition, food allergens are difficult to clearly identify on food labels, for trace amounts of allergens may be found in packaged foods. This contributes to the stress and confusion of patients with a FA (Athas, 2019). Furthermore, Congress does not yet require all "Big 9" allergens to be listed on ingredient labels. For example, identifying the presence of sesame will not be mandatory until 2023. This contributes to already high distrust of food labels and the need for consumers to engage in "risk assessments" (FDA, 2022; Athas, 2019). Moreover, children with FA often display increased anxiety, particularly when engaging in social activities such as eating at restaurants and attending birthday parties. Their parents often experience guilt for their child's anxiety (Cimons, 2021). Dietary restrictions, a common method of avoiding accidental consumption of food allergens, have also been found to contribute to poor HRQOL in adolescents with FA (Nowak-Wegrzyn *et al*., 2021). Moreover, caring for a child with FA costs families billions of dollars annually (FARE, n.d.-a), creating potential financial strain that may induce increased stress levels in another way. This can also undermine mental health. Ultimately, these factors influence the HRQOL of both the patient with a FA and their caregivers.

Numerous factors, such as FA development and novel treatments, have been investigated. However, HRQOL remains insufficiently researched. The CDC highlights the importance of investigating HRQOL, as this would promote better improvement of public health and provide a means to determine the burden of certain medical conditions (CDC, 2018b). Additionally, limited studies have investigated HRQOL, using a FA-specific scale rather than a general HRQOL questionnaire (Cummings, 2010). Unfortunately, although FA-specific HRQOL questionnaires have been developed, until recently, few had been statistically validated with a broad sample. The EuroPrevall project, which sought to create management strategies for patients with a FA, resolved this issue and successfully validated several FA-specific questionnaires for children, teenagers, and parents (Fernandez-Rivas *et al*., 2015). Two of these instruments were used in this study. These questionnaires have been effectively employed since Nowak-Wegrzyn *et al.* (2021) utilized the FA Quality of Life Questionnaire-Teenager Form to measure the HRQOL of adolescents with FA, and Franxman *et al*. (2015) utilized the FA Quality of Life Questionnaire-Parental Burden Form to conclude that caregivers of patients with a confirmed FA reported poor HRQOL.

An additional issue with current FA research is the narrow demographic band of participants. For instance, Cohen *et al.* (2004) sought to validate a disease-specific HRQOL questionnaire to determine the parental burden related to caring for a child with a FA, yet 96% of participants were mothers, 90.8% were White, and nearly half earned over $100,000, demonstrating a severe lack of gender, racial, and economic diversity within its participants. A study by Nowak-Wegrzyn *et al.* (2021) demonstrated similar problems in terms of its racial demographic, with a parent sample that was 69.4% White. However, these researchers *did* contribute to the limited research available regarding the connection between feelings of parental burden and caring for a food- allergic child, concluding that patients with peanut allergies and their caregivers self-reported poor HRQOL. Moreover, equal gender representation is lacking throughout the current research. In King *et al.* (2009), only 35% of food-allergic participants were female. In Penner Protudjer *et al*. (2016), only 31% of participants were female. It is vital to rectify these percentages, as females with FA are more likely to exhibit poorer HRQOL than males (Penner Protudjer *et al.,* 2016).

However, some suggest that other facets of FA research should be more thoroughly investigated before devoting significant time to HRQOL, given recent progress in treatment options. For instance, oral immunotherapy (OIT), which gradually exposes patients to a food allergen to increase their threshold for allergen consumption without a reaction, has shown promising results. Additionally, Palforzia, now a FDA-approved drug, can be used in the OIT process (Zarif, 2020). Some scholars propose strictly focusing on creating treatment options to completely resolve FA, such as immunotherapy vaccinations, approaches still undergoing clinical trials (Zarif, 2020). Despite these significant developments, HRQOL research remains equally important, for it promotes improved understanding of both the physical and mental burden associated with having a FA.

Thus, this study aims to determine how the presence of a FA influences the HRQOL of food-allergic adolescents and parents of a child with a FA compared to a non-food-allergic population, with a specific focus on increasing the

Vol. 2023 (2) 166 – 177
ISSN 2688-3651 [online]

gender diversity of participants. Within the non-food-allergic adolescent population, the HRQOL of participants with health impairments lasting over 8 months will be compared to the HRQOL of the food- allergic adolescent participants to determine whether a FA is comparable to other long-term health impairments, such as anxiety. Identifying variations between the HRQOL of these populations could aid in the development of effective management strategies for food-allergic patients and their families, with a specific focus on mental well- being. We are especially interested in whether chronic stress, a factor of HRQOL, may exacerbate existing and future health complications (Wright, 2005).

2.1 Hypotheses

It is hypothesized that:

1. Adolescents with a FA will report poorer HRQOL than adolescents without a FA and adolescents with long-term health impairments.
2. Adolescents with a FA will report feeling troubled, frightened, and disappointed by their FA, according to the FA-specific adolescent HRQOL questionnaire.
3. Parents of a child with a FA will report poorer HRQOL than parents of a child without a FA.
4. Parents of children with a FA will report feelings of limitation and fright as a result of their child's allergy, according to the FA-specific parental burden questionnaire.

3. Method

3.1 Participants

113 adolescents, 14-18 years of age, participated. Subjects all hailed from New York's Nassau, Suffolk or Queens counties, with the majority drawn from an all-girls Catholic high school located in Hempstead, NY (mean age = 16.39, standard deviation = 1.02). Participants were recruited via email and through snowball sampling, in which they were asked to share the surveys with other potential study participants. Of the adolescent participants, 5.3% identified as male, 92.9% identified as female, and 1.8% chose not to disclose their gender. The sample mostly aligned with the ethnic makeup of the school, though the percentage of Hispanic/Latina participants was greater (see Table 1).

Table 1. High School Ethnicity Statistics

Ethnicity	High School Population**	Nassau County	United States	Sample %
White	74%	72.2%	75.8%	66.7%
Black	10%	13.1%	13.6%	9.8%
Hispanic / Latino	4%	17.6%	18.9%	13%
Asian / Pacific Islander	11%	11.9%	6.1%	10.6%
Other	-	2.1%	2.9%	-

*Totals may not equal 100 due to those who reported themselves as multiracial.
**2018-19 data provided by Sacred Heart Academy Annual 2019 Report to NYSED + 2021 U.S. Census Estimate (United States Census Bureau, 2021)

274 parents, 22-83 years of age, participated. All participants also resided in Nassau, Suffolk, or Queens (mean age = 37.46, standard deviation = 11.76). 62% of parents/guardians identified as male and 38% as female. The sample reflected a greater percentage of Asian/Pacific Islander participants and a smaller percentage of Black and Hispanic/Latino participants in relation to the demographics of Nassau County or the United States (see Table 2). Additionally, although some parents were related to adolescent participants because their child shared the parent survey with them, the majority of parents had no relation to adolescent participants. The responses of any adolescents with parents who also happened to complete the survey were not linked to their parents' responses in any way.

Vol. 2023 (2) 166 – 177
ISSN 2688-3651 [online]

Table 2. Adult Ethnicity Statistics

Ethnicity	Nassau County	United States	Sample %
White	72.2%	75.8%	72.5%
Black	13.1%	13.6%	3.4%
Hispanic / Latino	17.6%	18.9%	4.1%
Asian / Pacific Islander	11.9%	6.1%	14.8%
American Indian / Alaska Native	.6%	1.3%	3.4%
Other	2.1%	2.9%	1.4%

*Totals may not equal 100 due to those who reported themselves as multiracial.
**2021 data provided by 2021 U.S. Census Estimate (United States Census Bureau, 2021)

Among the 113 total adolescent participants, 18 reported having a FA. This is approximately twice the 8% FA rate for children in the United States (CDC, 2020). Of these adolescents, whose ages ranged from 15-18 (mean = 16.5, standard deviation = 1.1), 5.6% identified as male and 94.4% identified as female. 68.4% identified as White, 10.5% identified as Black/African American, 5.3% identified as Hispanic/Latino, and 15.8% identified as Asian/Pacific Islander. Within the 274 total parent participants, 150 reported having a school-aged child with a FA. Their ages ranged from 22-67 (mean = 33.68, standard deviation = 10.), and 69.3% identified as male and 30.7% identified as female.

FA support groups, including Food Allergy Support and Education (FASE), were contacted to share the surveys within the food-allergic population of adolescents on Long Island and in Queens, as well as within the population of parents of children with FA. Amazon Mechanical Turk, a virtual survey-distribution platform, was also utilized, which provided monetary compensation for completion of the parent survey. Compensation was supported by a grant from the New York Institute of Technology to support student public health research. The survey participants accessed was titled "Health-Related Quality of Life-Student Form" or "Health-Related Quality of Life-Parent Form," and all participation was on a volunteer basis.

3.2 Design

This study utilizes a between-subject experimental design. The independent variable is the presence of FA in an adolescent participant, and the dependent variable is the HRQOL of participants, as well as the subscales of the FA-specific questionnaires (troublesome, frightened, disappointed, and limited). Adolescent participants completed their appropriate survey online via Google Forms after parent/guardian consent, and their assent was provided at the beginning of the form. Parents/Guardians provided their consent before completing their own appropriate form.

3.3 Measures

FA-Specific HRQOL.

The Food Allergy Quality of Life Questionnaire-Teenager Form (FAQLQ-TF) (Flokstra-de Blok *et al*., 2008) is a 23-item questionnaire designed to determine the HRQOL of adolescents with a FA. The survey utilizes a 7-point scale, from 0 (not) to 6 (extremely), and includes 4 domains: how troublesome the FA, with a focus on social situations (questions 1-12), how troublesome the FA is in regard to the packaging of foods (questions 13-18), how frightening the FA is (questions 19- 21), and 2 questions concerning discouragement and disappointment resulting from the FA (questions 22- 23). Table 3 displays a sample question from each domain.

Table 3. FAQLQ-TF Sample Questions - Flokstra-de Blok et al. (2008)

Subscale	**Sample Question**
Troublesome 1	How troublesome do you find it, because of your food allergy, that you must always be alert as to what you are eating?
Troublesome 2	How troublesome is it, because of your food allergy that you have to explain to people around you that you have a food allergy?
Frightened	How frightened are you because of your food allergy of an allergic reaction?
Disappointed	How disappointed are you when people do not take your food allergy into account?

Vol. 2023 (2) 166 – 177
ISSN 2688-3651 [online]

Parental Burden.

The Food Allergy Quality of Life-Parental Burden Questionnaire (FAQLQ-PB) (Cohen, B. L. et al., 2004) is a 17-item questionnaire that details the HRQOL of parents of children with a FA. This survey utilizes a 7-point Likert scale, with 0 indicating "not limited/troubled" and 6 indicating "extremely limited/troubled," and consists of two subscales. The first 3 questions, part of the first subscale, detail limitations parents/guardians face with a food-allergic child, and the following 14 questions of the second subscale detail how troubled they are for their child. Table 4 displays a sample question from each subscale.

Table 4. FAQLQ-PB Sample Questions – Cohen, B. L. et al. (2004)

Subscale	Sample Question
Limitation	If you and your family were planning a holiday/vacation, how much would your choice of vacation be limited by your child's food allergy?
Troublesome	In the past week, how troubled have you been by your concerns for your child's health because of their food allergy?

General HRQOL

The Healthy Days Measure (CDC, 2018b) is a 14-item survey detailing a participant's general HRQOL by evaluating the number of unhealthy days a participant had within the past 30 days before completing the survey[1]. It contains the Healthy Days Core Module (overall health), the Activity Limitations Module (mental health), and the Healthy Days Symptoms Module (physical health). The Healthy Days Module contains questions 1-4, the Activity Limitations Module contains questions 5-9, and the Healthy Days Symptoms Module contains questions 10-14. Question 1 is based on a 5-point scale, where 1 represents "excellent," and 5 indicates "poor." Questions 2, 3, 4, 10, 11, 12, 13, and 14 are fill-in questions in which participants indicate a number of unhealthy days. It should be noted that question 14 was reverse scored, as it originally asked for a number of healthy days. Questions 5, 8, and 9 are "yes/no" questions, question 6 allows participants to select an answer, and question 7 provides ranges of time to select. Table 5 displays a sample question from each module. All surveys utilized in this study are in the public domain.

Table 5. The Healthy Days Measure Sample Questions - CDC (2018b)

Module	Sample Question
Healthy Days Core Module	During the past 30 days, for about how many days did poor physical or mental health keep you from doing your usual activities, such as self-care, work, or recreation?
Activity Limitations Module	During the past 30 days, for about how many days have you felt worried, tense, or anxious?
Healthy Days Symptoms Module	During the past 30 days, for about how many days have you felt very healthy and full of energy?

3.4 Procedure

Surveys were first distributed at the beginning of 2022 (winter), followed by additional distribution during fall (2022). All surveys began with a consent form briefly explaining the purpose of the project, the time necessary to complete the survey, with a promise of anonymity and minimal risk in participation. Contact information for myself, my advisor, the science department chair, and the school social worker were also provided for those with questions. Consequently, the adolescent survey required electronic parent/guardian consent and student assent before beginning the questionnaire, followed by basic demographic information. All households in the single-sex school community are notified by robocall and email about upcoming student surveys. All student participants completed the Healthy Days Measure (CDC, 2018b). They were then asked to read the provided definition of "food allergy" (to differentiate with food intolerances) before responding if they had a clinically diagnosed FA. The language used is as follows:

> "Food Allergy: "An allergy occurs when your body's natural defenses overreact to exposure to a particular substance, treating it as an invader and sending out chemicals to defend against it" (American College of Allergy, Asthma, & Immunology, n.d.). Some symptoms include vomiting, hives, dizziness, and even anaphylaxis. Note that this is NOT a food intolerance, which is an issue the body has in digesting particular

Vol. 2023 (2) 166 – 177
ISSN 2688-3651 [online]

foods (e.g., lactose intolerance, gluten intolerance)."

The form displayed a debriefing page explaining the goal of the study and requesting that participants refrain from discussing their responses with others to maintain confidentiality and unbiased results, and was then submitted for those who responded "no." Those who answered "yes" were directed to complete the FAQLQ-TF (Flokstra-de Blok *et al*., 2008) to determine their FA-specific HRQOL. Finally, the participant stated their specific FA and was shown the debriefing form before submitting their response. Prior to data analysis, adolescents without FA were grouped into three categories based on their response to question 7 of the Healthy Days Measure: Group 1 included those who reported a health impairment lasting 0-3 weeks, Group 2's impairment lasted 1-7 months, and Group 3's impairment lasted over 8 months.

Parent participants completed a separate Google Form, first providing their own consent to participate, followed by demographic questions. All parents/guardians completed the Healthy Days Measure (CDC, 2018b) to determine their general HRQOL. The definition of "food allergy" was provided before indicating if their child has a FA. Those who responded "no" were shown the debriefing form and submitted their form. Those who answered "yes" were directed to complete the FAQLQ-PB (Cohen, B. L. *et al*., 2004), were then shown the debriefing page before their form was submitted.

Unpaired *t*-tests were used to investigate Hypotheses 1 and 3. In the adolescent sample, the responses to the Healthy Days Measure (CDC, 2018b) from those without FA were compared to the responses of those with FA to determine the differences in the two groups' HRQOL. The Healthy Days Measure (CDC, 2018b) responses from parents without a child with FA were also compared to the responses of parents with a food-allergic child to determine the differences in HRQOL. The *t*-test allowed for the comparison of the means of unhealthy days within the adolescent group and the means within the parent group, and allowed for accurate conclusions to be drawn regarding the differences in each group. The *t*-test also yielded *p*-values, or "probability" values. These offered insight into whether these conclusions were statistically significant and could be utilized to effectively answer the hypotheses. $P > .05$ was deemed statistically insignificant, $p < .05$ was deemed significant, $p < .01$ offered strong significance, and $p < .001$ was deemed the most significant.

Correlations tested for linear relationships to examine Hypotheses 2 and 4. This test yielded an R^2 value, the correlation coefficient, and a *p*-value was also calculated. The R^2 value reveals the variance between the independent and dependent variables tested. To test for the correlation between the presence of FA and poor HRQOL, the responses of the FAQLQ-TF (Flokstra-de Blok *et al*., 2008) and the Healthy Days Measure (CDC, 2018b) were compared. To test for the correlation between being a parent of a food-allergic child and having poor HRQOL, the FAQLQ-PB (Cohen, B. L. *et al*., 2004) and the Healthy Days Measure (CDC, 2018b) were compared.

4. Results

A series of unpaired *t*-tests were run to determine if the mean differences between each groups' Healthy Days Measure were significant. Adolescents without FA from the initial winter subsample were found to suffer from poorer HRQOL than adolescents with FA in terms of unhealthy mental and physical health days ($p<.05$). Adolescents without FA also reported a greater average number of overall unhealthy days, with an average of 22.2 days compared to the 16.36 days reported by adolescents with a FA, but this subscale was deemed only marginally significant ($p = .057$). (See Table 6, columns 1 and 2.)

Table 6. Means (Standard Deviation) of the Adolescent Healthy Days Measure

	Winter 2022 Subsample (N=80)		Fall 2022 Complete Sample (N=113)	
	FA	No FA	FA	No FA
Number of Overall Unhealthy Days	16.36 (11.72)[a]	22.2 (9.95)[a]	16 (11.2)[a]	21.6 (9.38)[a]
Number of Mentally Unhealthy	16.25 (12.54)[a]	23.2 (9.69)[a]	36.5 (14.14)[A]	23.59 (9.73)[B]
Number of Physically Unhealthy Days	20.7 (11.83)[a]	27.6 (6.47)[a]	48.7 (23.72)[A]	26.49 (7.79)[B]

Means with different superscripts (*lowercase*) differ at 95% confidence level. ($p<.05$)
Means with different superscripts (*uppercase*) differ at 99% confidence level. ($p<.01$)

Vol. 2023 (2) 166 – 177
ISSN 2688-3651 [online]

Table 7. Healthy Days Measure and FAQLQ-TF Correlation

	Winter 2022 Subsample (N=14)		Fall 2022 Complete Sample (N=18)	
	R^2	*p-value*	R^2	*p-value*
Troublesome 1/Overall Health	.197	.112	.076	.268
Troublesome 1/Mental Health	.197	.112	.158	.103
Troublesome 1/Physical Health	.282	.019	**.415**	**.004**
Troublesome 2/Overall Health	.222	.089	.144	.120
Troublesome 2/Mental Health	.229	.076	.230	.044
Troublesome 2/Physical Health	**.448**	**.009**	**.429**	**.003**
Frightened/Overall Health	.234	.080	**.389**	**.006**
Frightened/Mental Health	.146	.178	**.523**	**.001**
Frightened/Physical Health	.425	**.012**	**.453**	**.002**
Disappointed/ Overall Health	.179	.132	**.420**	**.004**
Disappointed/ Mental Health	.16	.156	**.557**	**<.001**
Disappointed/ Physic al Health	.229	.030	**.600**	**<.001**

Boldface type signifies robust and statistically significant data, with R^2 >.4 and p<.05.

In the complete fall sample, which included data from the previous winter subsample as well as additional responses solicited during September of the same year, adolescents with FA were found to report a greater average number of unhealthy mental and physical health days (p<.001). Adolescents without FA continued to demonstrate a greater number of overall unhealthy days compared to their food allergic counterparts (p<.05). (Table 6, columns 3 and 4.) Group 3 adolescents had poorer HRQOL, in terms of their overall unhealthy days (*M* = 26.75, *SD*= 6.73), than food-allergic adolescents, but reported fewer unhealthy mentally (*M* = 24.75, *SD* = 9.48) and physically (*M* = 24.56, *SD* = 8.73) unhealthy days. 81.8% of Group 3 participants reported having anxiety/depression, which are mental health conditions.

Among the Winter 2022 subsample, moderate correlations were found between physical health and feelings of trouble and fright among participants with FA, subscales of the FAQLQ-TF (Nowak-Wegrzyn *et al.*, 2021) and the Healthy Days Measure (CDC, 2018b). Within the complete Fall 2022 sample, moderate correlations were identified between physical health and feelings of trouble and fright, while a stronger correlation was identified between physical health and feelings of disappointment. Additionally, overall health and mental health were found to correlate with feelings of fright and disappointment (see Table 7).

The following Table 8 details the specific food allergies reported by adolescent participants, with the most common being peanut and tree nut allergies.

Table 8. Adolescent Reported FA

Allergy	Percentage
Milk	3.0%
Eggs	9.1%
Fish	3.0%
Shellfish	9.1%
Tree Nuts	24.2%
Peanuts	24.2%
Wheat	3.0%
Sesame	12.1%
Other	15.0%

Totals may not equal 100 due to those who reported having multiple FA.

Table 9. Mean (Standard Deviation) of the Parent Healthy Days Measure

	Winter 2022 Subsample (N=43)		Fall 2022 Complete Sample (N=231)	
	FA	No FA	FA	No FA
Number of Overall Unhealthy Days	7 (9.0)[a]	14.76 (10.57)[a]	**19.5 (12.38)***[A]**	**12.68 (11.38)***[B]**
Number of Mentally Unhealthy	9 (11.17)*[a]	13.79 (11.13)*[b]	**17.8 (11.47)***[A]**	**23.42 (10.42)***[B]**
Number of Physically Unhealthy Days	20.2 (12.3)*[a]	22.7 (9.88)*[b]	**27.09 (6.6)***[A]**	**22.7 (9.71)***[B]**

*p<.05; **p<.01; ***p<.001
Means with different superscripts (*lowercase*) differ at 95% confidence level. (p<.05)
Means with different superscripts (*uppercase*) differ at 99% confidence level. (p<.001)

Vol. 2023 (2) 166 – 177
ISSN 2688-3651 [online]

Within the Winter 2022 subsample, a series of unpaired *t*-tests found that parents of a child without a FA had poorer HRQOL in terms of overall health (M = 14.76, SD = 10.57), than parents of a child with a FA (M = 7, SD = 9, p<.05). Though parents of a child without a FA also reported a greater number of average unhealthy mental (M = 13.79, SD = 11.13) and physical (M = 22.7, SD = 9.88) health days, these results were deemed statistically insignificant, with *p*-values of .241 and .512, respectively (see Table 9, columns 1 and 2). The Fall 2022 Complete Sample, however, differed; parents with a child with a FA reported a greater number of overall (M = 19.5, SD = 12.38) and physically (M= 27.09, SD = 6.6) unhealthy days (p<.001). Parents of a child without FA displayed a greater average number of mentally unhealthy days (p<.001). (See Table 9, columns 3 and 4.)

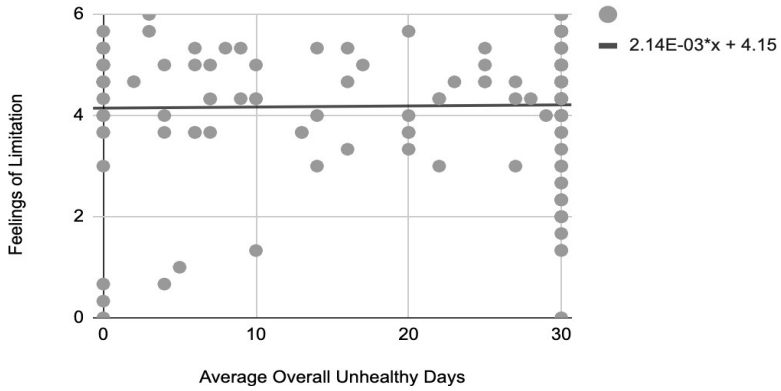


Figure 1. Fall 2022 Complete Sample Parent Overall Unhealthy Days vs. "Limitation" Correlation

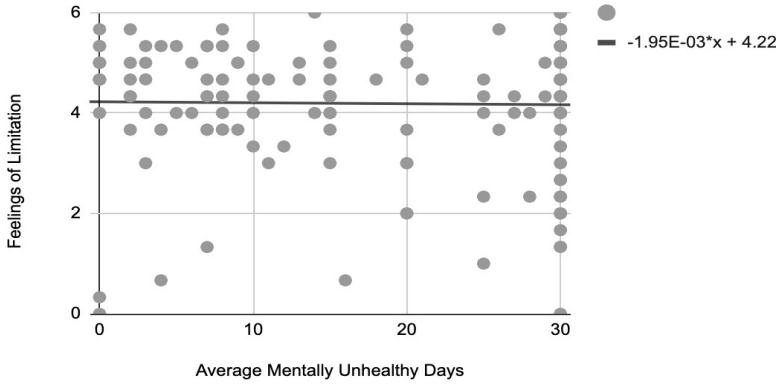


Figure 2. Fall 2022 Complete Sample Parent Mentally Unhealthy Days vs "Limitation" Correlation

A significant conclusion could not be drawn in either sample regarding the correlation between HRQOL and the parental burden experienced by parents of a child with a FA (p>.05), as displayed in Table 10. Numerous outliers in data were also identified (see Figures 1-3).

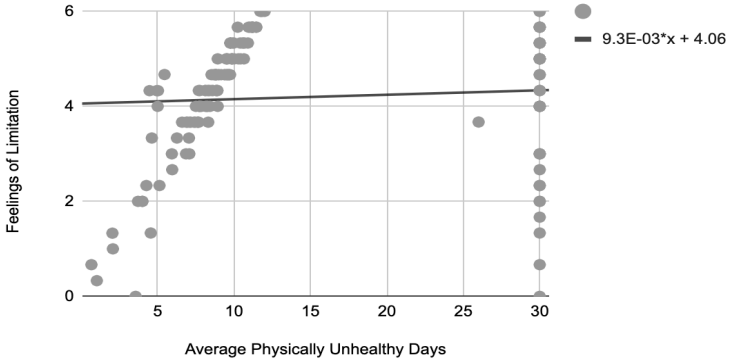


Figure 3. Fall 2022 Complete Sample Parent Physically Unhealthy Days vs "Limitation" Correlation

5. Discussion

Contrary to hypotheses 1 and 3, adolescents from the Winter 2022 subsample with a FA and parents of a child with a FA reported fewer average unhealthy days, and therefore, had better HRQOL than their allergy-free counterparts. Additionally, hypothesis 4 was unsupported, as significant feelings of parental burden could not be determined. Hypothesis 2 was partially supported in that unhealthy *physical* health days were related to feelings of trouble, fright, and disappointment; however, unhealthy overall and mental health days were not significantly associated to these feelings.

In the Fall 2022 complete sample, however, hypothesis 1 was better supported, as adolescents with a FA *did*

Vol. 2023 (2) 166 – 177
ISSN 2688-3651 [online]

Table 10. Healthy Days Measure and FAQLQ-PB Correlation

	Winter 2022 Subsample (N=10)		Fall 2022 Complete Sample (N=150)	
	R^2	*p-value*	R^2	*p-value*
Limitation/Overall	.048	.560	<.001	.808
Limitation/Mental Health	.01	.926	<.001	.836
Limitation/Physical Health	.017	.716	.005	.395
Troubled/Overall	.230	.160	.030	.034
Troubled/Mental Health	.287	.110	.008	.290
Troubled/Physical Health	.042	.570	.027	.045

exhibit poorer HRQOL in terms of both mentally and physically unhealthy days than those without FA. Additionally, hypothesis 2 was supported, as correlations were identified between physical health with feelings of disappointment, overall health with fright and disappointment, and mental health with fright and disappointment. Regarding hypothesis 3, parents of a child with a FA *did* display poorer HRQOL than parents with a child without FA in terms of a) overall and b) physically unhealthy days. Correlations from this complete sample also supported parts of hypothesis 4. Physical health, overall health, and mental health were all significantly linked with various components of the FAQLQ-PB.

The initial Winter 2022 results came as a surprise, as they proposed conflicting findings to previous research. Nowak-Wegrzyn *et al.* (2021), for instance, concluded that avoidance of allergens, a theme within the FAQLQ-TF, *did* contribute to poor HRQOL among adolescents with a FA. The Fall 2022 complete sample, however, better supported Nowak-Wegrzyn's insights, though slight variations remained. For instance, only certain categories of the Healthy Days Measure were associated with feelings of fright and disappointment among adolescents with FA. Moreover, Franxman *et al.* (2015) found that parents of a child with a FA reported poor HRQOL, while the present study could not draw significant conclusions regarding the HRQOL of parents of a child with a FA in either sample. These differing results ultimately display the necessity of continued FA-specific HRQOL research to better determine the physical and mental burden of having a FA on both the adolescent and their parents.

A source of these variations in results may stem from the demographics of participants in this study. Compared to the demographics of Cohen *et al.* (2004), in which 96% of parent participation was derived from mothers, over half of parent participants in this study were fathers (62%). Additionally, because of unequal participation of female and male adolescents within past FA research, female adolescents were intentionally oversampled, a potential contributing factor to the conclusions formulated herein. Unlike the 35% participation of females in King *et al.* (2009) and the 31% in Penner Protudjer *et al.* (2016), 92.9% of adolescent participants in this study self-identified as female. The demographic of participants allowed for improved gender diversity in this research, and allowed for a historically underrepresented population to become the focus of a novel FA study. In addition, a FA-specific questionnaire, the FAQLQ-TF, was successfully utilized to determine the HRQOL of adolescents with a FA, filling the gap in current FA-specific HRQOL research, which generally lacks the use of FA-specific HRQOL questionnaires. Furthermore, this study successfully addressed a goal for future FA-specific HRQOL research proposed by Teufel *et al.* (2007), as FA and food intolerances were clearly distinguished to participants in the questionnaires.

The findings also demonstrate the seriousness of continuing HRQOL research, both within the FA community and beyond. In particular, 27 of the 33 adolescents who reported long-term health impairments in question 6 of the Healthy Days Measure noted in question 7 that their major impairment was anxiety/depression. There is a dire need for confronting the state of HRQOL of adolescents today so that in the future, their mental and physical health may improve.

Moreover, it is important to consider the differences between the HRQOL of these adolescents with long-term anxiety/depression (the majority of Group 3) and adolescents with FA. Group 3's responses to the Healthy Days Measure (CDC, 2018b) reveal that they experience poorer overall HRQOL than adolescents with FA, while food-allergic adolescents reported a greater number of mentally and physically unhealthy days. Based on these differences, it is therefore evident that various long-term health impairments differ in the way they influence overall, physical, and mental health. A potential reason why FA appears to affect mental and physical health to a greater extent may derive

Vol. 2023 (2) 166 – 177
ISSN 2688-3651 [online]

from the nature of the condition, which revolves around constant hyperawareness of food allergens. Those with FA must constantly avoid consumption of these allergens lest they have a potentially deadly allergic reaction (Lange, 2014), which impacts their physical health. This undoubtedly creates a constant stressor in their lives, influencing their mental health.

5.1 Limitations

The large discrepancy between the number of adolescent participants without a FA and with a FA proved to be a significant limitation. The limited number of participants from the food-allergic population directly impacted the calculated *p*-values and r^2 values, which are highly dependent on sample size. Attaining adolescent participants with a FA was quite challenging, despite reaching out to various FA support groups and employing a snowball sampling method. Ethical concerns prevented direct payment to minor subjects, but our IRB permitted use of a university grant to "hire" adult survey-takers through Amazon's well-regarded Mechanical Turk. Thus, a larger sample size of food-allergic participants would allow for more significant conclusions to be drawn.

As previously mentioned, a snowball sampling technique was utilized to increase the sample size of adolescents with a FA and parents of a child with a FA. Though the benefit of circulating the questionnaires through participants was gained, participation was no longer truly random, and the possibility of greater sampling bias was introduced.

The use of Amazon Mechanical Turk also influenced sampling in that "semi-professional" survey-takers look to maximize their income per minute spent on each survey. Because the title of the survey (appropriately) mentioned food allergies, Mechanical Turk subjects skewed heavily towards parents with food-allergic children. If a greater number of participants read the entirety of the description provided, more would have seen we needed control subjects as well.

5.2 Future Research

The findings of this study spark numerous questions to be explored in future research. One possible direction of any additional study would be to investigate whether having *multiple* food allergies more significantly influences HRQOL. This would be an interesting avenue to consider, as it was proven that having at least one FA *does* contribute to feelings of trouble, fright, and disappointment. Additionally, researching how the HRQOL of adolescents with FA compares to autoimmune disorders would allow for a direct comparison of different chronic illnesses, and the comparison of the extent of the burden they may create.

Additionally, the majority of adolescent participants identified as female within this study. Although this allowed a population to be present in a field of research that has not adequately reflected them in the past, it would be beneficial to create a study that has an equal number of male and female adolescent participants. By making this a focus of a future study, findings may be applied to *all* those who suffer from FA, without gender bias.

Moreover, improving racial diversity in future works would allow for a wider, more inclusive demographic to be created, which better reflects the overall population. Prospective research may also aim to incorporate focus groups, as they would provide additional reasoning behind participant responses to the questionnaires administered, allowing for the better overall understanding of the HRQOL of a particular population.

References

American College of Allergy, Asthma, & Immunology. (n.d.). *Food Allergies | Causes, Symptoms & Treatment*. ACAAI. https://acaai.org/allergies/allergicconditions/food/#:%7E:text=Conditions%20Food%20Allergy-,Food%20Allergy,chemicals%20to%20defend%20against%20it.

Athas, E. (2019, January 28). Which Allergens Are in Your Food? You Can't Always Tell From the Labels. *The New York Times*. https://www.nytimes.com/2019/01/28/health/food-allergies-children-labeling.html

Vol. 2023 (2) 166 – 177
ISSN 2688-3651 [online]

Centers for Disease Control and Prevention. (2018, October 31a). *HRQOL Concepts*. CDC. https://www.cdc.gov/hrqol/concept.htm

Centers for Disease Control and Prevention. (2018, October 31b). *Healthy Days Core Module: HRQOL-14 Measure | HRQOL | CDC*. https://www.cdc.gov/hrqol/hrqol14_measure.htm

Centers for Disease Control and Prevention. (2020, June 8). *Food Allergies | Healthy Schools | CDC*. Centers for Disease Control and Prevention. https://www.cdc.gov/healthyschools/foodallergies/index.htm

Cimons, M. (2021, August 14). Fears over their food allergies can lead kids to shun normal childhood events. *The Washington Post*. https://www.washingtonpost.com/health/kids-food-allergy-anxiety/2021/08/13/9053c1b4-f491-11eb-9738-8395ec2a44e7_story.html

Cohen, B. L. *et al.* (2004). Development of a questionnaire to measure quality of life in families with a child with food allergy. *Journal of Allergy and Clinical Immunology*, *114*(5), 1159–1163. https://doi.org/10.1016/j.jaci.2004.08.007

Cummings, A. J. *et al.* (2010). The psychosocial impact of food allergy and food hypersensitivity in children, adolescents and their families: a review. *Allergy*, *65*(8), 933–945. https://doi.org/10.1111/j.1398- 9995.2010.02342.x

Fernández-Rivas, M. *et al.* (2015). The EuroPrevall outpatient clinic study on food allergy: background and methodology. Allergy, 70(5), 576–584. https://doi.org/10.1111/all.12585

Flokstra-de Blok, B. M. *et al.* (2008). Development and validation of the self-administered Food Allergy Quality of Life Questionnaire for adolescents. *Journal of Allergy and Clinical Immunology*, *122*(1), 139–144.e2. https://doi.org/10.1016/j.jaci.2008.05.008

Food Allergy Research & Education. (n.d.-a). *Facts and Statistics*. FARE. https://www.foodallergy.org/resources/facts-and-statistics

Food Allergy Research & Education. (n.d.-b). *What Is a Food Allergy?* Retrieved September 5, 2021, from https://www.foodallergy.org/resources/what-food-allergy

Franxman, T. J. *et al.* (2015). Oral food challenge and food allergy quality of life in caregivers of children with food allergy. The journal of allergy and clinical immunology. In practice, 3(1), 50–56. https://doi.org/10.1016/j.jaip.2014.06.016

Jones, R. (2020, November 25). *Why food allergies are on the rise*. BBC Future. https://www.bbc.com/future/article/20201023-food-allergies-why-nut-dairy-and-food-allergy-are- rising

King, R. M., Knibb, R. C., & Hourihane, J. O. (2009). Impact of peanut allergy on quality of life, stress and anxiety in the family. *Allergy*, *64*(3), 461–468. https://doi.org/10.1111/j.1398-9995.2008.01843.x
Lange, L. (2014). Quality of life in the setting of anaphylaxis and food allergy. *Allergo Journal International*, *23*(7), 252–260. https://doi.org/10.1007/s40629-014-0029-x

Nowak-Wegrzyn, A. *et al.* (2021). The Peanut Allergy Burden Study: Impact on the quality of life of patients and caregivers. *World Allergy Organization Journal*, *14*(2), 100512. https://doi.org/10.1016/j.waojou.2021.100512

Penner Protudjer, J. L. *et al.* (2016). Impaired health-related quality of life in adolescents with allergy to staple foods. *Clinical & Translational Allergy*, 6, 1–9. https://doi.org/10.1186/s13601-016-0128-5
Platts-Mills, T. A. (2015). The allergy epidemics: 1870–2010. *Journal of Allergy and Clinical Immunology*, *136*(1), 3–13. https://doi.org/10.1016/j.jaci.2015.03.048

United States Census Bureau. (2021). *U.S. Census Bureau QuickFacts: Nassau County, New York*. U.S. Census Bureau. https://www.census.gov/quickfacts/nassaucountynewyork

Vol. 2023 (2) 166 – 177
ISSN 2688-3651 [online]

United States Census Bureau. (2021). *U.S. Census Bureau QuickFacts: United States*. U.S. Census Bureau QuickFacts: United States. https://www.census.gov/quickfacts/fact/table/US/PST045221

Teufel, M. *et al.* (2007). Psychological burden of food allergy. *World journal of gastroenterology*, *13*(25), 3456–3465. https://doi.org/10.3748/wjg.v13.i25.3456

U.S. Food and Drug Administration. (2022, January 31). *Food Allergies*. https://www.fda.gov/food/food-labeling-nutrition/food-allergies

Wright, R. (2005). Stress and atopic disorders. *Journal of Allergy and Clinical Immunology*, *116*(6), 1301– 1306. https://doi.org/10.1016/j.jaci.2005.09.050

Zarif A. (2020). Advances in Food Allergy Treatment. The Yale journal of biology and medicine, 93(5), 749–758.

Journal of Research High School

Vol. 2023 (2) 178 – 188
ISSN 2688-3651 [online]

The Impact of Climate Change on Animal Behavior: The Arctic Marine and Northeast Pacific Ecosystems

Cecilia Yang[1] *

[1]The Harker School, San Jose, CA, USA
*Corresponding Author: cecilia.m.yang@gmail.com

Advisor: Carly Scott, cbscott@utexas.edu

Received February 11, 2023; Revised August 1, 2023; Accepted, September 19, 2023

Abstract

Climate change affects ecosystems worldwide and the entire biosphere. Although the effects of climate change have been extensively studied, comparisons of its effects on animal behavior across ecosystems are lacking. To narrow this gap, this review compared the changing behavior of marine mammals in the Northeast Pacific ecosystem to that in the Arctic marine ecosystem. This review highlighted the similarities and differences in the migration, predation, distribution, and reproductive behavior of selected animals in response to climate change. This review focused on marine mammal species central to each ecosystem: sea otters, seals, dolphins, orcas, and whales in the Northeastern Pacific Ocean and polar bears, seals, whales, orcas, and walruses in the Arctic. By comparing the Northeastern Pacific ecosystem and the Arctic marine ecosystem, researchers can gain an in-depth understanding of the changes in animal behavior due to climate change across wide geographies and suggest a new method of observing the impact of climate change globally to enable more accurate ecosystem predictions.

Keywords: Climate change; Behavioral ecology; Animal behavior; Arctic marine ecosystem; Northeast pacific ecosystem

1. Introduction

The effects of climate change on various environments and changes in animal behavior have been extensively investigated. Findings show that climate change has destroyed animal habitats, removed food sources, and resulted in many other serious ecological consequences, as demonstrated by melting ice caps and intensifying storms (Hardy, 2003). Changing environments cause some animals to become unfit for their habitat, forcing them to either adapt or migrate elsewhere. Many species acclimate to their new environment in situ, behaving differently to survive in unfamiliar habitats. These behavioral changes in various species also affect how they interact with each other (Harmon and Barton, 2013). The chain effects caused by climate change ultimately lead to major shifts in flora and fauna populations and significant alterations to ecosystems.

Climate change has caused major alterations in biomes and ecosystems worldwide (e.g., Hardy, 2003; Moore and Huntington, 2008; Wassman et al., 2011). When climate change affects animal behavior, it causes chain effects that also impact humans. Furthermore, the study of animal behavior reveals the reasons behind species interactions and animal responses to the environment. As climate change and other challenges prove to be a threat to animal survival, their behavior may change so that they can survive. Behaviors previously beneficial in an environment can become maladaptive, leading to a population die-off if they are not adjusted. Behavioral changes drastically transform interspecies relationships and ecosystems, as any shift in the behavior of a species can lead to a larger chain of reactions (Harmon and Barton, 2013). Additionally, changes in animal behavior mediate climate change and the environment. When a species alters its behavior, the species can in turn cause drastic environmental changes, and as a result, the

Vol. 2023 (2) 178 – 188
ISSN 2688-3651 [online]

combined impact of these changes can potentially lead to an ecological catastrophe in their ecosystem. In particular, these chain reactions often form a large-scale impact on community ecology and the trophic food chain, notably in consideration of keystone species (Harmon and Barton, 2013). Generally, patterns among species responding to climate variability arise through greater human-animal interactions, organisms migrating or dispersing as a consequence of habitat loss, and often, increased aggression or sociability towards humans (Beever et al., 2017).

This article reviews the literature on climate change and animal behavior and their close relationship in the Northeast Pacific and the Arctic ecosystems. The two ecosystems have been selected for their comprehensive similarities and differences; though both are sensitive to changes in the global climate and environment, the Arctic is exceptionally vulnerable due to global warming and ice melt as the rising sea level impacts the habitats of thousands of species. By contrast, the Northeast Pacific shares similarities in that it is also a marine ecosystem with overlapping organisms and similar human impacts on its environment. However, since it is less responsive to climate change, it may serve as a comparison to the extreme reactions within the Arctic marine ecosystem. Both ecosystems face the threat of rising sea levels, warming of the sea surface, and habitat destruction. Melting ice in the northern Pacific Ocean and the Arctic Ocean threatens the habitat of species such as seals and polar bears. Seals in the Arctic marine region and Northeast Pacific require ice to reproduce; therefore, the declining ice in the two areas reduces the population of seals (Gifford-Gonzalez, 2011; Moore and Huntington, 2008). The diminishing amount of sea ice also affects polar bears' distribution and predation behavior in the Arctic marine ecosystem (Atwood, 2016).

This research emphasizes the importance of observing the connection between climate change and changes in animal behavior and the major similarities in animal reactions to a changing environment in these ecosystems. Moreover, the differences between the responses of organisms in the two ecosystems allow for further understanding of the mechanism behind changes in animal behavior and pave the way for potential predictions of changes in animal behavior in other ecosystems.

2. Literature Review

2.1 Northeast Pacific Ecosystem

Northeast Pacific ecosystem is along the Pacific coastline of North America, serving as a habitat for thousands of species, including sea otters, seals, dolphins, orcas, and whales. Human fishing, boating, surfing, and increased pollution impact the region and threaten organisms in the area. According to statistics from the IPCC, the Intergovernmental Panel on Climate Change, marine heatwaves in the past forty years have doubled in frequency and become longer-lasting, more intense, and more extensive. Due to climate change, the sea surface temperature in the Northeast Pacific is rising rapidly (Chavez et al., 2017). Storm winds, particularly during the winter storm period, are intensifying, the sea level is rising, and seawater is becoming more acidic (Reiter, 2015). These factors all contribute to shifting animal behavior throughout the ecosystem, as many species are forced to adapt to their new environment.

The density of marine mammals along the North American Pacific coastline increases as the water temperature rises (Burrows et al., 2012). Throughout the Pacific Ocean, marine mammals, particularly those with limited thermal tolerances, have declined in population. Since most marine mammals cannot adapt quickly to a changing environment, those who survived the effects of the change to their habitats due to climate change had to adapt by changing their living and migratory habits. Table 1 provides a summary of animal behavioral changes in the Northeast Pacific ecosystem.

2.2 Species in Northeast Pacific

In the Northeast Pacific ecosystem, sea otters (*Enhydra lutris*), seals (*Pinnipedia* spp), dolphins (*Lagenorhynchus obliquidens*), orcas (*Orcinus orca*), and whales (*Cetaceae* spp.), the species examined in this article, are the major mammal species in this ecosystem, with otters being the keystone species. These five mammals adapt their behavior in different ways to changing environments caused by climate change. Specifically, these species primarily altered their migration, predation, distributive, reproductive, and social behavioral patterns.

Vol. 2023 (2) 178 – 188
ISSN 2688-3651 [online]

Sea otters in the Northeast Pacific ecosystem have increased diet and prey selection variations due to climate change. As a result of the fluctuation of a consistent food source, specifically, sea urchin populations, otters in the Northeast Pacific show increasing individual variability in diet choices. Many otters have sought alternative food sources, and some have demonstrated extreme inter-individual variation in their diets (Estes et al., 2003). Furthermore, their presence near the Pacific coastline sharply increased as water closer to land became warmer after 2002. Because otters prefer warmer climates, as the Pacific coastline's surface temperatures increase, the populations of sea otters also increase in such regions (Burrows et al., 2012). As a result of climate change, sea otters in the Northeast Pacific ecosystem must alter their diets and living patterns to survive, and their behavior has changed in response to the warming seas. Thus, sea otters must spread more sparsely in the northern Pacific Ocean, causing them to have an increased and more diverse habitat range.

In the Northeast Pacific region, seals, similar to sea otters, have an increased habitat range due to the warming of the northern Pacific Ocean. Similar to sea otters, elephant seals (*Mirounga angustirostris*) and harbor seals (*Phoca vitulina*) demonstrated a notable increase in their populations along the North American Pacific coastline. However, the fur seal (*Arctocephalinae* spp.) population in the Pacific Ocean decreased (Burrows et al., 2012; Hazen et al., 2013). Fur seal pups are born later and are significantly smaller due to chain effects from rising sea levels. The rising sea level leads to displaced storm and weather patterns, thereby altering the foraging trips of female fur seals and changing their breeding grounds into an environment that is much more hostile to birthing seal pups (Gifford-Gonzalez, 2011). Harbor seals have also adapted to melting ice in northern latitudes by dispersing broadly while breeding to access resources more easily (Nicholson, 2000). All three species of seals have been affected by climate change, and their behavioral patterns have changed, whether in migration, location, or birthing.

Pacific dolphin (*Lagenorhynchus obliquidens*) populations in the Northeast Pacific coastline have declined. Initially, many dolphin species populations along the coastline increased drastically around 1999; however, their regional densities began to decline after 2002. Large groups of dolphins moved toward land in search of prey, such as squid, and left when the population density of such prey declined due to mass predation (Burrows et al., 2012). The densities of Pacific white-sided dolphins in the Northeast Pacific ecosystems are directly related to the temperature of the water, causing their populations to increase with the warming of bay water (Black, 1994). Dolphins have been forced to shift their living and predatory habits to adjust to changing environments. As more dolphins migrate into coastal regions, they affect the species in the same environment, particularly those they prey on, often disrupting these species' behaviors as they reduce their populations and force new migratory patterns.

Killer whales, otherwise known as orcas, (*Orcinus orca*) have experienced population decline throughout the Pacific Ocean, much of which is due to unnatural causes such as malnutrition or reasons related to human interaction. This population decline is significantly, and negatively, impacting the population health of this species. Although killer whales are traditionally a largely migratory species, due to human interference in their environment, their natural patterns of migration and distribution have been disrupted, often dying out of their regular habitat as a result of unfamiliar tides or becoming stranded on land. In a study where 53 killer whales from various coastal regions in the Pacific North American coastline were examined, 42% demonstrated significant impacts on population health as a result of human interaction and climate change. (Raverty et al., 2020). Furthermore, as an aftermath of human contaminants spilling over into the Pacific, killer whales face an increased risk of disease and fatality from chemical pollution in their habitats (Buckman et al., 2011). As a whole, killer whales have needed to alter their migratory and distributive patterns, and consequently, their predation habits, as human interactions negatively impact killer whales themselves, as well as the habitat in which they reside.

Many whale species migrate between the Arctic marine ecosystem and the Northeast Pacific ecosystem. However, these very migration patterns have been disrupted as a result of the increasing temperatures. Although species like the North Pacific gray whale (*Eschrichtius robustus*) and the blue whale (*Balaenoptera musculus*) have always migrated to the Arctic during the summer in search of cooler environments, these species have changed their migration patterns as a result of global climate change. Most whale species that move between the Arctic and Pacific Oceans have shifted their behavior to remain in the northern ocean for much longer, rather than returning to the Pacific when the summer ends (Moore and Huntington, 2008). Furthermore, since whale species in both oceans have unnaturally modified their distributive behavior, their typical prey in warmer oceans and the Pacific have been affected, causing chain reactions

Vol. 2023 (2) 178 – 188
ISSN 2688-3651 [online]

throughout the trophic cascade and impacts throughout the food chain. In a study where baleen whales in the Northeast Pacific movements were tagged to study their migratory movement, only 29% of the whales spent their time in one familiar region, instead often opting to travel further north (Bailey et al., 2009). Since whales are a tertiary consumer, they tend to keep their prey's species' population in check; however, with their changed distributive behaviors, they instead excessively decrease their prey population in the Arctic marine ecosystem, and allow similar populations in the Northeast Pacific ecosystem to grow out of control. Thus, the changing whale species' migratory, distributive, and predation patterns affect species in both ecosystems, linking the impact climate change has on tertiary consumers to two different environments.

Table 1. Summary of animal behavioral changes in the Northeast Pacific ecosystem

	Migration	Predation	Distributive	Reproductive	Social
Sea Otters	Increased presence along Pacific coastlines	Increased diet and prey variation	Sparser spread and increased habitat range	n/a	More human-animal interactions and new interspecies interactions
Seals	Increased presence along Pacific coastlines	n/a	Increased habitat range, broader dispersal	Disrupted and hostile birthing conditions, weaker pups	More human-animal interactions and new interspecies interactions
Dolphins	Increased presence along Pacific coastlines	Increased diet and prey variation	n/a	n/a	More human-animal interactions and new interspecies interactions
Orcas	Migration outside of their natural habitats	Increased diet and prey variation	Increased habitat range, presence near humans	n/a	More human-animal interactions and new interspecies interactions
Whales	Migration outside of their natural habitats (Arctic)	Increased diet and prey variation	Increased presence in Arctic, increased habitat range	Disrupted reproduction times and patterns	More human-animal interactions and new interspecies interactions

2.3 Arctic Marine Ecosystem

The Arctic Ocean provides a home for thousands of species, including polar bears, seals, whales, orcas, and walruses, the species this article will focus on. This ecosystem is in the northernmost body of water in the world and is covered with glaciers while receiving precipitation as snow (Walsh 2008). As time passes, more and more humans enter the region for residency, tourism, and fishing. Consequently, pollution in the area increases, and marine traffic in this ecosystem affects the migration and behavior of animals. However, as a consequence of climate change, ice in the Arctic marine ecosystem melts, and the temperature of seawater is rising rapidly. Since 1980, more than 50% of permanent ice in the Arctic ecosystem has been lost and is decreasing by 13% per decade (Townhill et al., 2022). From statistics from the IPCC, ice sheets and glaciers lost mass at an average rate of 218 gigatonnes per year from 2006 to 2015. As a result, the zooplankton population has declined, altering their consumers' behavior and the rest of the marine food chain (Wassman et al., 2011).

Northward distribution shifts are forcing the positions of food sources to change and introduce new predators into the Arctic region. With the changing prey distribution, predators have to adjust their migration timing and destinations to improve their chances of survival (Davidson and Ruhs, 2021). Because so many marine mammals are migrating

Vol. 2023 (2) 178 – 188
ISSN 2688-3651 [online]

north, their new behavior will alter species interactions within the ecosystem and cause chain effects in all species in the Arctic marine ecosystem. Additionally, the melting sea ice in the region is shifting migration patterns as the literal shape of the ecosystem is changing. Table 2 provides a summary of animal behavioral changes in the Arctic Marine ecosystem.

2.4 Species in the Arctic

Polar bears (*Ursus maritimus*), seals (*Pinnipedia* spp.), whales (*Cetaceae* spp.), orcas (*Orcinus orca*), and walruses (*Odobenus rosmarus*) are the major mammal species in Arctic marine ecosystems. These five mammals are tertiary consumers, which are usually keystone species and help control the population of other species and maintain a balance in the food chains of ecosystems. Many Arctic marine mammals are ill-equipped to adapt quickly to climate change. By analyzing these species, we hope to understand how major predatory species adapt to the effects of climate change.

Polar bears have gradually adjusted their land use behaviors in response to their melting habitat in the sea. More of the polar bear population in the Arctic Ocean's southern Beaufort Sea has begun to stay onshore for longer because of melting sea ice. From 1986 to 2013, the average distribution of polar bears on the shore increased from 5.8% to 37% (Atwood et al., 2016). Polar bears spend much of their time on sea ice; however, with the diminishing amount of sea ice due to global warming, polar bears in the southern Beaufort Sea have become more dependent on their terrestrial habitat. These polar bears risk interacting with humans to access more food on land, as climate change has destroyed much of their water habitat (Atwood et al., 2016).

Similarly, various species of seals in the Arctic area are poised to migrate further north as the Arctic climate continues to warm. Harp (*Pagophilus groenlandicus*), hooded (*Cystophora cristata*), ribbon (*Histriophoca fasciata*), and spotted seals (*Phoca largha*), among others, rely on sea ice to give birth to young; without sea ice, their populations are likely to decline (Moore and Huntington, 2008). The dependence of seals on ice masses for parturition is changing and as the amount of ice on the seas is waning, seals must migrate further north or face the threat of declining reproduction rates (Laidre et al., 2008). Seal populations have slowly decreased over the past century with the shift to a warmer climate. Seals do not typically migrate or move long distances; however, they have to make this behavioral change to survive the rapidly warming climate of the Arctic marine ecosystem.

The North Pacific gray whale (*Eschrichtius robustus*) population has altered its migration timing and reproduction rates in response to major shifts in its environment due to climate change. While the amount of sea ice is reduced by global warming, whales in the Arctic marine ecosystem, including gray whales, are likely to travel further north and stay longer. In places such as the Bering Sea, many whales remain in the northern areas of the sea all year instead of migrating back south once summer is over (Moore and Huntington, 2008). Due to this change in distribution and shifted migration schedule, whales will also have to adjust their reproduction timing accordingly. These location changes may delay or even prevent whales from reproducing as usual. As a repercussion of the dramatic influence of climate change throughout this ecosystem, whales have had to completely shift their migration and living behaviors to combat habitat destruction in the ocean. Yet as whale species are forced to remain in northern areas for longer periods of time as global temperatures consistently rise, they more commonly begin to adapt in situ, finding ways to adapt to climate change in the Arctic rather than in southern oceans like the Pacific.

In the Arctic, killer whales (*Orcinus orca*) have experienced extreme changes in their distributive and migratory behaviors as ice melt, habitat destruction, warmer temperatures, and the rising sea level affect their environment. Sightings of killer whales have become more frequent closer to human residency overall; however, such sightings decline drastically during winter, suggesting that these populations opt to move even further north than they have typically gone in search of colder climates suited to their biology (Higdon et al., 2011). As a whole, killer whales, like whales, have an emerging pattern of moving more and more north as temperatures globally steadily rise, and these populations seek the temperatures that they are accustomed to. However, as killer whales leave familiar habitats for survival, they may find that such unfamiliar environments may threaten them instead. Formerly, sea ice had inhibited killer whale movement, preventing entrance into potentially dangerous areas, or regions with large amounts of human activity. As such 'choke points' open when sea ice continues to melt, killer whales may endanger themselves with

Vol. 2023 (2) 178 – 188
ISSN 2688-3651 [online]

increased human interaction with sighting reports increasing substantially since 1850 (Higdon and Ferguson, 2009). Though killer whales migrate in hopes of finding the colder environments that they are accustomed to and surviving despite climate change's extreme impacts on their habitats, these changing behaviors may prove to be fatal instead.

Walruses (*Odobenus rosmarus*) in the Arctic marine ecosystem face several threats to their predation and distributive patterns as a consequence of the sea ice melting. The recession of spring sea ice past the Arctic shelves into deeper basins causes the typically icy waters to be less of a resting reproduction environment with easy access to the walruses' benthic prey. Thus, while the reproductive chances of the Pacific walrus is reduced, its survival rate also decreases, unable to fully utilize the Arctic basins to their advantage in deeper waters (Bluhm and Gradinger, 2008). Furthermore, in some cases, as walruses overwinter in the marine Arctic, their distribution changes as their habitat shifts around them. Often, these animals are forced out of comfortable resting areas, instead having to stay in areas with restricted access to sea water, becoming an easy target for animals like polar bears. Isolated walruses outside their natural zones are extremely vulnerable to predators, and their traditionally distributive overwintering behaviors become a risk rather than a solace to these organisms (Kiliaan and Sterling, 1978). Altogether, the melting sea ice poses great threat to marine mammals like the walruses, who depend on patches of ice within the water for their typical reproductive, predation, and distributive behaviors. However, with the decreasing amounts of sea ice and habitat destruction due to climate change, these behaviors instead consistently put walruses at risk and threaten the species' survival as a whole.

Table 2. Summary of animal behavioral changes in the Arctic Marine ecosystem

	Migration	Predation	Distributive	Reproductive	Social
Polar Bears	n/a	n/a	Increased presence onshore, less presence on sea ice	n/a	More human-animal interactions and new interspecies interactions
Seals	Migration outside of their natural habitats and increasingly northern	n/a	n/a	Declining reproduction rates	More human-animal interactions and new interspecies interactions
Whales	Migration outside of their natural habitats and increasingly northern	Increased diet and prey variation	Increased presence in Arctic, increased habitat range	Disrupted reproduction times and patterns	More human-animal interactions and new interspecies interactions
Orcas	Migration outside of their natural habitats and increasingly northern	Increased diet and prey variation	Increased habitat range, presence near humans	n/a	More human-animal interactions and new interspecies interactions
Walruses	n/a	Increased diet and prey variation	Increased habitat range, presence near predators	n/a	More human-animal interactions and new interspecies interactions

3. Comparative Analysis

3.1 Similarities between the Arctic marine ecosystem and the Northeast Pacific Ecosystem

Vol. 2023 (2) 178 – 188
ISSN 2688-3651 [online]

In response to rapid climate change, marine mammals in both ecosystems have changed their behavior to survive in these new environments. To illustrate, the species examined have adjusted their migration, distribution, and predation habits, among others, to face the new challenges that climate change presents to them in their environment.

The changing climate in both ecosystems has altered many interspecific relationships and species interactions in the food chains. Many species have all encountered changes in their food sources and have adapted their diets to their new environment. Due to climate change, sea urchin populations along the Northeast Pacific coastline have fluctuated, and sea otters have shifted their predation behaviors accordingly to compensate for their unreliable food sources (Estes et al., 2003). Dolphins have adjusted their distribution patterns to search for more prey, and their densities in the Northeast Pacific have increased (Burrows et al., 2012). Similarly, polar bears have changed their distribution patterns on land and in the sea in the Arctic marine ecosystem to access more prey.

Species in both ecosystems are declining because of the effects of climate change on their environments. All the species are noticeably decreasing in population. The chain effects of their behavioral changes can be seen throughout the trophic structure of each ecosystem. Because of the rising seawater temperature and sea level, many species have altered their migration patterns. Marine mammals in the Northeast Pacific have shifted their distribution patterns in response to an increase in warmer waters of the Pacific Ocean. Predation, reproduction, and dietary behaviors are also shifting.

3.2 Differences between the Arctic marine ecosystem and the Northeast Pacific Ecosystem

Although the two ecosystems share many similarities in their topographical features and species, changes in species behavior are not always parallel because of climate differences between the two ecosystems. In the Arctic marine ecosystem, ice is a major part of many animal habitats, and the melting affects their migration, reproduction, and other patterns. In contrast, the Northeast Pacific ecosystem does not face this issue because its location is more southern. Because of this difference in habitat, animals in the Arctic marine ecosystem contend with habitat loss and alter their distribution, migratory habits, and reproductive patterns. In the two ecosystems, seals have changed their reproductive behavior in different ways. In the Northeast Pacific, seals spread over icy areas in the northern Pacific Ocean, and seals in the Arctic Ocean migrate further northward to access more ice (Nicholson, 2000; Laidre et al., 2008). Since the Pacific Ocean is much farther away from the northern sea ice and seals cannot migrate large distances, they compensate for their lack of ice by spreading sparsely across the available ice. In comparison, the seals in the Arctic marine ecosystem are significantly closer to the northern sea ice; therefore, they can migrate to access more ice.

Although climate change manifests itself in similar ways in the two ecosystems, the types of behaviors the animals alter in response vary. Specifically, while the living and eating patterns of the five species have changed in the Northeast Pacific, the five species in the Arctic marine ecosystem have mainly changed their migration patterns. In the Northeast Pacific ecosystem, sea otters have mainly changed their predation behaviors, seals have changed their reproductive behaviors, dolphins have changed their predation behavior (see Table 3 and Figure 1). By comparison, in the Arctic marine ecosystem, seals have changed their migration and reproductive habits, polar bears have changed their distribution patterns, and walruses have altered their distributive and predation behaviors. Within both ecosystems, the overlapping species of whales and orcas demonstrate that these two species both face changes in their migratory and distributive patterns, consequently causing major changes in their predation and reproductive behaviors. Whales and orcas move between the two ecosystems throughout the year, spending the warmer part of the year further north in the Arctic marine, and they return to the more southern Northeast Pacific ecosystem when temperatures cool.

Due to this overlapping migration pattern, the two ecosystems experience different perspectives on the effect of climate change on whale behavior. As whale species spend more and more time in Arctic regions as global temperatures steadily rise, they also in turn have a greater impact on the ecosystem itself. Though the Arctic's ecosystem reacts from the added presence of whales when they are typically in the Pacific, the Northeast Pacific reacts from the loss of the whale species that are usually present. In the Arctic marine ecosystem, this migratory change has led to shifting reproductive timings as well as other Arctic species having to contend with a new predator. Within the Northeast Pacific ecosystem, the ecosystem reels from the partial loss of one of its major tertiary consumers, causing

Vol. 2023 (2) 178 – 188
ISSN 2688-3651 [online]

secondary consumer populations to grow unchecked and disrupting the entire food web of the ecosystem. Altogether, both ecosystems must face the consequences of the whale species in both choosing to spend more time in the Arctic, though the resulting changes manifest in two drastically different manners in these ecosystems.

Table 3. The effect of climate change on types of behavioral change relative to species

	Migration	Predation	Distributive	Reproductive	Social
Arctic Only	n/a	*Walruses*	*Polar bears, walruses*	n/a	*Polar bears, walruses*
NE Pacific Only	*Sea otters, dolphins*	*Sea otters, dolphins*	*Seals*	n/a	*Sea otters, dolphins*
Both	*Whales, orcas, seals*	*Whales, orcas*	*Whales, orcas*	*Whales, seals*	*Whales, orcas, seals*

The two ecosystems also show differences in species density in response to climate change. The Northeast Pacific is witnessing an increase in the densities of its species, while the Arctic sees a decrease in its densities (Whitt et al., 2015; Laidre et al., 2008; Moore and Huntington, 2008). The Arctic's melting ice causes many species to migrate further north in search of more ice. The diminishing habitat of species living on ice causes the population of these species to decline. Furthermore, seals in the Arctic, which require ice to give birth, also decrease in population as their necessary surroundings for parturition decrease. By contrast, due to increasing temperatures in the Northeast Pacific, sea otter and dolphin population densities increase in this ecosystem as more sea otters and dolphins move into the bay for warmer waters and more prey.

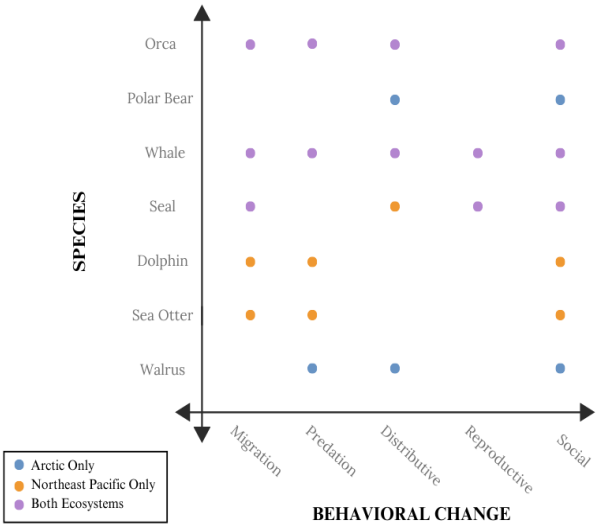


Figure 1. The effect of climate change on animal behavioral change across species. Here I show the five categories of behavior I identified from the literature and how the species analyzed may show behavioral changes for each category in each ecosystem. Some taxa have overlapping behavioral changes in both the Arctic and Northeast Pacific ecosystem. However, for many species, the effects of climate change impact them differently in the two regions, likely because the Northeast Pacific has a greater diversity of behavioral changes while the Arctic is experiencing more severe changes.

Although the species in the Northeast Pacific ecosystem have more variety in their behavioral changes, that is because they have the freedom to seek alternative means of survival. The behavioral changes as a result of climate change in the Northeast Pacific demonstrate greater complexity, as a result, it is more difficult to predict its subsequent outcomes on the overall ecosystem. The behavioral changes in the Arctic marine ecosystem are significantly more extreme, albeit less complex. The Arctic species' changing behaviors highlight the sensitivity and high vulnerability of the Arctic marine to the globally rising temperatures. More importantly, the discrepancy in the degree of complexity of behavioral changes reflects the striking variation in how different ecosystems respond to climate change. Based on the preceding analysis, it does not seem that the changes in the Northeast Pacific ecosystem will necessarily follow those changes observed in the Arctic marine ecosystem over time simply because one is more extreme than the other. The behavioral changes exhibit a variety of patterns, including increases or declines in population densities, and are driven by different underlying reasons, such as melting ice caps, water temperature, or availability of food sources, highlighting the importance of comparative analyses.

Vol. 2023 (2) 178 – 188
ISSN 2688-3651 [online]

4. Discussion

In the years to come, climate change will likely continue to affect animal behavior. Within the Northeast Pacific region, many species may move in greater numbers into regions along the coastline, potentially overpopulating these areas and eventually causing the population of these species to decline. This may lead to negative consequences for the Pacific fishing industry. Other species might seek habitats further north. Overall, climate change will greatly affect the distribution patterns of the species in the Northeast Pacific ecosystem. The Arctic marine ecosystem is also likely to show an overall decline in species density. Due to habitat destruction, animals in the Arctic marine area will continue to shift their migration patterns to adapt to their changing climate, potentially leading to altered predation habits and reproduction rates. Different species distributions, migration patterns, and predation behaviors will lead to new species interactions. It is important for researchers to be cognizant of the differences in how climate change transforms specific animal behaviors in each unique ecosystem to identify effective tailored short-term and long-term solutions to these negative impacts. In summary, this review has demonstrated that climate change has drastically altered the lifestyles of many species, and ultimately the lifestyles of humans, and it will continue to do so unless prevented.

The scope of the review was limited as only five marine species were chosen for each ecosystem. The species were selected to highlight the main ecological crises caused by climate change in these two marine ecosystems; however, many more comparisons can be made at the ecosystem level, particularly in regards to species that are not marine mammals. Additionally, it is important to be mindful of the discrepancy between the size and species count of the two ecosystems while drawing a direct comparison between them. This review was also limited by a lack of empirical data and up-to-date research on animal behavioral changes in response to climate change. Future research that tracks observational data on key behavioral change variables will be essential in performing additional comparative analyses and making reliable predictions for behavioral changes in each ecosystem.

Future review studies on a similar topic may select a larger variety of species beyond marine mammals and predators and potentially include microorganisms or plants in marine ecosystems. A larger scope would provide additional evidence and insights into the effects of climate change. Future research may examine food chains as a whole rather than individual species or select the most at-risk species in ecosystems. Field experiments along the North American Pacific coastline or behavioral experiments along the Pacific coast will also be promising areas of future exploration. In addition, as global temperatures continue to rise and shatter previous records in written history and considering the record-breaking temperatures of the summer of 2023, future research could study the impact of accelerated temperature increases on various animals globally and the relationship between the vulnerability of an ecosystem and its animals' behaviors.

5. Conclusion

This review examines the impact of climate change on ecosystems worldwide while focusing on its impact on animal behavior, particularly in marine mammals. Furthermore, this review illustrates overarching behavioral signatures of climate change. By comparing two unique marine ecosystems, this review highlights similar behavioral adjustments that demonstrate the consequences of climate change. The effects of climate change are evident in the behavioral adaptations that have appeared in species throughout the Arctic marine and Northeast Pacific ecosystems. Research has shown that species in both ecosystems have altered their migration, distribution, predation, and reproductive patterns to adapt to their rapidly changing environment. While there are key differences between the two ecosystems, similarities in species' behaviors show the effect climate change has on marine mammals. As the impact of climate change worsens, more and more species' populations in the world continue to decline and struggle to respond to climate change. Ideally, this initial analysis will inform future research and underline the importance of considering animal behavior in changing ecosystems.

Acknowledgment

I would like to acknowledge my mentor, Carly Scott for her guidance on this project.

Vol. 2023 (2) 178 – 188
ISSN 2688-3651 [online]

References

Atwood, T. C. et al. (2016). Rapid environmental change drives increased land use by an Arctic marine predator. *PLoS One*, 11 (6), e0155932. DOI: 10.1371/journal.pone.0155932

Bailey, H et al. (2009). Behavioural estimation of blue whale movements in the Northeast Pacific from state-space model analysis of satellite tracks. *Endangered Species Research*, *10*, 93-106. DOI: 10.3354/esr00239

Beever, E. A. et al. (2017). Behavioral flexibility as a mechanism for coping with climate change. *Frontiers in Ecology and the Environment*, *15*(6), 299-308. DOI: 10.1002/fee.1502

Bluhm, B. A., & Gradinger, R. (2008). Regional variability in food availability for Arctic marine mammals. *Ecological Applications*, *18*(sp2), S77-S96. DOI: 10.1890/06-0562.1

Breaker, L. C. (2006). "Nonlinear aspects of sea surface temperature in Monterey Bay. Prog. *Oceanography*, 69 (1), 61–89. DOI: 10.1016/j.pocean.2006.02.015

Buckman, A. H. et al. (2011). PCB-associated changes in mRNA expression in killer whales (Orcinus orca) from the NE Pacific Ocean. *Environmental Science & Technology*, *45*(23), 10194-10202. DOI: 10.1021/es201541j

Burrows, J. A. et al. (2012). Marine mammal response to interannual variability in Monterey Bay, California. *Marine Ecology Progress Series*, 461, 257–271. DOI: 10.3354/meps09712

Chavez, F. P. et al. (2017). Climate variability and change: response of a coastal ocean ecosystem. *Oceanography*, 30 (4), 128–145. DOI: 10.5670/oceanog.2017.429

Davidson, S. C., & Ruhs, E. C. (2021). Understanding the dynamics of Arctic animal migrations in a changing world. *Animal Migration*, 8 (1), 56–64. DOI: 10.1515/ami-2020-0114

Estes, J. A. et al. (2003). Individual variation in prey selection by sea otters: patterns, causes and implications. *Journal of Animal Ecology*, 72 (1), 144–155. DOI: 10.1046/j.1365-2656.2003.00690.x

Gifford-Gonzalez, D. (2011). Holocene Monterey Bay fur seals. Human impacts on seals, sea lions, and sea otters: Integrating archaeology and ecology in the Northeast Pacific: 221–242.

Higdon, J. W., Hauser, D. D., & Ferguson, S. H. (2012). Killer whales (Orcinus orca) in the Canadian Arctic: distribution, prey items, group sizes, and seasonality. *Marine Mammal Science*, *28*(2), E93-E109. DOI: 10.1111/j.1748-7692.2011.00489.x

Higdon, J. W., & Ferguson, S. H. (2009). Loss of Arctic sea ice causing punctuated change in sightings of killer whales (Orcinus orca) over the past century. *Ecological Applications*, *19*(5), 1365-1375. DOI: 10.1890/07-1941.1

Hardy, J. T. (2003). *Climate Change: Causes, Effects, and Solutions*. John Wiley & Sons.

Harmon, J. P., & Barton, B. T. (2013). On their best behavior: how animal behavior can help determine the combined effects of species interactions and climate change. *Annals of the New York Academy of Sciences*, 1297 (1), 139–147, DOI: 10.1111/nyas.12192

Hazen, E. L. et al. (2013). Predicted habitat shifts of Pacific top predators in a changing climate. *Nature Climate Change*, 3 (3), 234–238. DOI: 10.1038/nclimate1686

Kiliaan, H. P. L., & Stirling, I. (1978). Observations on overwintering walruses in the eastern Canadian High Arctic. *Journal of Mammalogy*, *59*(1), 197-200. DOI: 10.2307/1379895

Laidre, K. L. et al. (2008). Quantifying the sensitivity of Arctic marine mammals to climate-induced habitat change. *Ecological Applications*, 18 (sp2), S97–S125. DOI: 10.1890/06-0546.1

Vol. 2023 (2) 178 – 188
ISSN 2688-3651 [online]

Moore, S. E., & Huntington, H. P. (2008). Arctic marine mammals and climate change: impacts and resilience. *Ecological Applications*, 18 (2 Suppl.), S157–S165. DOI: 10.1890/06-0571.1

Nicholson, T. E. (2000). *Social structure and underwater behavior of harbor seals in southern Monterey Bay, California*. Diss; San Francisco State University.

Raverty, S. et al. (2020). Pathology findings and correlation with body condition index in stranded killer whales (Orcinus orca) in the northeastern Pacific and Hawaii from 2004 to 2013. *PloS one*, *15*(12), e0242505. DOI: journal.pone.0242505

Reiter, S. M. et al. (2015). Climate adaptation planning in the Monterey Bay region: an iterative spatial framework for engagement at the local level. *Natural Resources*, 06 (5), 375–379. DOI: 10.4236/nr.2015.65035

Townhill, B. L. et al. (2022). Pollution in the Arctic Ocean: An overview of multiple pressures and implications for ecosystem services. *Ambio*, 51 (2), 471–483. DOI: 10.1007/s13280-021-01657-0

Walsh, J. E. (2008). Climate of the Arctic marine environment. *Ecological Applications*, 18 (sp2), S3–S22. DOI: 10.1890/06-0503.1

Wassmann, P. et al. (2011). Footprints of climate change in the Arctic marine ecosystem. *Global Change Biology*, 17 (2), 1235–1249. DOI: 10.1111/j.1365-2486.2010.02311.x

Whitt, A. D. et al. (2015). Abundance and distribution of Marine mammals in nearshore waters off New Jersey, USA. *J. Cetacean Res. Mana*g., 15, 45–59.

Journal of Research High School

Vol. 2023 (2) 189 – 193
ISSN 2688-3651 [online]

Chemical Composition Analysis on Amethyst

Yinghan Zhang[1] *

[1]Beijing No.2 Middle School, Tongzhou Campus, Beijing China
*Corresponding Author: yinghanzhang2005@163.com

Advisor: Yuanyuan Liu, yuanyuanliu198705@163.com

Received March 16, 2023; Revised July 29, 2023; Accepted, September 13, 2023

Abstract

Amethyst is one of the most popular ornaments in Chinese culture. Here, aiming at clarifying the component of amethyst, chemical analysis by energy dispersive X-ray spectrum and powder X-ray diffraction, combined with Raman and infrared absorption spectrum, was carried out. Energy dispersive X-ray spectrum suggested that its dominant elements is silicon and oxygen in the atomic concentration ratio of 18.98:43.01, with tiny iron, cobalt or manganese in the mass percentage less than 1%. Further, X-ray diffraction demonstrated that its main phase is crystalline quartz, with the transition metals embedded in the crystal lattice, and the conclusion of which was also confirmed by lattice dynamics analysis based on Raman and infrared absorption spectra. Through our study, the component of amethyst was clarified.

Keywords: Amethyst; Chemical Composition; Crystal Structure; Quartz

1. Introduction

In Chinese traditional culture, purple symbolizes nobility, elegance, happiness and peace, and amethyst is one of the most popular ornaments in the living and office room in China (Ye, 1994). In our living room there is an amethyst hole. The question about the component of amethyst has been interesting to me, since my dad bought the amethyst hole and placed it there seven years ago.

Figure 1 The amethyst ornament in our living room. The inset displays the part used for chemical component analysis.

This query has been asked to many people around, and the gotten answers covered from crystal glass, gemstones, diamond to even jade, etc. To clarify these controversies and get an academic response, a comprehensive literature research was implemented. Accordingly, in these previous studies, by X-ray diffraction, the main phase of amethyst was indexed to be quartz (SiO_2) with a small amount of defect (Wu, 2016 and Suastika 2017). However, the deviation of crystal structure from pure quartz induced by the defects, has never been concerned, and the exact lattice constant of amethyst is still unclear. Besides, pure quartz is colorless, and it is anticipated that the purple color is attributed to the optical absorption from the defects. By utilizing absorption spectra, the defects, that give rise to the purple color, was indirectly assigned to the tiny iron atoms and some other transition element (Liu, 2022). It should be noted that absorption spectrum is an indirect element analysis approach. The direct evidence that strongly support the origin of the purple in amethyst is still inadequate. The objective of this study is to disclose the exact chemical composition in amethyst. According to literature review, it is anticipated the main component is quartz (SiO_2) in a high probability, and here the exact defect

Vol. 2023 (2) 189 – 193
ISSN 2688-3651 [online]

element, that result in the purple color, will be explicitly determined. Moreover, the values of lattice constant modified by the defect will be refined as well. Here, for this purpose, a small piece was cut from the ornament, and element analysis and atomic arrangement determination is carried out on it.

2. Materials and Methods

2.1 Sample Preparation

The amethyst cut from the ornament was firstly soaked by dilute nitric acid for 24 hours to dissolved the impurity in the surface, and then washed by deionized water. After being kept in drying oven for 6 hours to remove the surface water, by an agate mortar the amethyst was well grinded in to powder. The powder sample was used for all the measurement.

2.2 Chemical Element Analysis

Chemical element analysis on amethyst was performed by the method of energy dispersive X-ray spectroscopy (EDX), implemented by a Hitachi S4800 scanning electron microscopy. Under the electron beam irradiation, X-ray is emitted from the constituent atoms in the sample. The elements and corresponding contents can be determined from the characteristic energy and intensity of the emitted X-ray (Zieba, 2000).

2.3 Powder X-Ray Diffraction (PXRD)

Powder X-ray diffraction (PXRD) data is an effective tool to determine microscopic structure in atomic scale. It was collected by a PXRD apparatus of Rigaku smartlab SE, with a Cu Kα radiation (λ=1.5406Å). The 2θ range was set to 5° ~ 90° with the scanning angular step of 0.02°, and the scanning rate was set to a slow value of 0.2s/step to guarantee a high intensity, so as to make the tiny components in the sample show up. The angles of diffraction peaks are different from one component to another in PXRD spectrum, and according to the angles where diffraction peaks occur, the components in the sample can be identified (Gualtieri, 1997). JADE 6.0 was employed to index the PXRD spectrum. The PXRD spectrum was refined by FULLPROF Suit (Rodriguez-Carvajal, 2008).

2.4 Raman and Infrared Absorption Spectrum

The atomic type and arrangement can also be reflected in the atomic vibrational frequency, which can be studied by Raman and infrared absorption spectrum. Raman spectrum was collected by a LabRAM Aramis confocal Raman spectroscopy system (Renishaw, Invia Raman microscope). A 532nm laser serves as the excitation wavelength, and the spectrum was collected in the range of 100 ~ 4000cm^{-1}, with the interval of 1cm^{-1}. The spectrum was integrated for five times to improve the signal-to-noise ratio. Infrared spectrum was recorded by a Thermo Nicolet 6700 infrared spectrometer in the range of 400~4000cm^{-1} with a spectral resolution of 0.5cm^{-1}. In the infrared absorption measurement, a flake tableted by mixture of powder amethyst and KBr in the mass ratio 1:99 was used. The peaks in Raman and infrared absorption spectrum characterize the atomic vibrational frequency in the sample, by which the types of the constituent atomic unit can be identified.

3. Results and Discussion

In order to determine the chemical composition, it is prerequisite to identify the constituent elements in amethyst. Energy dispersive X-ray spectrum performed by scanning electron microscopy, which distinguish the element via the emitted X-ray under electron radiation, tells that silicon, oxygen, iron, cobalt and manganese elements are included in the sample. As displayed in Table 1, the quantitative analysis unveils that silicon and oxygen dominantly occur in the sample, with the mass percentage of ~99%, and those of transition metals, *i.e.*, iron, cobalt and manganese, are less

Vol. 2023 (2) 189 – 193
ISSN 2688-3651 [online]

than 1%. Remarkably, the atomic concentration ratio between silicon and oxygen are 18.98:43.01, which is close to 1:2 by considering the measured error. This suggests that the main phase of amethyst is quartz (SiO_2), and iron, cobalt and manganese may exist in the form of impurity. It should be emphasized that, conventionally, the pure quartz is colorless and absolutely transparent. It is speculated that the purple of amethyst may arise from the light absorption of the constituent transition metal elements (Marzouk, 2006), and the uneven distribution of purple is attributed to their distribution inhomogeneity in the main phase of quartz (Figure 1).

Table 1 Element analysis result of energy dispersive X-ray spectrum

Element	Concentration(a.u.)	weight%
O	43.01(32)	55.93(42)
Si	18.98(26)	43.19(59)
Fe	0.15(8)	0.68(36)
Co	0.04(9)	0.19(43)
Mn	0.00(16)	0.00(71)
Total		100.00

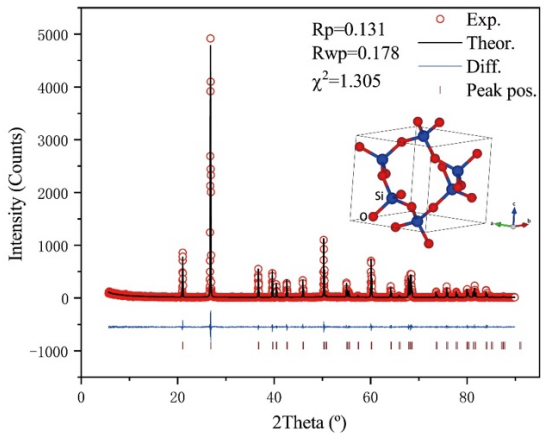


Figure 2. Powder X-ray diffraction spectrum of amethyst. Rp, Rwp and χ^2 are the profile R-factor, weighted profile R-factor, and goodness-of-fit for the refinement, respectively. Red circles and black line are the experimental and theoretical XRD spectra respectively, blue line is the difference between the experimental and theoretical spectra, and red bars indicate the theoretical peak positions.

Powder X-ray diffraction (PXRD) can distinguish whether a solid is in crystalline- or non-crystalline states, via sharp diffraction peaks or broad scattering bump occurring in the spectrum. If it is the former, the chemical composition can be determined by the angles where the diffraction peaks emerge. Accordingly, sharp peaks are observed in the spectrum, and a smooth scattering background without bump is manifested, as plotted in Figure 2. This demonstrates that the compositions are all crystalline and no non-crystalline is included. All the diffraction peaks can be indexed to the quartz in the space group $P3_221$ with two independent cell parameters a=4.913Å and c=5.405Å (PDF card No. 99-0088, Le Page, 1976). Moreover, except the diffraction peaks belonging to quartz, no peaks, that are indexed to the compounds containing iron, cobalt or manganese, are detected. This demonstrates that these transition elements are embedded in the crystalline lattice of quartz, but do not exist in the form of isolated compounds. The cell parameters are refined to be a=4.92068(3) Å and c=5.41291 (6) Å on the PXRD spectrum, which agree with the values in standard PDF card. The slight deviation between the theoretical and refined values may be due to the doping of transition metals with the different atomic radii from silicon and oxygen atoms.

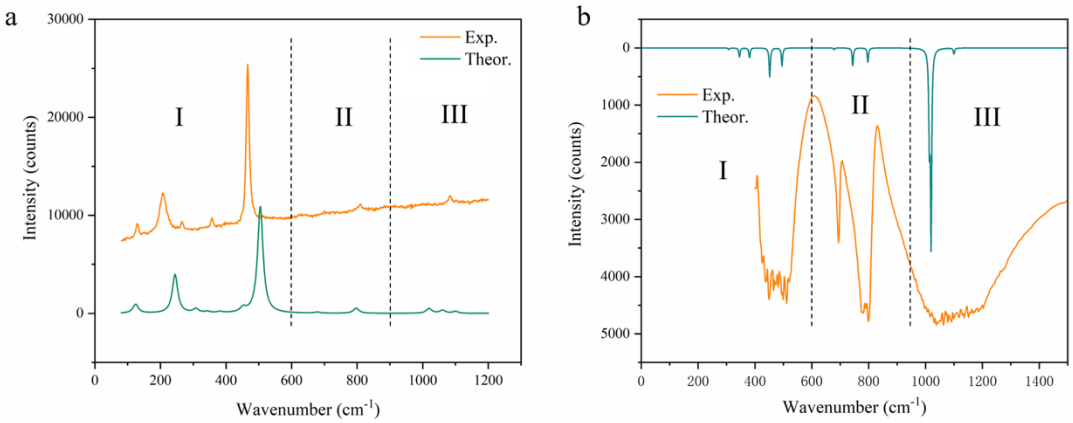


Figure 3. Raman (a) and infrared (b) absorption spectrum of amethyst

Vol. 2023 (2) 189 – 193
ISSN 2688-3651 [online]

To further confirm the composition determined by energy dispersive X-ray spectroscopy and powder X-ray diffraction, the atomic vibration spectrum analysis based on Raman and infrared absorption spectrum are implemented. Raman and infrared absorption spectrum are two sets of complementary approach to investigate the intrinsic atomic vibrational frequency and the corresponding intensity, by which all the atomic vibrational frequency can be detected (Dove, 1993). As displayed in Figure 3, both the Raman and infrared absorption spectrum are consistent with those measured by Briggs and Soda respectively (Briggs, 1977; Soda, 1961). Roughly, these peaks can be divided to three types: (I) those below ~ 600cm^{-1} are attributed to the atomic vibration of the whole skeleton as a quasi-rigid unit; (II) The vibration of the silica structure, i.e., the twisting and rotating of the SiO_4 units predominantly account for the peaks between ~600 cm^{-1} and 900cm^{-1}; (III) Remarkably, a peak is detected at ~1100cm^{-1}. The previous study performed by Tekippe et al has explicitly disclose that this Raman peak directly characterize the stretching vibration of Si-O bonds in quartz (Tekippe, 1973). All these observation in Raman and infrared spectrum confirm that the main phase of amethyst is quartz.

4. Conclusion

In summary, to disclose the composition of amethyst, the energy dispersive X-Ray spectrum, powder X-ray diffraction, Raman and infrared absorption spectrum was performed. Energy dispersive X-ray spectrum suggests that its main component is quartz with tiny (less than 1%) transition metal, including iron, cobalt and manganese. By powder X-ray diffraction, it is demonstrated that amethyst has a crystalline state, and can be indexed to quartz. The slight deviation between the refined and theoretical values of the lattice constants verifies the dopant of iron, cobalt and manganese with the different radii from silicon and oxygen atoms. Further, the conclusion that main component is quartz, is confirmed by atomic vibration analysis based on Raman and infrared absorption spectrum analysis. Our study answers the question what is the exact composition amethyst, and clarifies that amethyst is quartz with tiny impurity of iron, cobalt and manganese inserted in the crystal lattice.

Acknowledgment

The author acknowledges the guidance and supervision of Yuanyuan Liu, and the financial support from the scientific and technological innovation project for high school students

References

Briggs, R. J., & Ramdas, A. K. (1977). Piezospectroscopy of Raman-spectrum of alpha-quartz. *Physical Review B*, *16* (8), 3815-3826.

Dove, M. T. (1993). *Introduction to Lattice Dynamics*. Cambridge University Press: Cambridge.

Gualtieri, A. F., Immovilli, S., & Prudenziati, M.(1997). Powder X-ray diffraction data for the new polymorphic compound omega-Bi_2O_3. *Powder Diffraction*, *12* (2), 90-92.

Le Page, Y., & Donnay, G. (1976). Refinement of the crystal structure of low-quartz. *Acta Crystallographica Section B*, *32* (8), 2456-2459.

Liu K., & Guo Y. (2022). Comparative study of mineralogical characteristics of natural and synthetic amethyst and smoky quartz, Crystal, 12(2022), 1735.

Marzouk, S. Y., et al. (2006). Optical absorption of gamma-irradiated lithium-borate glasses doped with different transition metal oxides. *Physica B: Condensed Matter*, *382* (1), 340-351.

Rodriguez-Carvajal, J. (2008). The determination of magnetic structures by simulated annealing using the FullProf Suite. *Acta Crystallographica a-Foundation and Advances*, *64*, C107-C107.

Vol. 2023 (2) 189 – 193
ISSN 2688-3651 [online]

Soda, R. (1961). Infrared absorption spectra of quartz and some other silica modification. *Bulletin of the Chemical Society of Japan*, *34* (10), 1491-1494.

Suastika G. K., et al. (2017). Characterization of central kalimantan's amethysts by using X-ray diffraction, Journal of Physics: Conference Series, *846*, 012024.

Tekippe, V. J., Ramdas, A. K., & Rodriguez S. (1973). Piezospectroscopic study of the Raman spectrum of alpha-quartz. Physical Review B, 8, 706.

Wu, W., & Yue S. (2016). Influence of heat treat on colour of amethyst and its spectral characterisitic. *Journal of Gems and Gemmology***,** *18* (05), 47-55.

Ye, Y. (1994). How to appraise amethyst. *China Gems & Jades***,** 1(03), 20-21.

Zieba, P. (2000). Recent progress in the energy-dispersive X-ray spectroscopy microanalysis of the discontinuous precipitation and discontinuous dissolution reactions. *Materials Chemistry and Physics***,** *62* (3), 183-213.

Journal of Research High School

Vol. 2023 (2) 194 – 202
ISSN 2688-3651 [online]

Presence of Phosphorus and its Effects on the Habitability of Exoplanets

Shreyan Deo[1 *]

[1]Delhi Public School, Delhi, India
*Corresponding Author: shreyandeo011@gmail.com

Advisor: Mritunjay Sharma, mritunjay@athenaeducation.co.in

Received May 4, 2023; Revised August 21, 2023; Accepted, September 22, 2023

Abstract

Could phosphorus be the key to unlocking life beyond our planet? In this review paper, the various dimensions of phosphorus's significance in Exoplanetary habitability are comprehensively explored. Starting with fundamental concepts, including the definition of Exoplanets and the array of factors shaping their habitability, the paper proceeds to elucidate phosphorus's pivotal role in supporting potential life. Contemporary methods for estimating phosphorus content through parent star observations are examined, alongside an intricate analysis of the phosphorus cycle's interrelation with carbon and oxygen cycles. The discourse extends to the feasibility of sustaining a phosphorus cycle on Waterworlds devoid of extensive landmasses. The critical phosphorus compound, phosphine gas, is highlighted as a key biosignature gas, with detection techniques and limitations expounded upon. The review also briefly engages with the phosphine detection controversy in Venus's clouds and its potential implications. In conclusion, current limitations and future prospects for advancing phosphorus-related studies in extraterrestrial contexts are discussed, underscoring the pursuit's significance. Overall, this comprehensive review seeks to establish phosphorus as a potential indicator of Exoplanetary habitability.

Keywords: Phosphorus, Phosphine, Biosignature, Phosphorus Cycle, Habitability, Astrobiology, Exoplanet, Venus, Waterworld

1. Introduction

This research paper revolves around a holistic understanding of Phosphorus and its implications for Exoplanetary environments, particularly concerning habitability and the emergence of life. This paper delves into several key themes that encompass the challenges associated with detecting Phosphorus on Exoplanets directly and its significance when studying the composition of their parent stars. Our investigation also delves into the intricate interplay between the Phosphorus cycle and the oxygen and carbon cycles, shedding light on the potential ramifications for planetary habitability.

The Waterworld debate constitutes a central part of our study, as it evaluates the viability of a Phosphorus cycle on Exoplanets dominated by water, challenging prior assumptions about such cycles' feasibility. Furthermore, our research underscores the importance of Phosphine gas as a distinctive biosignature and its potential role in identifying signs of life on Exoplanets.

In this context, our research contributes by systematically reviewing and consolidating the existing body of knowledge on Phosphorus within Exoplanetary systems. This paper aims to bridge the gap between diverse studies and ongoing debates, providing a comprehensive overview of the challenges, possibilities, and implications related to Phosphorus detection, distribution, and its potential role as a biosignature. By organizing and analyzing multiple facets of Phosphorus in Exoplanetary contexts, this paper offers a cohesive framework that benefits researchers and astronomers seeking to unravel the intricacies of Exoplanetary habitability and the search for potential extraterrestrial life forms.

Vol. 2023 (2) 194 – 202
ISSN 2688-3651 [online]

Planets that lie beyond our solar system are referred to as Exoplanets. Most orbit another star, but some are also free-floating. Since the 1990s, scientists have confirmed over 5000 heavenly bodies as Exoplanets, and thousands of others are under review. (*NASA Exoplanet Exploration*) One of the essential features that this paper explores is the habitability of Exoplanets.

Habitability is the capability of a heavenly body to sustain at least one known organism. A habitat must ensure metabolic activity for an organism and allow its reproduction, but it is not necessary (Cockell et al., 2016).

For a long time, humanity has been conflicted over whether it is alone in the universe or not. By looking at the habitability of Exoplanets one can offer one possible solution to this conflict. Another reason scientists are interested in looking at an Exoplanet's habitability is to make it a potential "backup" option for humans. It is well-known that the resources on Earth are dwindling day by day and conversely, pollution is on the rise. This may result in Earth becoming incapable of supporting life soon. Humans must shift to another planet to ensure survival in such an event. Thus, by looking at habitability, researchers can pinpoint which world may have Earth-like conditions and be able to sustain human population. Lastly, systematically looking at habitability factors helps us eliminate several planets. Our resources are limited and it is only possible to visit a limited no. of Exoplanets. Therefore, the possibilities can be efficiently narrowed down using certain constraints to find the "perfect" planet to sustain life.

Now, let us discuss some of the requirements for habitability. For the purpose of this paper, it is not possible to go in detail about each condition, nor individually look at the different categories of habitability (i.e., instantaneous and continuous habitability), but rather this paper provides a brief review of the main requirements:

a. Solvent- The function of a solvent is to dissolve various substances and allow particular essential reactions to occur. Presently, water is the only liquid solvent known that is capable of supporting life. Water availability on a planet is affected by several factors, such as temperature, distance from its star, etc. It is worth noting that a planet with a significant amount of liquid water is known as a liquid-water world. Based on the location of water, they are further categorized as surface liquid water world and interior fluid water world (Cockell et al., 2016).
b. Physicochemical conditions- Temperature, radiation, pressure, pH, salinity, aridity and toxic metals are some of the various physicochemical conditions which must be within a certain limit for specific organisms to survive. However, it is seen that some microorganisms can survive (i.e., perform metabolic activities and may even be capable of reproduction under such conditions) in even harsh physicochemical conditions such as high or low temperatures, pressure, pH etc (Cockell et al., 2016).
c. Available Energy- Energy needs vary significantly from organism to organism, but what can be deduced is that the environment must provide them with sufficient energy in some form (for example, light) to carry out their life processes (Cockell et al., 2016).
d. Major Elements- There are six essential elements for life: Carbon, Hydrogen, Nitrogen, Oxygen, Phosphorus and Sulphur (together abbreviated as CHNOPS). They are present in various forms (for example, CO_2, SO_2, O_2, etc.), and each performs a unique function to ensure life. These elements (specifically Phosphorus) and their importance are studied later in the paper in depth (Cockell et al., 2016).
e. Specific Elements- Certain organisms also require certain specific elements to survive. These include iron, manganese, potassium, etc. Life will inevitably change the availability of various elements in a habitat over time (Cockell et al., 2016).
f. Atmosphere and other Planetary Factors- These include mass, density, atmosphere, plate tectonics, magnetic field, etc., of the planet. By looking at the planetary factors, the planet can be generally categorized as whether it can sustain life or not, but on its own, it is not sufficient to claim that life can be present on a planet or not (Cockell et al., 2016).
g. Astronomical Factors- These include planetary rotation, orbital characteristics, the type of star the planet is orbiting, impact events, etc. These factors predominantly influence the type of elements present on the planet, its temperature, the type of the planet itself, and various other vital factors (Cockell et al., 2016).

Of all the significant elements required for life (CHNOPS), Phosphorus is the most unique. Termed the "Staff of Life", (Karl, 2000) Phosphorus is present in the "blueprint" of all forms of life- the DNA- and forms the base of ADP and ATP, necessary for providing energy to all organisms, which is used during their metabolic processes. It is an

Vol. 2023 (2) 194 – 202
ISSN 2688-3651 [online]

indispensable element which played a significant role in the origin of life itself. Furthermore, Phosphorus is essential for terrestrial biological productivity and is critical to carbon balance in terrestrial ecosystems.

It is present in the form of various compounds in the environment, which may be in the soil, dissolved in water or even as a gas in the atmosphere. Despite its ability to aggregate into various compounds (PO_4^{3-}, $Ca_5(PO_4)_3OH$, PH_3, etc.), Phosphorus is the limiting nutrient out of all the significant nutrients. This can be explained by looking at the Phosphorus cycle. Simply put, fresh Phosphorus available for life is released majorly only during soil weathering. Transporting this weather Phosphorus by the rivers and other water bodies is the only appreciable source of Phosphorus in the oceans (Filippelli, 2002).

Ultimately, the presence of Phosphorus can limit the primary production by marine life, thus making it the "limiting nutrient." Examining whether the Phosphorus cycle exists on every planet may be impossible. However, the conclusion that one can draw from here is that depending upon the Phosphorus present on the top surface of a planet (soil, weathering rocks, etc.), it can be classified as potentially habitable, while the others can be rejected.

However, recent studies show that ocean weathering- instead of surface weathering- may also provide sufficient concentration of Phosphorus in the oceans in the Waterworlds under certain conditions. One may wonder what a Waterworld is. Simply put, Waterworld is a planet with oceans- either surface or subsurface- covering a significant portion of its lithosphere. Earth is also a Waterworld. It is not possible detect liquid water on an Exoplanet at present, so instead water vapor is used as an indicator. Before, scientists believed Waterworlds whose entire surface is covered by oceans cannot have Phosphorus and thus are not fit for life. However, new studies suggest otherwise.

Another importance of Phosphorus is that one of its compounds- Phosphine (PH_3)- is a biosignature gas in an anoxic environment. No abiotic false positives on terrestrial planets have been found for Phosphine, making it an excellent biosignature gas. With the help of the latest technology of the James Webb Telescope, detecting Phosphine gas will become easier (Sousa-Silva et al., 2020).

Thus, Phosphorus is a unique element and a great indicator of whether life could be present on an Exoplanet.

2. Discussion

This research paper aims to systematically review different aspects of Phosphorus relating to its presence, detection and uses on an Exoplanet. The following themes are discussed ahead in order:

a. Exploring Exoplanets through Phosphorus Clues: The first section discusses how and why Phosphorus is presently detected on Exoplanets and the implications of it. It includes talking about detecting Phosphorus on an Exoplanet by looking at the Phosphorus content of its parent star, and what insights are gained about Exoplanets from it.
b. The Phosphorus Cycle and Its Relation with Oxygen and Carbon Cycles: This section mainly elaborates on the Phosphorus cycle, giving a general overview and the dependence of the oxygen and carbon cycles.
c. The Waterworld Debate: In the past years, life on Waterworlds was rejected because a Phosphorus cycle was not possible on them. However, scientists have found this false in recent years and suggest otherwise. This section addresses the Water world debate regarding the Phosphorus cycle.
d. Phosphine Gas- A Unique Biosignature: Phosphine is a unique biosignature gas. This section discusses its importance, its mechanism of formation and destruction, and its uses and limitations as a biosignature gas.
e. Debate Regarding Phosphine on the Cloud Decks of Venus: Another famous debate involving Phosphorus- specifically the Phosphine gas- was its detection in the Venusian Atmosphere. This last section reviews and summarizes the debate on the presence of Phosphine on the clouds of Venus.

2.1 Exploring Exoplanets through Phosphorus Clues

Presently, it is beyond reach to look directly at the Phosphorus present on Exoplanets. That is, more specialized tools are required to determine whether Phosphorus is present on a planet, how it is present on it and in what forms. Scientists have devised an intelligent solution to this problem- looking at the star's composition that the Exoplanet is revolving. With the current technology, it not only impossible to determine the exact the exact amount of Phosphorus

Vol. 2023 (2) 194 – 202
ISSN 2688-3651 [online]

present on an Exoplanet, but also challenging even to detect the ratios of the major elements on the planet. Thus, looking at the Exoplanet's host star and determining the ratios of the major limiting nutrients such as Nitrogen, Phosphorus, Silicon, and Carbon can provide useful insights (Hinkel et al., 2020).

The reason is that the presence of Phosphorus in a star can provide information about the chemical composition of the Exoplanet. For example, suppose a star has a high abundance of Phosphorus. In that case, it is more likely that the Exoplanet will also have a high abundance of Phosphorus, which can affect the planet's physicochemical conditions and potential habitability. (Hinkel et al., 2020).

Recent research utilizes the data provided by the Hypatia Catalog to plot N/Si to C/Si, N/Si to P/Si, and P/C to Si/C graphs (here, N= Nitrogen, Si= Silicon, C= Carbon, P= Phosphorus) of various stars of the FGKM category (temperature range of stars) with Phosphorus abundance and compare the aspects of Sun, Earth's crust, bulk silicate Mars, bulk silicate Earth, and marine plankton; however, the research does not compare the data to the Earth's Core. The Sun's overall Carbon to Nitrogen ratio appears comparable to the Hypatia average, but its Carbon to Phosphorus and Nitrogen to Phosphorus ratios are significantly lower than the Hypatia average, indicating that our Sun is comparatively Phosphorus-rich. Compared to the Sun and the Hypatia stars, the Earth and Mars ratios are Nitrogen-poor and Phosphorus-rich (Hinkel et al., 2020).

It is challenging to discern distinct patterns in the stellar data, let alone the function of Phosphorus in developing an Exoplanet, due to the paucity of data on Phosphorus, which is available for only 1% of all stars and 1% of planetary host stars. The Phosphorus content of our Sun is comparatively large, whereas the Phosphorus needed by Earth life is tiny but limited (Hinkel et al., 2020).

The intense partitioning of Phosphorus into the core on rocky planets that develop around host stars with significantly less Phosphorus may rule out the possibility of surface Phosphorus and, subsequently, of life on that planet's surface. The study also concludes that if a star has low Phosphorus abundance, its orbiting planets may have almost no Phosphorus, and it may be reasonable to rule that planet out entirely from the possibility of hosting any form of life (Hinkel et al., 2020).

Apart from the ratios, it is also important to know in which state the elements are present. For example, whether Phosphorus is present as PO_4^{3-} -a useful form- or Fe_3P -not so useful form. But this is almost impossible to know because of the lack of information; therefore, there is little to no possibility of predicting the geological cycles that may be taking place on the planet.

2.2 The Phosphorus Cycle and Its Relation with Oxygen and Carbon Cycles

It is established that one cannot predict Phosphorus cycle taking place on an Exoplanet. However, understanding the Phosphorus cycle can provide us with some helpful insight. Since Phosphorus can change a planet's physicochemical conditions, certain Exoplanets may have such physicochemical conditions that the Phosphorus cycle may not be possible entirely on that planet. If the planet does not have a Phosphorus cycle, life cannot exist on it. In brief, researchers use the physicochemical conditions of the planet as an indicator of the Phosphorus cycle. Thus, let us look at the impacts of the Phosphorus cycle and the cycle itself.

The Phosphorus cycle on Earth can be divided into several parts:

a. Weathering: Phosphorus is found in rocks and minerals and released into the environment through surface weathering.
b. Transport: Phosphorus is transported through the environment through erosion, runoff, and leaching.
c. Uptake by plants: Phosphorus is taken up by plants through their roots and incorporated into biomass.
d. Decomposition: When plants and animals die, their biomass decomposes, releasing Phosphorus into the environment.

When only primitive life forms were present on Earth, the Phosphorus cycle would have been much simpler and less affected by human activities. The primary sources of Phosphorus would have been volcanic activity and weathering of rocks, which would have released Phosphorus into the environment. Compared to today's cycle, the amount of Phosphorus would have been limited, and cycling would have been slower due to the low population of primitive organisms. In addition, cycling would have been mainly driven by geologic processes rather than by human

Vol. 2023 (2) 194 – 202
ISSN 2688-3651 [online]

activities, so the impact of human activities on the cycle would have been nonexistent. Over the years, human activities, particularly agriculture, have affected the Phosphorus cycle. Using fertilizers in agriculture has dramatically increased the amount of Phosphorus in the environment. Since scientists are often looking at planets with faint or no life traces at all, looking at the primitive Phosphorus life cycle is more important.

Having acquired a basic understanding of the Phosphorus cycle, the focus now shifts to the correlation of phosphorus cycle with the oxygen and carbon cycles.

The Phosphorus cycle is closely related to the oxygen and carbon cycle. Indirectly and through various feedback mechanisms, Phosphorus controls the oxygen and carbon cycles, which is why it is so important to look at it.

There are several complex feedback mechanisms of Phosphorus in anoxic conditions, all of which point towards the correlation between Phosphorus and oxygen. Analyzing the Phosphorus cycle over various geological periods on Earth, several feedback mechanisms may have operated together over a specific time frame, while others might not have been that active.

Let us look at a feedback mechanism that relates Phosphorus, Carbon burial and atmospheric Oxygen. First, if the anoxic environment contracts, less Phosphorus is attached to the sediments containing active carbon. This process, in turn, increases the amount of organic carbon buried per unit of Phosphorus. As a result, the buried carbon sediments release significant Oxygen, further increasing atmospheric Oxygen levels. In the same way, if the anoxic environment expands, the amount of carbon buried per unit of Phosphorus decreases, causing atmospheric Oxygen levels to decrease. Ultimately, the amount of Phosphorus limits the production of such Oxygen, limiting the atmospheric oxygen concentrations (Canfield et al., 2020).

Another feedback mechanism operates so that Phosphorus binding onto water-column-produced iron oxides in a ferruginous ocean represents a significant Phosphorus sink. The amount of Phosphorus in marine waters and the amount of Phosphorus available for primary production are decreased due to this sink. This happens because Phosphorus-rich particulates are removed into deep-sea sediments. This Phosphorus sink ultimately reduces the generation of Oxygen and the deposition of organic carbon (Canfield et al., 2020).

There are feedback mechanisms apart from these also. However, while all these feedback mechanisms related to Phosphorus acted to different extents at different times, all of these affected the carbon and oxygen cycles (Canfield et al., 2020; Filippelli, 2002).

2.3 The Waterworld Debate

This leads to our next topic. If the planet has no appreciable surface and is covered mainly by oceans, will the Phosphorus cycle be possible on such a planet?

For a long time, it was assumed that life is impossible in Water Worlds because it lacks a mechanism for delivering essential nutrients such as Phosphorus. Nevertheless, recent research studies present a new front, trying to show that it is possible that enough Phosphorus can be weathered from the seafloor basalts to sustain an ecosystem for evolution in a Water World.

Many scientists thought that Phosphorus might be lacking for consumption in a water world because the conditions of the deep ocean were similar to that of the Earth- they were considered oxic (rich in oxygen). If the deep ocean is oxic, the Phosphorus combines with iron oxides and settles down, causing a significant loss of Phosphorus. Thus, it was earlier considered that if the terrestrial weathering of Phosphorus is not possible, it must not be readily available for consumption on Water Worlds. However, studies proved that Phosphorus might be sufficiently available if the deep ocean is anoxic in the Water Worlds. For this, the environment of a Water World was replicated with submarine basalt seafloor. Several lab experiments were then performed, which revealed that a sufficient amount of Phosphorus was released from these and then cycled under anoxic conditions. The reaction of dissolved CO_2 being consumed during submarine basalt weathering is balanced by the release of Phosphorus during weathering and the subsequent uptake and ultimate burial of organic carbon. O_2 is formed during this process (Filippelli, 2022; Holstege, 2019).

Vol. 2023 (2) 194 – 202
ISSN 2688-3651 [online]

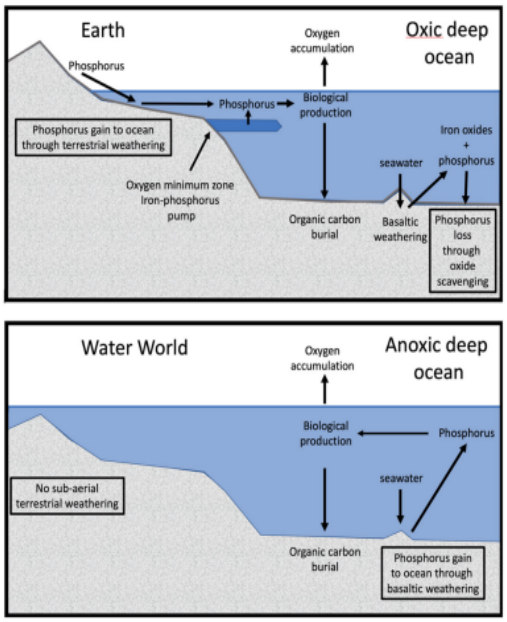


Figure 1: The following diagram briefly describes the Phosphorus cycle on Earth and a Waterworld with an Anoxic deep ocean. The major difference is that Earth has terrestrial surface weathering, whereas Waterworld doesn't (Filippelli, 2022).

There are two significant implications of this study. Firstly, although some scientists considered the cycling of Phosphorus under anoxic conditions, they did not consider it being released from submarine basalt seafloors. Secondly, an Exoplanet with notably more water than Earth may also have a carbon-silicate cycle along with a Phosphorus cycle, having a similar response to that of Earth's cycles.

2.4 Phosphine Gas: A Unique Biosignature

Phosphine is a gas that holds particular importance. It has three unique features- that is, it shows strong bands around 2.7-3.6, 4.0-4.8 and 7.8-11.5 micrometers- and at least one of the three bands is dominant and distinguishable when compared to other gases, such as CO_2, H_2O, CO, CH_4, NH_3, and H_2S (Fig. 2).

Phosphine has no substantial abiotic false positives from any source that could generate fluxes high enough for detection on terrestrial planets with surface temperatures below 800K. If Phosphorus is produced at a high-enough rate at the planet's surface, it can rapidly build up to amounts that are observable in the atmosphere of an Exoplanet. Another advantage of Phosphine is that it can be detected at no extra cost because its spectra lie in the same wavelength as that of other critical atmospheric molecules and biosignature gases (Sousa-Silva et al., 2020).

There are primarily two possible ways in which Phosphine may be produced:

a. Anaerobic bacteria can form Phosphine using Phosphorus in its surroundings through a direct process
b. Anaerobic microbes can generate Phosphorus indirectly too. These microbes make certain acid products from the anoxic fermentation of organic matter, and these acid products then interact with inorganic metal phosphides, such as those found as residual elements in waste metal, to produce Phosphine.

The researchers emphasize that the first way may be more probable, although the exact process of Phosphine production in anoxic environments remains a mystery.

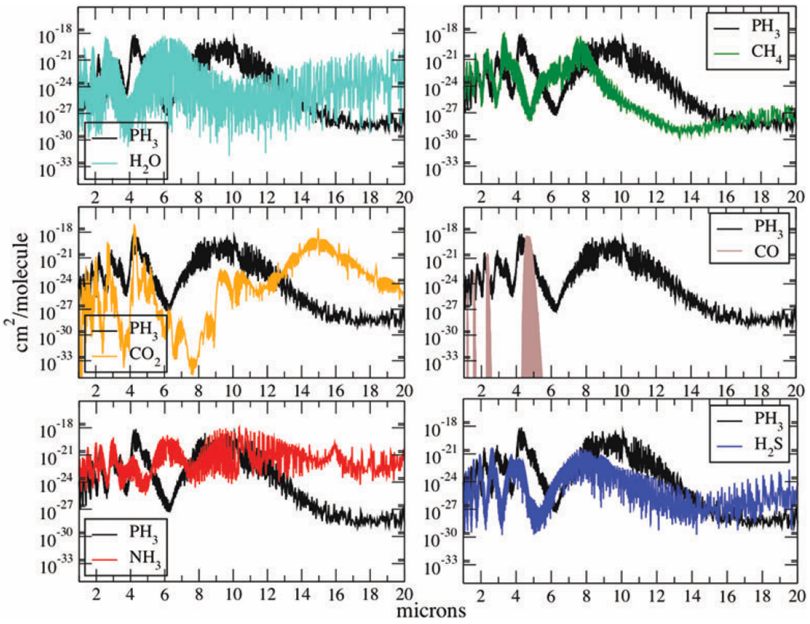


Figure 2: These figures compare the spectral cross-sections of Phosphine with other molecular gases at room temperature. Intensity on the y-axes is given in a log-scale with units of cm2 /molecule, and wavelength is represented on the x-axes in microns (Sousa-Silva et al., 2020).

Vol. 2023 (2) 194 – 202
ISSN 2688-3651 [online]

Apart from production, Phosphine is destroyed by OH, O or H molecules. Out of these, OH is by far the most common way Phosphine is destroyed, whereas O is the fastest.

After its destruction, Phosphine can also form again by recombining with H molecules, which can be a significant source of Phosphine, especially if the concentration of PH_2 is relatively high in the surroundings.

Apart from these molecules, UV radiation is also a source of destroying Phosphine. It is primarily because less UV radiation makes Phosphine gas easier to detect at night on Earth. Although Earth has an ozone layer protecting it from UV radiation, planets with little to no UV protection and are anoxic tend to have a UV as a major destroying Phosphine (Sousa-Silva et al., 2020).

On rocky planets, Phosphine production without life is thermodynamically unfavorable. Phosphine cannot be produced by any abiotic processes at the rates required for its discovery on habitable Exoplanets. Therefore, scientists come to the conclusion that unlike other compounds like ammonia and methane, the discovery of Phosphine on a temperate world is most likely to be explained by the existence of life. Production of Phosphine through exogenous delivery, lightning and volcanism is also negligible, making false positives almost impossible (Sousa-Silva et al., 2020).

However, Phosphine, too, has its disadvantages. Its vulnerability to UV photolysis presents a significant problem when detecting it. Moreover, it may take a long duration of time to detect Phosphine spectrally.

If detected on a terrestrial planet, Phosphine is an extremely promising biosignature gas since its lack of high-flux false positives would be a solid reason to hypothesize production by life.

2.5 Debate Regarding Phosphine on the Cloud Decks of Venus

In 2020, an announcement was made by a team of astronomers that sparked a lot of excitement and debate around the potential for life on Venus - the discovery of Phosphine gas in the cloud decks of Venus. Phosphine gas is a molecule associated with life forms on Earth, and its presence on Venus has raised many questions about the possibility of life existing elsewhere in our Solar System. The evidence for the presence of Phosphine gas on Venus comes from observations made by astronomers using the James Clerk Maxwell Telescope (JCMT) and Atacama Large Millimeter/submillimeter Array (ALMA) Telescope. This is the first time that Phosphine gas has been detected in the atmosphere of another planet, and it has sparked a debate among astronomers about the origin of the gas and its implications for the potential for life on Venus. While some astronomers believe that the presence of Phosphine gas indicates the presence of some form of life on Venus, others argue that it could be the result of some other process.

A few possible sources of Phosphine gas on Venus explain its presence in the atmosphere. One possibility is that some form of life on Venus, such as bacteria or other single-celled organisms, produces the gas. Another possibility is that chemical reactions in the atmosphere, such as the breakdown of organic matter or the oxidation of Phosphorus-containing minerals, produce the gas. It is also possible that the Phosphine gas is produced by the photochemical reactions of atmospheric gases or by extraordinary volcanism (Bains et al., 2022), lightning, or meteorite impacts. All of these possibilities must be considered when looking for potential life indicators on Venus.

Certain scientists believe that due to the presence of gases such as Phosphine, hydrogen sulfide, nitrous acid (nitrite), nitric acid (nitrate), hydrogen cyanide, and possibly ammonia- which they detected in the atmosphere of Venus, indicate that the Venus clouds are not yet in equilibrium. Additionally, their research raises the possibility of an anaerobic Phosphorus metabolism signature, along with important chemical contributors in photosynthesis without oxygen and the land nitrogen cycle (Mogul et al., 2021).

Some recent studies also suggest that the Phosphine may not be present on Venus at all and that the detection made is simply a result of errors. One cause of such an error can be contamination of SO_2 gas while detection, along with a few minor errors which ultimately caused a skewed result. (Akins et al., 2021; Villanueva et al., 2020) But these arguments were again countered and so on (Greaves et al., 2020).

The detection of Phosphine gas on Venus has sparked a lot of excitement and debate about the potential for life on other planets in our Solar System. Many groups of respectable scientists have written several research papers and articles contradicting each other's claims.

Vol. 2023 (2) 194 – 202
ISSN 2688-3651 [online]

It is obvious that more research into the potential for PH_3 on Venus is required through ground-based or space-based telescopes at millimeter, submillimeter, and infrared wavelengths (Akins et al., 2021).

While the presence of Phosphine gas is not definitive proof of life on Venus, it is an important indicator that could help to guide our exploration of the planet and our search for life on other planets. In the coming years, astronomers will continue to look for more evidence of life on Venus, such as signs of organic molecules or signs of metabolic activity. These observations will help to shed light on the origin of the Phosphine gas and its implications for the potential for life on Venus and beyond.

3. Conclusion

This comprehensive research paper delves into the multifaceted realm of phosphorus in relation to exoplanet exploration and astrobiology. By proposing innovative methods for indirectly detecting phosphorus on exoplanets through the analysis of their parent stars, the study offers a pioneering approach that could revolutionize our understanding of distant worlds. Furthermore, the identification of phosphine gas as a unique biosignature presents a significant leap in our search for life beyond Earth. This paper's interdisciplinary nature, integrating concepts from astronomy, geology, chemistry, and biology, showcases the power of collaboration and cross-disciplinary insights. The debates and discussions surrounding the detection of phosphine in Venus's atmosphere underscore the dynamic nature of scientific inquiry, driving progress and encouraging a reevaluation of our understanding. In its entirety, this research enriches education by providing a compelling real-world context for learning about astrobiology, planetary science, and spectroscopy techniques. It inspires technological advancement by spurring the development of innovative tools for remote sensing and analysis. Additionally, it opens avenues for public engagement, igniting curiosity and interest in space exploration through popular science communication, exhibits, and outreach programs.

The study's findings also have the potential to influence policy decisions related to funding priorities for space missions and scientific research. There are drawbacks to utilizing Phosphorus as a constraint, such as the difficulty of detecting Phosphorus from a distance and the present paucity of available data on Phosphorus. Future research should consider the pros and cons of developing specialized equipment for locating different chemicals and Phosphorus-related features on Exoplanets. The financial costs of developing such technology, the time value aspect, and the skills, education, and knowledge needed to operate and interpret the output from such machinery must all be taken into account in this assessment. By doing this, it may be possible to assess if the advantages of employing Phosphorus as a limitation outweigh the disadvantages and difficulties involved.

Ultimately, this paper lays a robust foundation for future scientific endeavors in exoplanet research, planetary exploration, and the captivating quest for extraterrestrial life.

Acknowledgement

I want to thank my parents and my research mentor for supporting me in writing this research paper.

References

Akins, A. B., et al. (2021). Complications in the ALMA Detection of Phosphine at Venus. https://iopscience.iop.org/article/10.3847/2041-8213/abd56a/pdf

Bains, W., et al. (2019). Trivalent Phosphorus and Phosphines as components of biochemistry in anoxic environments. https://dspace.mit.edu/bitstream/handle/1721.1/124611/ast.2018.1958.pdf?sequence=2&isAllowed=y

Bains, W., et al. (2022). Only extraordinary volcanism can explain the presence of parts per billion Phosphine on Venus. https://www.pnas.org/doi/full/10.1073/pnas.2121702119

Canfield, D. E., et al. (2020). The modern Phosphorus cycle informs interpretations of Mesoproterozoic Era Phosphorus dynamics. https://findresearcher.sdu.dk/ws/files/174538690/Manuscript_revised_Plain_text.pdf

Vol. 2023 (2) 194 – 202
ISSN 2688-3651 [online]

Cockell, C.S., et al. (2016). Habitability: A Review. https://www.researchgate.net/profile/Toby-Samuels/publication/289528782_Habitability_A_Review/links/569e121a08aed27a70319bb8/Habitability-A-Review.pdf

Filippelli, G. M. (2002). The Global Phosphorus Cycle. https://www.researchgate.net/publication/236628350_The_Global_Phosphorus_Cycle

Filippelli, G. M. (2022). Phosphorus and Life on a Water World. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2021GL097346

Greaves, J. S., et al. (2020). Re-analysis of Phosphine in Venus' Clouds. https://arxiv.org/pdf/2011.08176.pdf

Greaves, J. S., et al. (2021). Phosphine Gas in the Cloud Decks of Venus. https://arxiv.org/pdf/2009.06593.pdf

Grenfell, J. L. (2017). A Review of Exoplanetary Biosignatures. https://arxiv.org/pdf/1710.03976.pdf

Hinkel, N. R., Hartnett, H. E. & Young, P. A. (2020). The Influence of Stellar Phosphorus on Our Understanding of Exoplanets and Astrobiology. https://iopscience.iop.org/article/10.3847/2041-8213/abb3cb/pdf

Holstege, C. (2019). Toward an Understanding of Phosphorus Cycling on Waterworlds. https://earth.yale.edu/sites/default/files/files/SeniorEssays/Holstege_thesis_final.pdf

Karl, D. M. (2000, July 6). Phosphorus, the staff of life. *Macmillan Magazines Ltd*. https://www.nature.com/articles/35017686.pdf

Mogul, R., et al. (2021). Venus' Mass Spectra Show Signs of Disequilibria in the Middle Clouds. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL091327

NASA Exoplanet Exploration. Exoplanet Exploration: Planets Beyond our Solar System. Retrieved February 25, 2023, from https://Exoplanets.nasa.gov/

Seager, S. & Deming, D. (2010). Exoplanet Atmospheres. https://arxiv.org/pdf/1005.4037.pdf

Seager, S. (2013). Exoplanet Habitability. http://www.chriscunnings.com/uploads/2/0/7/7/20773630/qualifications_of_habitability.pdf

Sousa-Silva, C., et al. (2020). Phosphine as a Biosignature Gas in Exoplanet Atmospheres. https://www.liebertpub.com/doi/pdf/10.1089/ast.2018.1954

Trujilo, J. C., et al. (2021). Computational Infrared Spectroscopy of 958 Phosphorus-bearing Molecules. https://arxiv.org/pdf/2105.08897.pdf

Villanueva, G., et al. (2020). No evidence of Phosphine in the atmosphere of Venus by independent analyses. https://arxiv.org/abs/2010.14305

Figure 1: Filippelli, G. M. (2022). Phosphorus and Life on a Water World. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2021GL097346

Figure 2: Sousa-Silva, C. (2020). Phosphine as a Biosignature Gas in Exoplanet Atmospheres. https://www.liebertpub.com/doi/pdf/10.1089/ast.2018.1954

Journal of Research High School

Vol. 2023 (2) 203 – 213
ISSN 2688-3651 [online]

Gradient-Based Neural Model Prediction Control for Continuous-Time Control of Quadrotors

Joonyong Choi[1] *

[1]Korea International School Jeju Campus, Republic of Korea
*Corresponding Author: jychoi24@kis.ac

Advisor: Jae Myong Ryu, jmryu@kis.ac

Received June 9, 2023; Revised September 24, 2023; Accepted, October 5, 2023

Abstract

This research proposes a novel deep learning-based control model for quadrotors utilizing gradient-based neural model prediction. The primary objective of this research is to enhance the accuracy and efficiency of quadrotor motion prediction through optimization techniques related to deep neural networks. The proposed model in this research outperforms conventional approaches for quadrotor motion control regarding the accuracy of quadrotor trajectory prediction, which would eventually contribute to precise autonomous control of quadrotors prospected to be widely implemented in drone-related businesses and industries. The model especially becomes more accurate as it adjusts the system to reduce the deviations calculated from the cost function, which occurs during the process of quadrotor motion prediction. Using these techniques, the quadrotor's trajectory and routes were accurately measured through multiple trials, driving the quadrotor to the predetermined destinations accurately with high precisions as well. Therefore, the developed model successfully fulfilled this research's primary objective of developing a deep neural network-based deep learning model to accurately predict quadrotor trajectory for autonomous control. The results gained in this research are highly impactful in both academia and industries as they provide direct and significant benefits on the quadrotor motion development endeavors, aiding control system development and broadening the range for applications of quadrotors across various sectors of industries.

Keywords: Cost function, Deep neural networks, Multi-agent reinforcement learning, Neural model prediction, Proportional integral derivative control

1. Introduction

The emerging advancements in the drone industry holds promising potential as a cornerstone of modern technological and economic landscape. Within the modern landscape, quadrotors are considered one of the most prominent drone models due to the symmetrical characteristics represented by their arranged rotors. The primary objective of this research is to develop a continuous-time neural model prediction control for quadrotors by implementing deep neural networks within the model training process. There are several methods used to build model prediction control models. The first method is the proportional integral derivative control, which contains factors of quick response time, removing offsets, and anticipating the environment. The second method is multi-agent reinforcement learning, which trains multiple agents separately by providing respective rewards and punishments for each agent based on their performances. Finally, the third method, which is the proposed novel approach in this research, is to utilize deep neural networks. By proposing this novel approach for quadrotor motion prediction and control, this research aims to significantly reduce errors in quadrotor motion prediction, thus allowing the reduction of control errors and elevating the accuracy and safety standards for autonomous quadrotor operations. These

Vol. 2023 (2) 203 – 213
ISSN 2688-3651 [online]

advancements would be pivotal in both academia and a myriad of applications within industries such as delivery, agriculture, videography, and many more.

A distinct gap is observed in current quadrotor control models, in which a lack of precise trajectory prediction diminishes the reliability and possibility for broader application of quadrotors. Addressing this gap, the proposed model in this research endeavors to significantly enhance quadrotor trajectory prediction accuracy by leveraging deep learning algorithms in the training process. It is hypothesized that a notable reduction in costs, or deviations, in the quadrotor's motion prediction can be achieved by optimizing the cost function within this model.

Anticipated findings of this research could provide a model that notably improves quadrotor control accuracy, which is indispensable for industries aiming to integrate drones in the most pivotal and critical sectors. The validation of the model's efficacy is demonstrated through rigorous sets of repeated trials, with a keen focus on reducing the cost function, which represents the deviations in the quadrotor's motion prediction.

Furthermore, the research also clearly exhibits the methodologies employed to help the foundational understanding for the readers, including the deep neural networks and the optimization techniques. In the subsequent sections, a detailed exploration of the model's development, experimental setup, and findings are laid out. Eventually, this research demonstrates that the proposed model could potentially propel the drone industry forward and provides anticipations for further research and development in this domain.

2. Related Works

2.1 Model Predictive Control (MPC)

Model Predictive Control (MPC), which is a set of advanced control methods that predicts future behavior of the system, is deeply examined and tested in several research. This review summarizes and provides and comprehensive overview on the latest developments, applications, and strategies to handle computational burden of the Model Predictive Control technique. For the automation of systems, feedback controllers can be divided into three categories: classical, bang-bang, or state controllers. All these controllers use a system model to predict future behaviors and anticipate the deviations of that prediction as well. The PID controller is the best-known controller with significance even in its industrial applications. What makes the MPC method valuable is that it anticipates the behavior of the system and considers hard constraints. Furthermore, MPC relies on models, which allows it to make use of a long range of knowledge and leaves the trivial formulation of an explicit control law. Today, the development of MPC theory is pulled forward by its variety of applications in areas such as the manufacturing technology (Schwenzer, et a l., 2021). Nevertheless, the classical MPC methodologies for devising machine learning or deep learning models tend to be computationally intensive, generating a high possibility of lack of ability to manage non-linear dynamics effectively. The gradient-based deep neural network-based model predictive control proposed in this research aspires to address this deficiency.

2.2 Quadrotor Control Planning

This research identifies the challenges of planning quadrotor trajectories in indoor environments. The research proposes a way to extend the existing work on polynomial trajectory generation through presenting a method of jointly optimizing polynomial path segments in an unconstrained quadratic program that is stable for high-order polynomials and large numbers of segments. Furthermore, this research also suggests a technique for an automation of selecting the allocated time for each segment, which allows the quadrotor to speed along the path. Applying polynomial trajectories eliminates the need for computationally intensive sampling and simulation when handling motion. The proposed approach creates high-quality trajectories significantly faster than purely sample-based optimal planning methods. The performance of the algorithm of efficiently generated trajectories is also provided (Richter, et al., 2013). However, despite these advancements, producing optimal trajectories for quadrotors based on prediction in dynamically changing environments remains a challenging task, which is an issue that this research aims to address with the novel deep-learning based quadrotor trajectory prediction model for autonomous control.

Vol. 2023 (2) 203 – 213
ISSN 2688-3651 [online]

2.3 Learning-Based Motion Control

Learning-based methods have shown favorable performance for accelerating motion planning, but one shortcoming is that they were mostly created for the setting of static environments. For a more challenging planning of dynamic environments, such as multi-arm assembly tasks or human-robot connection, motion planners should consider the trajectories of the dynamic obstacles and reason the temporal-spatial interactions in a very large extent. This research proposes GNN-based approach that uses temporal encoding and imitation learning with data aggregation for learning embeddings and the edge prioritization policies. The experiments performed show that proposed methods significantly accelerate online planning. Also, the models reduce costly collision, and therefore accelerate planning by up to 95% while achieving high success rates on challenging cases at the same time (Zhang, et al., 2022). Nonetheless, conventional learning-based approaches often lack the real-time adaptive capabilities essential for safe and efficient quadrotor operation in complex and dynamic environments. They are often deficient on their ability to make quick real-time predictions as well. As a result, this research endeavors to overcome this deficiency of conventional learning-based motion control through the integration of gradient-based neural model prediction and deep learning algorithms.

3. Theoretical Background

3.1 Proportional Derivative (PD) Control

PD control is a type of controller in which the output is printed proportionally to the error signal together with the derivative of the error signal. This controller provides the functions of both the proportional controller and the derivative controller.

The proportional controller is a feedback controller that minimizes the fluctuation in the process variable, but it often forms deviation from the set point. The advantage of proportional controllers is that it provides a faster response than most controllers. The P-only controller could respond even faster as the system has more complicated algorithms. However, the deviation, known as offset, from the set point increases, which could eventually result in an error.

The derivative controller is a form of forward control. The D-control analyzes the change of error and predicts process conditions. The advantage of a D-controller is to resist change, which is especially important when facing oscillations (Brogan, 1990).

$$\begin{bmatrix}\tau_\phi\\ \tau_\theta\\ \tau_\psi\end{bmatrix}=\begin{bmatrix}-I_{xx}\left(K_d\dot{\phi}+K_p\int_0^T\dot{\phi}\,dt\right)\\ -I_{yy}\left(K_d\dot{\theta}+K_p\int_0^T\dot{\theta}\,dt\right)\\ -I_{zz}\left(K_d\dot{\psi}+K_p\int_0^T\dot{\psi}\,dt\right)\end{bmatrix} \quad (1) \qquad \tau_B=\begin{bmatrix}Lk(\gamma_1-\gamma_3)\\ Lk(\gamma_2-\gamma_4)\\ b(\gamma_1-\gamma_2+\gamma_3-\gamma_4)\end{bmatrix}=\begin{bmatrix}-I_{xx}\left(K_d\dot{\phi}+K_p\int_0^T\dot{\phi}\,dt\right)\\ -I_{yy}\left(K_d\dot{\theta}+K_p\int_0^T\dot{\theta}\,dt\right)\\ -I_{zz}\left(K_d\dot{\psi}+K_p\int_0^T\dot{\psi}\,dt\right)\end{bmatrix} \quad (2)$$

$$T=mg \quad (3) \qquad T_{proj}=mg\cos(\theta)\cos(\phi) \quad (4)$$

$$T=\frac{mg}{\cos(\theta)\cos(\phi)} \quad (5) \qquad T=\frac{mg}{\cos(\theta)\cos(\phi)}=k\sum\gamma_i \Rightarrow \sum\gamma_i=\frac{mg}{k\cos(\theta)\cos(\phi)} \quad (6)$$

$$\gamma_1=\frac{mg}{4k\cos(\theta)\cos(\phi)}-\frac{2be_\phi I_{xx}+e_\psi I_{zz}kL}{4bkL} \quad (7) \qquad \gamma_2=\frac{mg}{4k\cos(\theta)\cos(\phi)}+\frac{e_\psi I_{zz}}{4b}-\frac{e_\theta I_{yy}}{2kL} \quad (8)$$

$$\gamma_3=\frac{mg}{4k\cos(\theta)\cos(\phi)}-\frac{-2be_\phi I_{xx}+e_\psi I_{zz}kL}{4bkL} \quad (9) \qquad \gamma_4=\frac{mg}{4k\cos(\theta)\cos(\phi)}+\frac{e_\psi I_{zz}}{4b}+\frac{e_\theta I_{yy}}{2kL} \quad (10)$$

3.2 Proportional Integral Derivative (PID) Control

PID control is a type of controller that is a combination of all three types of control methods: proportional, integral, and derivative control. So, it contains integral control in addition to the proportional control and derivative control priorly introduced.

Vol. 2023 (2) 203 – 213
ISSN 2688-3651 [online]

The integral control is also a type of feedback controller which is used to remove deviations. The controller makes the system return to its steady state. However, the I-only controllers have much slower response speeds compared to the P-only controllers because they are more dependent on the parameters.

The PID control uses the P-only control for a quick response time and removes the offset created by the P-only control using the I-only control. Furthermore, the D-control increases the controller's response because it anticipates the conditions. The PID controller is one of the best controllers which is highly accurate and stable (Brogan, 1990).

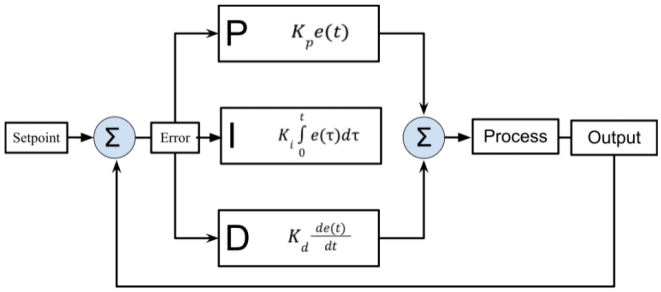


Figure 1. The Diagram of the PID Control (Mehta, Nikunj., et al., 2017)

$$e_\phi = K_d\dot{\phi} + K_p\int_0^T \dot{\phi}\,dt + K_i\int_0^T \dot{\phi}\,dt\,dt \quad (11)$$

$$e_\theta = K_d\dot{\theta} + K_p\int_0^T \dot{\theta}\,dt + K_i\int_0^T \dot{\theta}\,dt\,dt \quad (12)$$

$$e_\psi = K_d\dot{\psi} + K_p\int_0^T \dot{\psi}\,dt + K_i\int_0^T \dot{\psi}\,dt\,dt \quad (13)$$

$$F_i = k_F \cdot P_i^2 \quad (14)$$

$$T = \sum_{i=0}^{3}(-1)^{i+1} k_T \cdot P_i^2 \quad (15)$$

$$\ddot{x} = \left(R \cdot \left[0,0,k_F \sum_{i=0}^{3} P_i^2\right] - [0,0,mg]\right)m^{-1} \quad (16)$$

$$\dot{\psi} = J^{-1}\left(\varkappa \times (L,k_F,k_T,[P_0^2,P_1^2,P_2^2,P_3^2]) - \psi \times (J\psi)\right) \quad (17)$$

3.3 Reinforcement Learning (RL)

Reinforcement Learning is a technique of training machine learning models in order to make a series of decisions. The agents of the models learn to achieve a goal in a certain environment. In reinforcement learning, the AI attempts several trials and experiences failures until it achieves the goal, which is in most cases solving a problem. To set the AI to do a certain task, the AI receives rewards or penalties depending for its actions and achievements. Its goal is to maximize the total reward received. Reinforcement learning has been widely applied in the robotics fields for decades, and it had confined to higher-level decisions. However, recent research proves that reinforcement learning could also serve as a basis for serving aerial vehicles, such as quadrotors. A model-based reinforcement learning could be used to train a locally weighted linear regression policy, which could lead to a more dynamic motion of the quadrotor (Hwangbo, et al., 2017).

$$E[\sum_{t\geq 0}\gamma^t R(s_t,a_t,s_{t+1})|a_t\sim\pi(\cdot\,|s_t),s_0] \quad (18)$$

$$\varrho_\pi(s,a) = E[\sum_{t\geq 0}\gamma^t R(s_t,a_t,s_{t+1})|a_t\sim\pi(\cdot\,|s_t),a_0=a\,] \quad (19)$$

$$V_\pi = E[\sum_{t\geq 0}\gamma^t R(s_t,a_t,s_{t+1})\,|a_t\sim\pi(\cdot\,|s_t),s_0=s] \quad (20)$$

3.4 Multi-Agent Reinforcement Learning (MARL)

Multi-agent reinforcement learning is a type of reinforcement learning, which is a field of training machine learning models to make a series of decisions. For many years, reinforcement learning was enacted through single agents. However, since there are many situations where agents should make independent decisions, it is better to apply multi-agent reinforcement learning in most cases. The advantage of multi-agent reinforcement learning is that it allows us to explore all different alignments appearing when multiple agents share an environment and know how those alignments affect each agent. Each agent is motivated by its own rewards, similar to the single-agent reinforcement learning. However, compared to the single-agent reinforcement learning, multi-agent reinforcement learning is more complex as the technique combines the pursuit of finding ideal algorithms that maximizes awards along with more sociological concepts. Since there are multiple agents, machine learning might not be effective as a whole since each

Vol. 2023 (2) 203 – 213
ISSN 2688-3651 [online]

agent works for its own rewards. Therefore, it is also important to evaluate the social metrics of the agents such as cooperation, reciprocity, equity, and social influence (Panerati, et al., 2021; Zhang, et al., 2021).

3.5 Neural Differential Equations

Neural Differential Equations are continuous depth generalizations on standard layer to layer propagation. They use the neural network to parameterize the vector field. The purpose of using neural differential equations is to make the computation graph continuous so that it is actually differentiable. The residue network is formulated as $y_{j+1} = y_j + f_\theta(j, y_j)$, where $f_\theta(\mathrm{j},\cdot)$ is the j-th residual block. The ordinary differential equation was $\frac{dy}{dt}(t) = f_\theta(\mathrm{t}, \mathrm{y(t)})$.

By discretizing this via the Euler method, $\frac{y(t_{j+1})-y(t_j)}{\Delta t} \approx \frac{dy}{dt}(t_j) = f_\theta(t_j, y(t_j))$ is obtained, which shows that neural differential equations are the continuous limit of residual networks. There is also the physics-informed neural network (PINN), which is effective for the AI to recognize physical conditions of the environment similar to that of the human intuition process. However, when working with a model, there is a structural problem called the 'inductive bias.' There is always some gap between the theoretical prediction and the actual values. As a result, the high-capacity function approximation is used as a solution to close the gap between theory and actual conditions.

There are certain ways to train the universal differential equations. One option is to use the parameters to initialize and only train the network parameters. Another option is to regularize the neural network so that it only fits to the residual between the theoretical model and actually observed data values (Kidger, 2021; Sangalli, et al., 2022).

3.6 Model Predictive Control (MPC)

Model Predictive Control is an advanced method of process control used to make prediction about future outputs in a set of constraints. It is a widely used concept in process industries such as petroleum and chemistry industries. The model predictive controllers are dependent on dynamic models. The primary advantage of model predictive control is that it allows the optimization of current timeslots, while considering future timeslots for prediction. As a result, this makes the model predictive control to anticipate future events and automatically control the object's motion. This method is a widespread digital control model used in many fields in modern days.

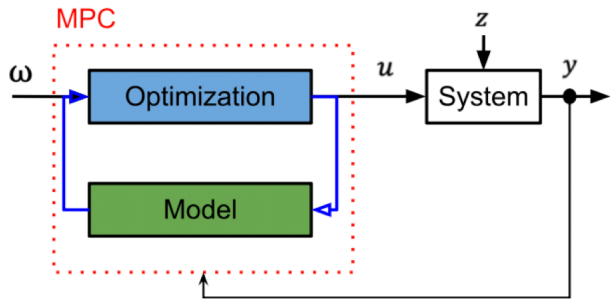


Figure 2. The Diagram of the MPC Process (Schwenzer, M., et al., 2021)

3.7 Gradient Descent

Gradient descent is an optimization algorithm used to train machine learning/deep learning models and neural networks in the manner of minimizing possible errors. Gradient descent method is important because it enables the model to be trained efficiently and effectively to achieve the goal. The gradient descent algorithm acts similarly to a commonly known quadratic function. The gradient descent method is used based on the cost function of a model, which is a function that measures the deviation, or error, between the actual output and the predicted output. From the starting point of the function, the derivative (slope) of the graph is determined to measure the steepness of the graph. Moreover, the slope would inform the parameters, possibly weights and bias, of the changes to be applied. Then, this process is repeated until it finds the lowest point on the cost function curve, the point of convergence, which would have a derivative nearly zero and the lowest cost. For the gradient descent to work, it requires two data points, direction and learning rate, which determine the partial derivative that would contribute to finding the minimum, which would

Vol. 2023 (2) 203 – 213
ISSN 2688-3651 [online]

be the point of convergence. Learning rate is the step size taken in the process to find the minimum. High learning rate means the step size is large, which indicates a possibility of more error. On the other hand, low learning rate means the step size is small, which means it would have more precision but less efficiency since it would require more computation. There are primarily three types of gradient-descent. The first one is the batch gradient descent. This method takes the sum of the error of each training set, and it is called a training epoch. Batch gradient descent provides computation efficiency, but in exchange it takes long processing time. Another type of gradient descent is stochastic gradient descent. This method runs a training epoch for each sample and updates parameters once at a time. It has faster processing speed compared to the batch gradient descent, but it can have losses in computation efficiency. Finally, the mini-batch gradient descent method combines the concept of batch and stochastic gradient descent. It divides the dataset into small batches and updates parameters of each of those batches, creating a balance between processing speed and computation efficiency.

As an example of a gradient descent model, the cost function of the training set is as the following equation (21), where L is the loss function per example, $f(x;\theta)$ is the predicted output for input x, and $\hat{p}_{data}$ is the empirical distribution. The optimization of the model is represented as equation (22), where m is the number of training examples.

$$J(\theta) = E_{(x,y)\sim\hat{p}_{data}} L\big(f(x;\theta), y\big) \quad (21)$$

$$E_{x,y\sim\hat{p}\,data(x,y)}[L(f(x;\theta),y)] = \frac{1}{m}\sum_{i=1}^{m} L\big(f\big(x^{(i)};\theta\big), y^{(i)}\big) \quad (22)$$

The maximum likelihood estimation problems decompose into a sum over each example.

Maximizing the sum is equal to maximizing the expectation over empirical distribution defined by the training set, which is the following in equation (24). The most commonly used property is the gradient in equation (25).

$$\theta_{ML} = \arg max \sum_{i=1}^{m} logp_{model}\big(x^{(i)}, y^{(i)};\theta\big) \quad (23)$$

$$J(\theta) = E_{x,y\sim\hat{p}_{data}} logp_{model}(x,y;\theta) \quad (24)$$

$$\nabla_\theta J(\theta) = E_{x,y\sim\hat{p}_{data}} \nabla_\theta \log p_{model}\,(x,y;\theta) \quad (25)$$

The stochastic gradient descent, an optimization algorithm for an objective function to find the model parameters that correspond to the best fit between predicted and actual values, minimizes the generalization error faster than other gradient descent algorithms. Also, it is especially suitable for large-scale datasets. The generalization error could be written as the following equation (26). The exact equation for the gradient is represented in equation (27).

$$J^*(\theta) = \sum_x \sum_y p_{data}(x,y) L(f(x;\theta),y) \quad (26)$$

$$g = \nabla_\theta J^*(\theta) = \sum_x \sum_y p_{data}(x,y)\, \nabla_\theta L(f(x;\theta),y) \quad (27)$$

For the experiment of direct optimal control of inverted pendulum with a torsional spring and the gradient-based control experiment for quadrotors, which would be articulated in section 4, both uses the *Adam* optimizer.

Adam optimizer is the extended version of stochastic gradient descent which is expected to be applied in various deep learning fields such as computer vision and natural language processing in the future. This algorithm can serve as an alternative option of the stochastic gradient descent. *Adam* uses estimations of the first and second moments of the gradient in order to adapt the learning rate for each weight of the neural network. So, since the *Adam* optimizer has better generalizing performance, it was adopted in both experiments.

The main component of the *Adam* optimizer is the Momentum algorithm. The momentum algorithm considers the exponentially weight average and expedites the gradient descent process. The momentum algorithm is divided into two parts. The first part calculates the displacement, or the position change, while the second part updates the prior position. The change in position is given as the equation (28), where α is the stepsize, or in other words, the hyperparameter which controls the movement in search space, also called the learning rate.

$$update = \alpha \times m_t \quad (28)$$

$$w_t + 1 = w_t - update \quad (29)$$

The new position of the or weights at time t is given by the equation (29). Furthermore, $f'(x)$ is the derivative or aggregate of gradients at time t. The equation (30) below shows the aggregate gradients at time t, represented as m_t and $m_t - 1$. The momentum algorithm dampens down the change of the gradient and the step size with each new point of the search space (Goodfellow, et al., 2016).

$$m_t = \beta_1 m_t + (1-\beta_1)\left(\frac{\delta L}{\delta w_t}\right) \quad (30)$$

Vol. 2023 (2) 203 – 213
ISSN 2688-3651 [online]

4 Materials and Methods

This section shows experimental data, measurements, and observations. No explanations or interpretations are expected in this section and those information needs to be addressed in the discussion section. All tables, figures, and equations should be located in the proper positions and all descriptive explanation needs to be referred from the context in the body.

The experiment described in this section was conducted by an M1 MacBook Pro Max 2021 laptop, a 14-inch version, which has 10 cores of Central Processing Unit (CPU) and 32 cores Graphics Processing Unit (GPU). Regarding the software, this research utilized the Visual Studio Code Integrated Development Environment (IDE) and specifically employed the Python3 programming language. The experiment was only based on simulation settings, and results were measured by how closely the quadrotor reaches to the predetermined destination point in a 3-dimensional coordinate space. After the experiment trials were completed, the range of the predicted position within the predetermined destination and the error bound derived from that range were obtained to quantify the accuracy of the model's prediction.

4.1 Direct Optimal Control of an Inverted Pendulum with a Torsional Spring

Using the theories listed in prior sections, it is also possible to control an inverted pendulum with a torsional spring and stabilize it in the upright position. The equation (31), which is in Hamiltonian form represents the inverted pendulum with a torsional spring.

$$\begin{bmatrix}\dot{q}\\ \dot{p}\end{bmatrix} = \begin{bmatrix}0 & 1/m\\ -k & -\beta/m\end{bmatrix}\begin{bmatrix}q\\ p\end{bmatrix} - \begin{bmatrix}0\\ mgl\sin(q)\end{bmatrix} + \begin{bmatrix}0\\ 1\end{bmatrix}u \tag{31}$$

First, to set up the experiment for the inverted pendulum, *matplotlib* and *torch* should be imported to the system. Also, the device should be set up according to the configuration. Then, to control the pendulum, the proper integral cost function must be derived so that the loss to be minimized through training could be accurately represented. The cost function is defined as the following equations (32) and (33), where x is the state and u_θ is the controller. x^* and u^* are the desired position, and the matrices P, Q, R represent the weights for altering the performance (Markus, et al., 1976).

$$\min J = \left(x(t_f) - x^*\right)^T \boldsymbol{P}\left(x(t_f) - x^*\right) \tag{32}$$

$$\int_{t_0}^{t_f}[(x(t) - x^*)^T \boldsymbol{Q}(x(t) - x^*) + (u_\theta(t) - u^*)^T \boldsymbol{R}(u_\theta(t) - u^*)]\, dt \tag{33}$$

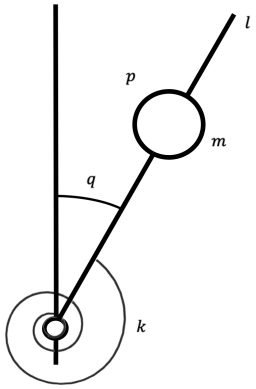


Figure 3. Diagram of Inverted Pendulum with Torsional Spring (Pierre, Coullet., Gilli, J.M., & Rousseaux, Germain., 2009)

After setting the proper integral cost function, the initial conditions such as the time span, step size, and initial and final time should be declared as variables and assigned input values. For this inverted pendulum experiment, a box-constrained controller would be used since there should be a given limited control input for the parameters p and q, which respectively represent the angular position and angular speed of the inverted pendulum. Finally, the optimization loop would be run. Specifically, the stochastic gradient descent method would be used along with *Adam* to optimize the parameters.

4.2 Gradient-Based Control for Quadrotors

This research uses Model Predictive Control (MPC) in order to enable the quadrotor to predict the future behavior of its system to predict the trajectory. The formulation for this Model Predictive Control (MPC) could be written as the following equations (34), (35), and (36), where J is the

$$min \sum_{k=0}^{T-1} J\,(x_k, u_k) \tag{34}$$

$$subject\ to\ \dot{x}(t) = f\left(t, x(t), u(t)\right),\ \ t \in ך \tag{35}$$

$$x(0) = x_0 \tag{36}$$

Vol. 2023 (2) 203 – 213
ISSN 2688-3651 [online]

cost function and $T \in \mathbb{T}$ is the model predictive control over which predicted future trajectories are optimized.

The quadrotor model is an appropriately modified version of explicit dynamics for batched training in *PyTorch*. The following equations (37) and (38) are for the accelerations of the quadrotor model, where $\boldsymbol{x} = [x, y, z]$ represents the quadrotor's 2-dimensional position and $\boldsymbol{\psi} = [\phi, \theta, \psi]$ represents the quadrotor's angular position. $\boldsymbol{R}$ and $\boldsymbol{J}$ are the quadrotor's rotational and inertial matrices respectively, and the function τ is a function that calculates the torque created by the motor generated with the speed of w_i, arm length of l, mass m, acceleration due to gravity g, and quadrotor's physical properties K_F, K_T. The initial conditions and some constants such as g are determined before the model is trained.

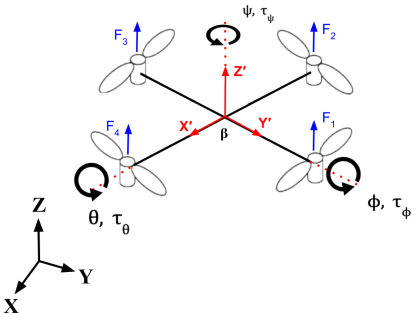


Figure 4. Diagram of the Quadrotor Dynamics

$$\ddot{x} = \left(\boldsymbol{R} \cdot \left[0,0,k_F \textstyle\sum_{i=0}^{3} \omega_i^2\right] - [0,0,mg]\right)m^{-1} \quad (37)$$

$$\ddot{\psi} = J^{-1}\left(\tau(l, k_F, k_T, [\omega_0^2, \omega_1^2, \omega_2^2, \omega_3^2]) - \dot{\psi} \times \left(\boldsymbol{J}\dot{\psi}\right)\right) \quad (38)$$

Then, the cost function is programmed. First, I set a goal that the quadrotor should be reached. The goal given to the quadrotor was to reach the position of $[x, y, z] = [3,3,3]$ under the circumstance that the quadrotor is initially at rest and has no rotation and velocities at the end. The cost function is programmed so that it could measure the deviations between the actual quadrotor's behavior and the predictions made by the Model Prediction Control (MPC) and contribute into making a more accurate prediction.

The initial conditions should be programmed after setting the cost function. Since the motion of the quadrotor initially begins at rest, all variables of the quadrotor motion and position should be set to 0. This includes the $[x, y, z]$ coordinates of the quadrotor's 3-dimensional position and the $[\phi, \theta, \psi]$ coordinates of the quadrotor's angular position.

The controller that would be used is a simple controller based on the Neural Network with bounded control inputs of the quadrotor. The input dimension, output dimension, hidden dimension, and the initialization of the last layer of the Neural Network to zero are included in these bounds.

The next step is the MPC initialization. First, the time step, noted as dt, is set to 0.02, and the final time would be 2 seconds. Then, a horizontal line would be programmed. This means that the quadrotor would be tested for 2 seconds with 0.02 seconds of intervals. Also, the learning rate would be set up, which means that the training would be repeated until the error is lower than the value that is set up. After that, the parameters would be optimized by the optimizer known as *Adam*. Finally, the MPC receives the data of every component and sends it to the device. At the end, the MPC finally goes through a simulation which produces the trajectories through repeated multiple trials.

5 Results

5.1 Inverted Pendulum with a Torsional Spring

As the optimization loop is run, the training loop is run, and the losses are plotted as a result with respect to the progress of epochs. In the result, represented in Figure 5, the diminishing trend of losses as the epochs are progressed are shown.

The final plot results of the pendulum trajectories are derived after the losses. The graphs for each of the variables p and q, respectively representing the position and angle of the pendulum, are obtained after the experiment. The final plot results for those variables are as the following.

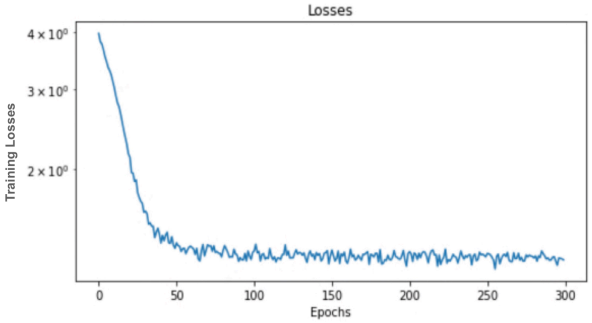


Figure 5. The Subplot of Losses with respect to the Epochs

Vol. 2023 (2) 203 – 213
ISSN 2688-3651 [online]

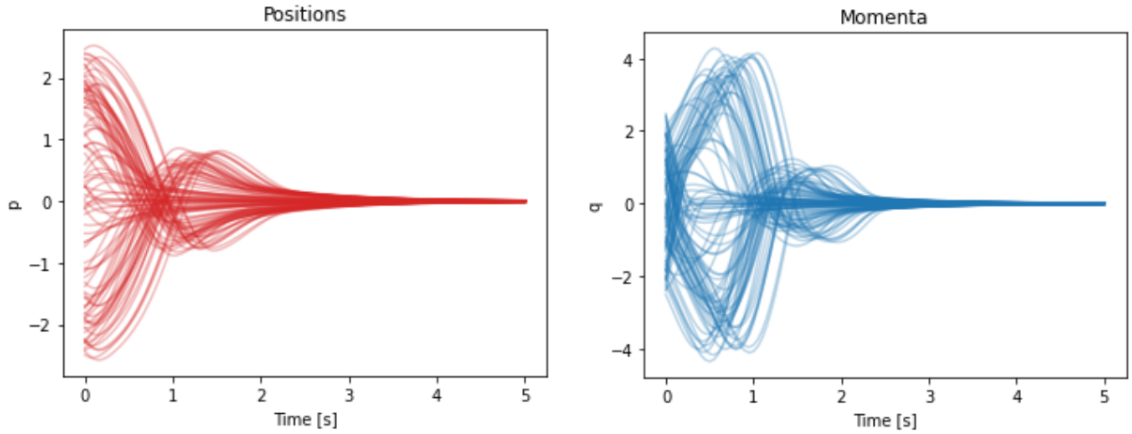


Figure 6. Plot Results of Pendulum Trajectories of each Variables

Ultimately, considering Figure 6, the overall phase space of the pendulum could be solely represented by p and q, as shown in Figure 7.

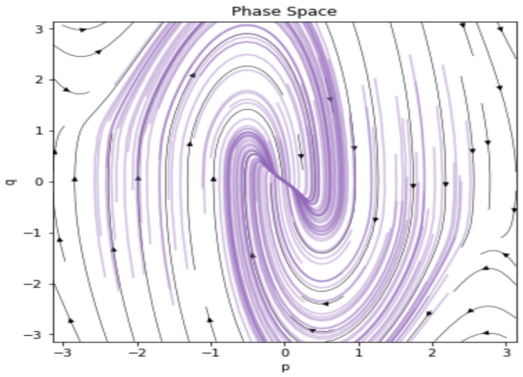


Figure 7. Plot Result of Pendulum for p and q

5.2 Gradient-Based Control for Quadrotors

During the simulation, the four propellers of the quadrotors are controlled. The change of the control of each propeller, in RPM, is shown as the following in Figure 8.

Due to the change of control inputs in the time interval, the trajectories for each of the twelve variables (linear positions x, y, z; angular positions ϕ, θ, ψ; linear velocities $\dot{x}, \dot{y}, \dot{z}$; angular velocities $\dot{\phi}, \dot{\theta}, \dot{\psi}$) were derived as the following in Figure 9.

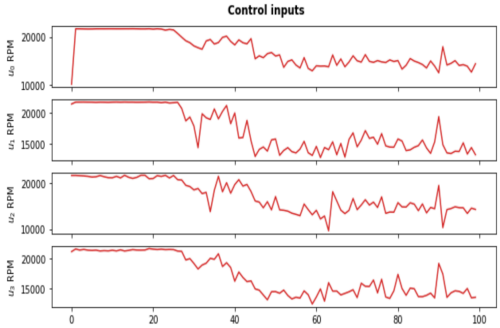


Figure 8. Change in Control Inputs of Each Propeller

Finally, as these trajectories for all twelve variables were derived, the complete trajectory of the quadrotor in the 3-dimensional space was obtained considering these changes in the variables in given time. The final trajectory of the quadrotor was derived as the following in Figure 10.

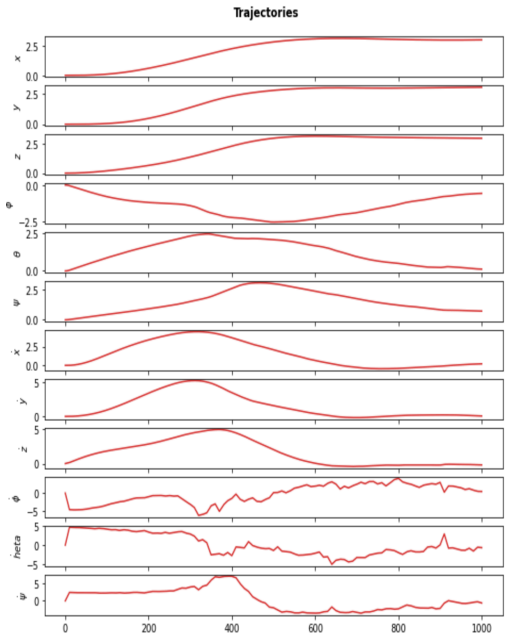


Figure 9. Trajectories of Twelve Variables of Quadrotor Motion after MPC simulation

Vol. 2023 (2) 203 – 213
ISSN 2688-3651 [online]

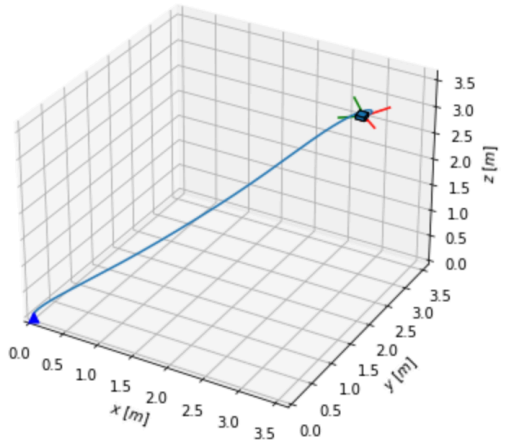


Figure 10. The 3-D Trajectory of the Quadrotor

6. Discussion

The result of the experiment was successfully derived as demonstrated visually in Figure 10. Using methods of MPC, cost function, initial conditions, controller, MPC initialization, and simulation, the quadrotor reached about a final destination of approximately $[x, y, z] = [2.974, 3.015, 3.009]$ compared to the predetermined destination of $[x, y, z] = [3, 3, 3]$, which is both in the units of meters. As a result, the error bounds of the axes are approximately 0.026, 0.015, and 0.009 meters respectively for each x, y, z axis respectively. Calculating the percent error based on these error bounds in meters, the absolute percentage error for each x, y, z axis is approximately 0.872%, 0.498%, and 0.299%, respectively. As a result, the mean absolute percent error across all the dimensions is approximately 0.556%. Therefore, based on these statistics, the devised model provided a significantly accurate final position of the quadrotor utilizing the techniques, control inputs, and trajectories of the twelve variables that represent the quadrotor motion, which are represented in Figure 9.

The experiment was successful, as a gradient-based neural model prediction control for continuous time-control of quadrotors was created and verified through simulation. The supposed benefit of this model is that it improved the accuracy and efficiency of quadrotor motion prediction using optimization techniques related to deep neural networks. Also, the model has reduced deviations calculated from the cost function, promising more accurate quadrotor trajectory prediction essential for autonomous control. Furthermore, as the model implements deep neural networks, it is likely to be more effective in complex environments as well with reduced costs and control errors. Eventually, this would enable broader applications of quadrotors in academia and industries, contributing to the advancement of quadrotor implementation in various areas such as delivery, agriculture, videography, and many more.

The limitation of this research is that this model still experiences significant computational intensity because it uses MPC as one of its methods for training. Also, since the experiment was only done at a simulation-based setting, it might be difficult for immediate and direct implementation in real-world, dynamic environment settings. Although it is possible to use in those kinds of environments, it is likely to demonstrate diminished performance compared to the experimental results shown in this paper. Finally, the research needs some reinforcement in indoor setting environments in the real-world as well.

Some potential solutions to these limitations would be utilizing more advanced deep learning and optimization techniques in the training process to significantly enhance quadrotor trajectory prediction accuracy. Also, combining the MARL technique along with the deep neural network technique implemented in this research could also serve as a feasible solution to address the limitation. Furthermore, extending the existing work through joint optimizing polynomial path segments in an unconstrained quadratic program or utilizing imitation learning along with data aggregation could potentially aid quadrotor motion prediction in dynamic environments and help improve quadrotor control planning.

7. Conclusion

The reason why this development is important is because it creates an accurate automatic quadrotor control model using the techniques that enable the system to anticipate the motions of the quadrotor precisely. Eventually, the suggested quadrotor control model could be applied in many areas of research or industries in the future.

For further development, it would be good to increase the accuracy of the model through constant training of the model. Moreover, as mentioned priorly, the results derived from this research could be applied in many areas, such as the delivery industry and agriculture. As many companies are developing drone deliveries, the model for accurate prediction of quadrotor trajectories and destination would be the most critical technology required for them. Also, in

Vol. 2023 (2) 203 – 213
ISSN 2688-3651 [online]

terms of agriculture, people are attempting to use drones, especially quadrotors, to check water levels and farmland conditions as well as any risks in surrounding areas. The neural model prediction control for quadrotors would be significant during this as it would allow the users to know the accurate trajectories and the final location of their quadrotors. As suggested in these examples, the results of this research would be highly influential and contributive to the development of the drone industry and society.

References

Brogan, W. L. (1990). Modern Control Engineering Theory Third Edition.

Conti, R., Markus, L., & Olech, C. (1976). Control Theory and Topics in Functional Analysis.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Deep Learning.

Hwangbo, J., et al. (2017, July 17). *Control of a quadrotor with reinforcement learning*. arXiv.org.

Kidger, P. (2022, February 4). *On neural differential equations*. arXiv.org.

Mehta, N., et al. (2017). Design of HMI Based on PID Control of Temperature. International Journal of Engineering Research and Technology. V6. 10.17577/IJERTV6IS050074.

Panerati, J., et al. (2021, July 25). *Learning to fly -- a gym environment with pybullet physics for reinforcement learning of Multi-agent Quadcopter Control*. arXiv.org.

Pierre, C., Gilli, J. M., & Rousseaux, G. (2009). On the critical equilibrium of the spiral spring pendulum. Proceedings of The Royal Society A Mathematical Physical and Engineering Sciences. 466. 407-421. 10.1098/rspa.2009.0393.

Richter, C., Bry, A., & Roy, N. (2013). Polynomial Trajectory Planning for Aggressive Quadrotor Flight in Dense Indoor Environments.

Sangalli, S., et al. (2022, January 5). *Constrained Optimization to train neural networks on critical and under-represented classes*. arXiv.org.

Schwenzer, M., et al. (2021, August 11). *Review on Model predictive control: An engineering perspective - the International Journal of Advanced Manufacturing Technology*. SpringerLink.

Zhang, K., Yang, Z., & Başar, T. (2021, April 28). *Multi-Agent Reinforcement Learning: A selective overview of theories and algorithms*. arXiv.org.

Zhang, R., et al. (2022, December 6). *Learning-based motion planning in dynamic environments using GNNS and temporal encoding*. Advances in Neural Information Processing Systems.

Journal of Research High School

Vol. 2023 (2) 214 – 225
ISSN 2688-3651 [online]

Recent Advancements of Cherenkov Radiation for Cancer Imaging and Treatment

Andrew Lee[1] *

[1]Avon Old Farms School, Avon, CT, USA
*Corresponding Author: joonseoandrewlee@gmail.com

Advisor: Federico Nova, federico.nova@stfc.ac.uk

Received January 17, 2023; Revised August 23, 2023; Accepted, September 25, 2023

Abstract

Since successful bioimaging in 2009, many studies have explored Cherenkov Radiation's potential value in medical diagnostics and treatment. By highlighting major accomplishments to date, this review proposes that Cherenkov Radiation has strong potential to become a novel class of modality in cancer molecular imaging. The basic principles of Cherenkov Radiation are explained. Variable conditions such as threshold energy and refractive index of tissues are described in the context of cancer diagnostics. Cherenkov luminescence imaging and tomography are outlined, as well as the milestone clinical trials with respective strengths and weaknesses. Application of Cherenkov in radiation dosimetry is discussed. Current research in Cherenkov photodynamic therapy is introduced, with many studies implying better penetrance than conventional therapy. Finally, current limitations of Cherenkov Radiation, especially its inability to reach deeper anatomical structures are examined. Despite significant challenges to overcome, applications of Cherenkov Radiation are gaining momentum in the field of oncology.

Keywords: Identity, Gender, Socioeconomic class, Education, Politics

1. Introduction

Fundamentally, molecular imaging allows for visualization of the biochemical processes of an active cell, without disturbance of the target cell and its environment (Rowe & Pomper, 2022). Molecular imaging is especially useful for cancer diagnostics and treatment monitoring, as cancer cells behave very differently compared to normal cells in metabolism. Cancer cells are characterized by their ability to rapidly grow and divide. In order undergo uncontrolled proliferation, cancer cells must consume additional nutrients to meet the increased energy demands (Chen et al., 2007). Molecular imaging utilizes the abnormal metabolic rates by using tracers that participate in the metabolic processes. Tracers may be radiolabeled or have other intrinsic properties that can be detected (Rowe & Pomper, 2022). Once a tracer is thought to be taken within the cell, a sensor or a scanner can detect the tracer and translate the detection into spatial representations such as images or graphs. The most ideal molecular imaging will have a high-affinity tracer that follows metabolic process unique to cancer, that can be detected by a scanner with high sensitivity, and high spatial, contrast and temporal resolutions. In addition, the imaging must be affordable to allow continuous, repetitive use for each patient. The current molecular imaging modalities used in oncology are limited in variability. Clinically, the most common are Magnetic Resonance Spectroscopy (MRS) and radionuclide modalities such as Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET). Although widely available, with low sensitivity and limited metabolites, MRS is used mostly for brain and prostate cancers in clinical settings (Hernot et al., 2019). SPECT is also widely available but is challenged by low resolution and low sensitivity (Wilson et al., 2020). PET is highly sensitive, and carries potential for high specificity, with more tracers being developed (Jadvar, 2016). However, its current high cost prevents widespread use. Therefore, improving current modalities and

Vol. 2023 (2) 214 – 225
ISSN 2688-3651 [online]

developing new imaging methods are necessary to advance the field of oncology.

Cherenkov radiation (CR) is a form of light emission that occurs when a charged particle exceeds the local speed of light in a dielectric medium. CR biomedical imaging was first successful in 2009, and many innovative studies produced high quality animal and human tissue imaging (Robertson et al., 2009). CR's unique broadband emission spectrum, spectral weight in the ultraviolet and blue wavebands, and local generation of light within a given tissue, have made it an attractive new modality in cancer imaging and treatment. This review will first introduce the general concept of CR, and then highlight the major milestones of CR applications in the field of oncology. Most articles were chosen from keywords "Cherenkov Radiation", "oncology", "Cherenkov luminescence imaging" searches in Pubmed and OVID databases. Others were excavated from articles found from the databases. All articles were accessed for credibility and were peer reviewed. By summarizing the notable accomplishments from 2009 to date, the authors intend to demonstrate that CR has the full potential to become significant and permanent class of modality in molecular imaging for oncology.

2. Cherenkov Radiation Theory and Required Conditions

In 1988, Oliver Heaviside predicted that a point charge would produce a conical wavefront whenever it was to travel at a speed greater than the speed of light in a medium as shown in Figure 1 (Ciarrocchi & Belcari, 2017; Das & Boruah, 2022). This theory later explained the 1934 phenomenon when Cherenkov observed a blue glow surrounding a radium salt solution.

Light can be slowed by the medium it passes, and its speed may be significantly less than speed of light in vacuum, c = 299,792,458 m/s. For example, the speed of light in water compared to in vacuum is reduced by 25%. Cherenkov radiation occurs when a charged particle passes through a polarizable dielectric medium with a speed greater than light's speed in that medium. As the charged particle travels, it rapidly polarizes the nearby molecules, aligning them in a polarizing field. As these polarized molecules return to their ground state, they emit light photons, observed as luminescence radiation. The emitted light will travel as a wavefront in the direction at an angle θ from the direction of particle travel, shown in Figure 1.

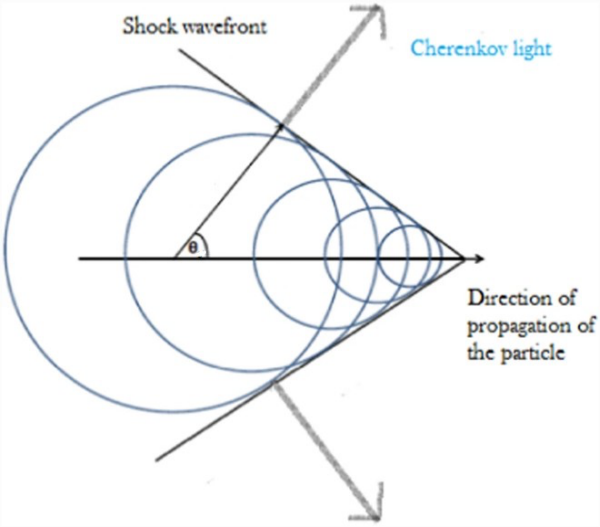


Figure 1. Cherenkov radiation is generated at an angle to the direction of the travelling particle, defined as θ, which is related to the energy of the particle. The diagram taken from (Das & Boruah, 2022) and adapted by the authors of this study.

The kinetic energy of the particles must exceed certain thresholds to produce CR. The required threshold energy can be obtained through the following equation, where m is the mass of the particle, c is the speed of light, and n is the refractive index (Tanha et al., 2015):

$$E_{min} = mc^2\left(\frac{1}{\sqrt{1-\frac{1}{n^2}}} - 1\right)$$

In addition, the intensity of Cherenkov radiation is related to the velocity and kinetic energy of the particles and the r efractive index of the environment. Figure 2 shows the inverse relationship between the energy threshold and the index of refraction. It is known through experimental trials that most biological tissues require at least 219 keV of energy when assuming a refractive index of 1.4.

Vol. 2023 (2) 214 – 225
ISSN 2688-3651 [online]

The intensity of produced photons increases with the refractive index. Since different tissues have different refractive indices as shown in Figure 3, it is possible to observe different intensities of CR and differentiate between healthy and tumor tissues.

The charged particles originate from radioisotopes with unstable atoms which decay in order to achieve stability. During this decay there is a release of radiation in the form of energy and a particle: alpha (α), beta (β) or gamma (γ). Although theoretically all charged particles can produce CR as long as they exceed the threshold, radioisotopes that can produce useful CR in biomedicine are limited to beta emitters (Mc Larney et al., 2021). Gamma particles cannot produce CR due to low energy of secondary electrons from Compton scattering and tissue interference. Additionally, alpha particles used in medicine cannot exceed the energy threshold required to produce CR. Fortunately, many beta emitter radionuclides currently used in the industry contain greater energy than threshold, and thus can produce Cherenkov photons in water and tissue.

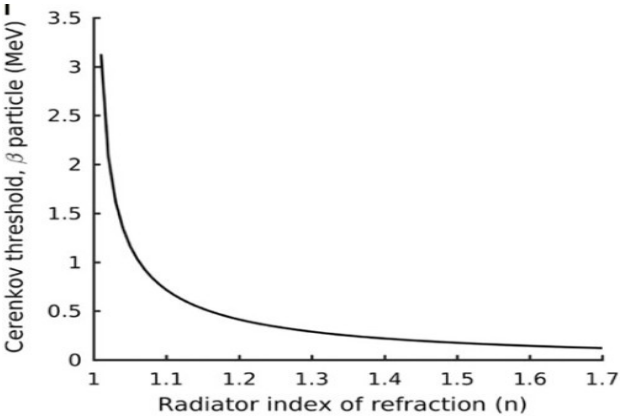


Figure 2. Cherenkov energy threshold for β particles to produce CR vs the refractive index of the environment. This graph assumes a β particle with a kinetic energy of 0.511 MeV (McLarney et al., 2021).

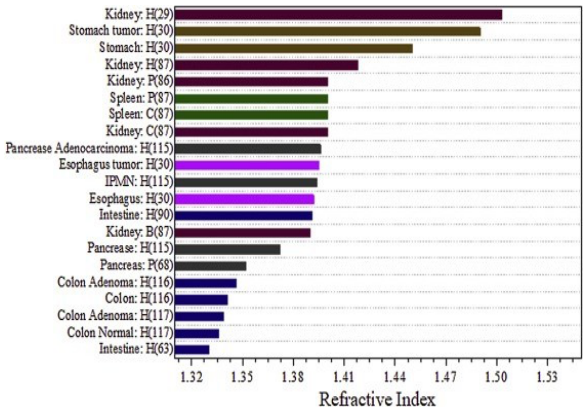


Figure 3. Refractive index values for miscellaneous tissues, including cartilage, heart, lungs, thyroid, breast, and neck. Note differences between malignant and benign breast, thyroid tissues (Khan et al., 2021).

In addition, high energy x-rays emitted by linear accelerators (LINACs) are also able to produce CR (Bianfei et al., 2022). This is relevant as LINACs are commonly used during clinical radiotherapy. During the external beam radiation therapy (EBRT), Compton scattering occurs where x-ray photons transfer energy to electrons, causing some to be ejected from their orbitals. If the ejected electron travels faster than the velocity of light in the tissue, CR will be generated (Axelsson et al., 2011). This property gave the potential for Cherenkov to be used to monitor EBRT in real time. Similarly, high-energy electron beams, heavy ion and proton radiation are recorded to produce CR during cancer therapy, opening the potential for CR to be further utilized in other high energy sources (Andreozzi, Brůža, et al., 2018; Masuda et al., 2018).

3. Cherenkov luminescence imaging (CLI)

3.1 Clinical CLI imaging utilizing Radioisotope emission

The current industry defines Cherenkov luminescence imaging (CLI) as any imaging technology that utilizes the Cherenkov radiation. The most comparable modality to CLI among widely used imaging techniques is PET. PET is able to provide images that are highly sensitive and its readings encompass the whole body. However, it is limited by poor spatial-temporal resolution, and high cost. In comparison, CLI products can achieve higher resolution at superficial depths, utilize non-radioactive dyes and are significantly cheaper and faster to produce. However, images of deeper tissues suffer with poorer resolutions. Since luminescence via CR is mostly in the ultraviolet and blue visible blue range, the light is easily absorbed and scattered by nearby tissues (Cao et al., 2015; Ciarrocchi & Belcari, 2017).

Vol. 2023 (2) 214 – 225
ISSN 2688-3651 [online]

Therefore, although CR is thought to be useful, the current technology is limited by the type and condition of the tissue assessed.

Robertson et al. (2009), were the first to produce CR bioimages after injecting mice with 2-deoxy-2-[18F]fluoro-d-glucose (^{18}F-FDG). Highly metabolic cells such as cancer cells increase their uptake of ^{18}F-FDG as it is an analogue of glucose. ^{18}F-mediated CLI has been further applied in various studies. Notably CLI was employed to define glioblastoma tumor margins (Lewis et al., 2018). Glioblastoma is a brain tumor in which the majority of the patients require brain surgery. For the operation to be successful, the tumor border where excision occurs must include all malignant cells. Rats implanted with human glioblastoma cells were injected with either the ^{18}F-FDG probe or the control fluorescent probe. The study by Lewis et al. (2018) has found that CLI performed better than the control in determining the accurate tumor boundary.

The feasibility, safety, and performances of ^{18}F-FDG probes in determining tumor margins were also evaluated in breast-conserving surgeries (Grootendorst et al., 2017). Patients with grade 3, estrogen receptor–negative/Her-2–negative tumors all received wide excision with sentinel lymph node biopsy or axillary lymph node dissection. CLI assessed tumor margins in 10 of the 12 patients with a significant correlation found between CLI and margin histopathology. Furthermore, the procedure was performed safely through low radiation exposures to clinical staff (Grootendorst et al., 2017). ^{18}F-FDG was also used for the first clinical application of endoscopic CLI. The CCD attached to endoscopes were used to identify cancerous gastrointestinal lesions, where imaged patients received diagnostic doses of ^{18}F-FDG. CL emitted by the imaging agent showed good correlation with clinical whole-body PET imaging and allowed for the quantification and differentiation of tumors from healthy tissue (Hu et al., 2015). ^{18}F has been similarly utilized to study tumor margins of lymphoma, liver cancer, and prostate cancer (Ciarrocchi et al., 2021; Costa et al., 2022; Ritter et al., 2021).

Spinelli et al. were the first to capture CR images of a human living tissue by using ^{131}I (Spinelli et al., 2013). Thyroid specific uptake of iodine has made ^{131}I an attractive radioisotope for the image. For other tumor cells, other radioisotopes are preferred. ^{68}Ga has become especially popular due to its favorable properties, such as a short half-life of 67.7 min and 89% beta emission, and relatively low threshold compared to other radioisotopes (Naji & AL-Nahhas, 2012). A major advantage of ^{68}Ga radioisotope is the ability of the metallic ion to chelate to various proteins and antibodies. Antibodies that target specific tumor markers can be designed, allowing for greater specificity. For example, a Swedish group developed a novel PET/magnetic resonance/CR triple-modality imaging agent, ^{68}Ga-SPIONs, by labeling superparamagnetic iron oxide nanoparticles with ^{68}Ga (Madru et al., 2013). ^{68}Ga-SPIONs demonstrated accurate and sensitive Cherenkov luminescence imaging of sentinel lymph nodes (SLN) - lymph nodes near cancerous tissue. ^{68}Ga-SPION shows potential for Cherenkov luminescence to be used alongside conventional PET and MRI. In addition, the novel imaging probe ^{68}Ga-DOTAKEK-(GX1)2 was applied for PET/CR dual-modality imaging, enabling simultaneous PET and Cherenkov diagnosis for gastric cancer (Yin et al., 2021). Furthermore, a recent study reported that [^{68}Ga]Ga-PSMA-11 CR imaging could correctly identify tumor margin status in 83% of 15 prostate cancer patients, suggesting that CR imaging is a promising technique for obtaining rapid pathology results in image-guided prostate cancer surgery (J. O. Heuvel et al., 2022). Specificity and sensitivity of ^{68}Ga-chelated radioisotopes show promise and are anticipated to enter further clinical trials. In addition to ^{18}F and ^{68}Ga, other radioisotopes such as ^{47}Sc, ^{64}CU, ^{89}Zr, ^{90}Y and ^{124}I are being studied for CR mediated diagnosis and bioimaging, especially in the scope of dosimetry and monitoring radiotherapym (Bianfei et al., 2022).

In general, CRs emitted by radioisotopes are low in intensity, and most rely on charged coupled devices (CCD) or PET scanners to obtain discernable image. CCDs are popular photodetectors commonly used for digital imaging and video since the 1980s. They are designed to detect low intensity photons, through many photoactive regions made of metal-oxide-semiconductor (MOS) capacitors (Lesser, 2014). Once a capacitor interacts with photons, a surface charge is generated for that capacitor. By calculating the specific capacitors activated, and the surface charges generated, CCDs are able to translate the positional and intensity values of the photon to an image (Hui, 2020; Lesser, 2014). However, even with CCDs, the luminescence produced by the CR is dimmer than the ambient room light, and therefore requires absolute darkness to be detected. This absolute dark environmental condition is a significant disadvantage against widely utilizing CR during clinical practice.

Vol. 2023 (2) 214 – 225
ISSN 2688-3651 [online]

3.2 Cherenkov Luminescence Tomography (CLT)

CLI is a method of two-dimensional (2D) planar imaging, which cannot provide the three-dimensional (3D) spatial distribution of the radionuclide probes (Hui, 2020; Wang et al., 2021). This problem can be solved by its 3D derivatives, such as CLT (Z. Hu et al., 2010). CLT reconstructs the spatial distribution of the internal radionuclide probes by integrating the CL images measured from the body surface with the structural information and other prior information. Li et al. first proposed the concept of CLT, reconstructing the 3D distribution of ^{18}FDG in a homogeneous mouse model (Li et al., 2010). The homogenous model is relatively simple, and is quite different from the real imaging organisms, resulting in inaccurate results. This can be solved by a more complex heterogeneous model with more defined data points as shown in Figure 4 (H. Guo et al., 2017).

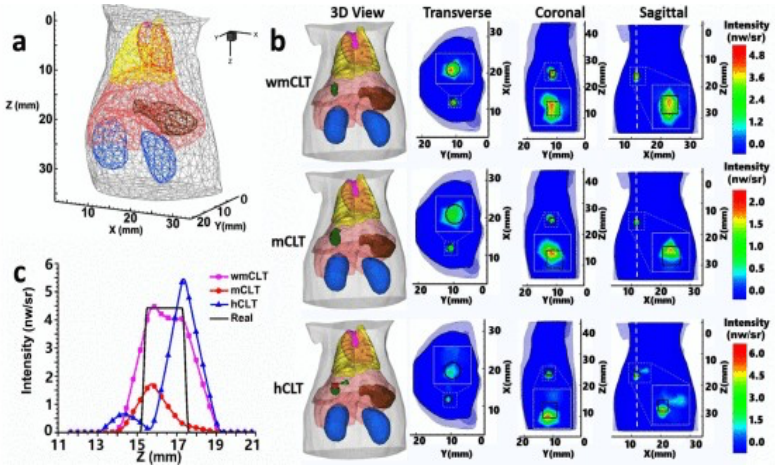


Figure 4. The multispectral representation of a xenograph tumor of various regions, a reconstruction that was possible due to weight mediated data representation. (H. Guo et al., 2017)

However, increasing the amount of measured data has the disadvantages of increasing the reconstruction time and limiting the reconstruction efficiency. Using a small amount of data to obtain accurate reconstruction results is an important problem to be addressed (Wang et al., 2021).

3.3 Cherenkov Radiation mediated Dosimetry

CR can be used for monitoring the dose and dose distribution of radiotherapy to prevent radiotherapy-related adverse events. Patients whose radiotherapies approach the eyes often experience flashes of blue and white light during the treatment (Bianfei et al., 2022). In 2008, Newman et al. concluded that the source of this light is the high-energy x-ray radiation that results in the production of CR inside the eyes (Newman et al., 2008). Since this observation, CR induced by external beam radiation therapy (EBRT) is regarded as a possible treatment and monitoring modality. EBRT delivers radiation to tumors using a linear accelerator (LINACs). The production of CR from LINACs is the same as radioisotope decay although for LINACs electrons pass through the tissue at two to three orders of magnitude higher kinetic energy. In addition, it is possible to increase photon beam energy at will, and apply beam hardening filtration for greater depth penetration than radioisotopes. These advantages strengthen LINAC's candidacy for optical molecular imaging with CR.

In 2011, Axelsson et al. captured CR images during EBRT in tissue-mimicking media and noted that the intensity of CR increased with increasing radiation energy in the same medium (Axelsson et al., 2011). Furthermore, the study has found that CR could excite the photosensitizer protoporphyrin IX, and stated the possibility of concurrent monitoring of EBRT by CR induced molecular fluorescence. Another study has utilized luminescent probe platinum oxyphor G4 (PtG4) to measure tumor oxygenation in vivo by Cherenkov-excited luminescence scanned imaging. The produced image showed submillimeter resolution and nanomolar sensitivity (Pogue et al., 2018). The use of PtG4 also allowed for the distribution values to be translated via Monte Carlo simulations to show three dimensional dose distributions. (Glaser et al., 2013). Subsequent studies have further demonstrated that dosimetry monitoring in real time for EBRT with both MV electron and x-ray beams might be possible for superficial tissues (Decker et al., 2021) Zhang et al used a time-gated intensified CCD (ICCD) that was coupled with a commercial lens to access canine oral

Vol. 2023 (2) 214 – 225
ISSN 2688-3651 [online]

tumor (Zhang et al., 2012). The subsequent images showed that under irradiation, the intensity of Cherenkov emission is directly proportional to radiation dose (Figure 5).

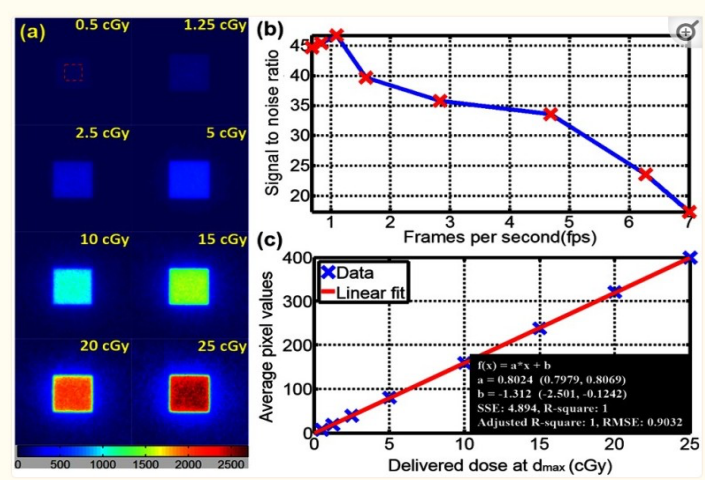


Figure 5. Images that compare the delivered dose compared with light intensity. Average pixel values and delivered dose at dmax shows direct relationship. Image obtained from (Zhang et al., 2013)

Since the use of CR produced by EBRT was established as feasible, further studies were conducted to further study the relationship between CR and radiation dose, dose depth and dose rate for accurate translation (Helo et al., 2014). Experimentally it was confirmed as the previous study that CR was linear with dose and independent of dose rate. More importantly, Cherenkov beam direction and transverse profiles showed great potential to be used to check the range and field width constancy of electron beams. This showed the potential use of CR as a monitor during radiotherapy for precise targeting of electron beams. The correlation between the diffusion of CR and its optical signal was further proven, introducing a more accurate method for verifying dose delivery in real time (Decker et al., 2021). CR imaging exhibited accurate intensity matching, suggesting that CR is suitable for quality assessment of radiation therapy that relies heavily on precise intensity (Andreozzi, Mooney, et al., 2018; Black et al., 2019; Miao et al., 2019).

The first clinical application of CR imaging in EBRT was performed on a breast cancer patient. Jarvis et al. acquired a CR video during whole-breast radiation and demonstrated the correlation between the intensity of CR and the superficial dose (Jarvis et al., 2014). Hachadorian et al. continued the whole breast imaging and improved the accuracy of CR images by reducing the noise of ambient light (Hachadorian et al., 2018). Currently CR imaging is being studied as part of the protocol to reduce variability and errors in whole-breast radiation therapy, minimizing the risk of recurrence and healthy tissue toxicity caused by inaccurate coverage (Hachadorian et al., 2021). In addition to conventional EBRT, MRI guided radiotherapy, bremsstrahlung x-rays and positron emitters were explored for CR use, with positive results (Darr et al., 2020; Jarvis et al., 2014).

Table 1. Summary of human trials involving Cherenkov Radiation

Delivery	Source	Application	Highlight	Reference
^{131}I	^{131}I	Thyroid cancer	First CR tumor diagnostic image	Spinelli et al., 2013
^{68}Ga-PSMA	^{68}Ga	Prostate cancer (after tumor cell as excised)	Tumor cells were successfully detected on the incised prostate CLI images as confirmed by histopathology.	Darr et al., 2020
^{68}Ga-PSMA	^{68}Ga	Prostate cancer, in vivo	First to demonstrate ability to distinguish between a positive and negative surgical margin, imaging within 45 min and low radiation exposure to staff	Heuvel et al., 2020
[^{68}Ga]Ga-PSMA-11	^{68}Ga	Prostate cancer	Imaging can correctly identify margin status, including close margins, in 83% of cases	Heuvel et al., 2020
LINAC	x-ray	Breast cancer	The first clinical application of CR imaging in EBRT	Zhang et al., 2013
LINAC	x-ray	Breast cancer	Whole breast imaging with more reduced visual noise.	Hachadorian et al., 2018
LINAC	x-ray	Breast cancer	Use of CR as a coupled monitoring system for accuracy of radiation beam	Hachadorian et al., 2021

Vol. 2023 (2) 214 – 225
ISSN 2688-3651 [online]

4. Photodynamic Therapy (PDT)

Photodynamic therapy (PDT) is a minimally invasive therapeutic modality that is primarily used to treat cancer (Brown et al., 2004). PDT uses two entities, a photosensitizer (PS) drug and electromagnetic radiation. PS drug's specific activation only in the presence of radiation allows for high precision (Ellis-Davies, 2007; Mayer & Heckel, 2006). Once activated PS drugs can generate cytotoxic reactive oxygen species (ROS) such as singlet oxygen (1O2) to destroy cancer cells (Clement et al., 2017). However, the majority of photoresponsive compounds react most effectively to irradiation with light showing wavelengths < 500 nm, typically showing a limiting penetration depth of only micrometers into biological tissue. Moreover, highly energetic UV light can cause harm to cells, possibly causing apoptosis or DNA damage (D'Orazio et al., 2013; Matsumura & Ananthaswamy, 2004). Cherenkov radiation (CR), which exhibits an emission spectrum in the near-UV range (250–600 nm) carries the advantage of covering the wavelength range of most photoresponsive compounds currently known (Elrick & Parker, 1968; Jelley, 1955).

Ran et al. investigated the possibility of activation based on CR generated by ^{18}F (Ran et al., 2012). Other studies have attempted to introduce novel ways to produce CR with greater penetrance and photoactivation of the PS drugs (Bulin et al., 2013; Ran et al., 2012). One study utilized ^{18}F but as a PET radiotracer, which was able to activate luciferin, a possible breast cancer treatment (Krebs et al., 2021). By demonstrating the activation and subsequent detection, the study has suggested that CR induced by therapeutic radiopharmaceuticals such as ^{90}Y could offer a possibility for a synergistic combination of radiotherapy with targeted photopharmacology in future applications. In addition, some studies employed ^{137}Cesium and ^{90}Y as a possible source for therapeutic CR (Y. Guo et al., 2020; Hartl et al., 2016). Most studies have suggested that PDT induced by Cerenkov radiation has deeper tissue penetration than traditional PDT. However, the strategy of coupling radionuclides with photosensitizers may cause severe side effects. In response, Qian et al, designed a new modality of PS drug delivery via nanovesicles, for additional protection from inaccurate activation CR due to tissue scattering (Qian et al., 2022).

5. Challenges and Future Prospects

Cherenkov Radiation is limited by half-life, faint intensity, and penetrance. As the CR process is dependent on emissions of the charged particles during decay, the speed of decay is an important factor to consider (Mc Larney et al., 2021). For example, half-life of ^{18}F (110 minutes) is not suitable for longer experimental investigations that might require days. Therefore, the choice of radioisotope to be used for clinical studies and therapy may be limited by the individual half-life. To compensate for limitation of radioisotope choice, more materials are being tested for potential use. CR is also limited by low intensity of the radiation - in the nanoWatt to picoWatt per cm2 range (Tanha et al., 2015). This characteristic is especially challenging during radioisotope-mediated CLI. The faint luminescence requires the environment to be completely dark, reducing the practicality (Bianfei et al., 2022; Tanha et al., 2015). This limitation has driven the effort to develop more sensitive CCDs that can differentiate ambient light, as well as Monte Carlo modeling to estimate minimum potential dose necessary for CLI, within the safety range (Shu et al., 2018; Tian et al., 2022). The dependence of dark has also driven exploration of CLI use in laparoscopy or endoscopy, which does not require light. Additionally, CR generated by energy modalities such as EBRT and other high energy yields does not require light control as greater kinetic energy generated results in intensity near six-fold of emitting radioisotopes (Tanha et al., 2015). Finally, the spectrum produced by CR is continuous and ranges from the UV to NIR wavelength, and its intensity is related to wavelength by $1/\lambda^2$. The most intensive CL is in the spectrum of the UV/blue band, which is easily scattered and absorbed by biological tissues (Ma et al., 2014). Therefore, only a small portion of CL emitted from the radionuclide in vivo can be detected. This resulted in most of the clinical trials currently limited to surface tumors, such as thyroid, breast, and laparoscopically, prostate and gastric cancers. However, penetrance is shown to be greater for EBRT, and further research will aim towards evaluating new probes that produce more penetrating light.

6. Conclusion

Since its first use in biomedical sciences, Cherenkov radiation has emerged as an attractive new modality for safe

Vol. 2023 (2) 214 – 225
ISSN 2688-3651 [online]

and cost efficient bioimaging compared to conventional imaging technologies. While the first decade was limited mostly to animal models, an increasing number of studies have conducted human clinical trials to better marginalize the tumors for surgical excisions. There is a growing trend of evolving radioisotopes that can simultaneously be detected by PET, MRI and CR related CCD. Further, three-dimensional imagery of CR is being refined through a new modality known as Cherenkov luminescence tomography (CLT). In addition to vast research dedicated to diagnostics, photodynamic therapy utilizing CR is also receiving more attention, due to their higher penetrance compared to conventional therapies. Currently, all the advanced CR related applications in oncology are in the clinical trial stages, and more time is needed to see full clinical potential. Developing tracers or probes that can produce longer lasting CR with greater intensities, and developing scanners that can detect faint signatures will further cement the place of CR in the growing field of cancer molecular imaging.

References

Andreozzi, J. M., et al. (2018). Improving treatment geometries in total skin electron therapy: Experimental investigation of linac angles and floor scatter dose contributions using Cherenkov imaging. *Medical Physics*, *45*(6), 2639–2646. https://doi.org/10.1002/mp.12917

Andreozzi, J. M., et al. (2018). Remote Cherenkov imaging-based quality assurance of a magnetic resonance image-guided radiotherapy system. *Medical Physics*, *45*(6), 2647–2659. https://doi.org/10.1002/mp.12919

Axelsson, J., et al. (2011). Cerenkov emission induced by external beam radiation stimulates molecular fluorescence. *Medical Physics*, *38*(7), 4127–4132. https://doi.org/10.1118/1.3592646

Bianfei, S., et al. (2022). Application of Cherenkov radiation in tumor imaging and treatment. *Future Oncology*, *18*(27), 3101–3118. https://doi.org/10.2217/fon-2022-0022

Black, P. J., et al. (2019). An investigation of clinical treatment field delivery verification using cherenkov imaging: IMRT positioning shifts and field matching. *Medical Physics*, *46*(1), 302–317. https://doi.org/10.1002/mp.13250

Brown, S. B., Brown, E. A., & Walker, I. (2004). The present and future role of photodynamic therapy in cancer treatment. *The Lancet. Oncology*, *5*(8), 497–508. https://doi.org/10.1016/S1470-2045(04)01529-3

Bulin, A.-L., et al. (2013). X-ray-Induced Singlet Oxygen Activation with Nanoscintillator-Coupled Porphyrins. *The Journal of Physical Chemistry C*, *117*(41), 21583–21589. https://doi.org/10.1021/jp4077189

Cao, X., et al. (2015). Intensity Enhanced Cerenkov Luminescence Imaging Using Terbium-Doped Gd2O2S Microparticles. *ACS Applied Materials & Interfaces*, *7*(22), 11775–11782. https://doi.org/10.1021/acsami.5b00432

Chen, Z., Lu, W., Garcia-Prieto, C., & Huang, P. (2007). The Warburg effect and its cancer therapeutic implications. *Journal of Bioenergetics and Biomembranes*, *39*(3), 267–274. https://doi.org/10.1007/s10863-007-9086-x

Ciarrocchi, E., & Belcari, N. (2017). Cerenkov luminescence imaging: Physics principles and potential applications in biomedical sciences. *EJNMMI Physics*, *4*(1), 14. https://doi.org/10.1186/s40658-017-0181-8

Ciarrocchi, E., et al. (2021). In-vitro Study for Cerenkov Luminescence Imaging Application in Hepatic Metastasectomy. *Frontiers in Physics*, *9*. https://www.frontiersin.org/articles/10.3389/fphy.2021.635540

Clement, S., et al. (2017). Nanoparticle-mediated singlet oxygen generation from photosensitizers. *Journal of Photochemistry and Photobiology A: Chemistry*, *332*, 66–71. https://doi.org/10.1016/j.jphotochem.2016.08.009

Costa, P. F., et al. (2022). F18-PSMA Cerenkov luminescence and flexible autoradiography Imaging in a prostate cancer mouse model and first results of a radical prostatectomy feasibility study in men. *Journal of Nuclear Medicine*. https://doi.org/10.2967/jnumed.122.264670

Vol. 2023 (2) 214 – 225
ISSN 2688-3651 [online]

Darr, C., et al. (2020). Intraoperative 68Ga-PSMA Cerenkov Luminescence Imaging for Surgical Margins in Radical Prostatectomy: A Feasibility Study. *Journal of Nuclear Medicine*, *61*(10), 1500–1506. https://doi.org/10.2967/jnumed.119.240424

Das, P. R., & Boruah, K. (2022). Gamma hadron separation method in ground-based gamma ray astronomy using simulated data. *Indian Journal of Physics*. https://doi.org/10.1007/s12648-022-02423-y

Decker, S. M., et al. (2021). Estimation of diffuse Cherenkov optical emission from external beam radiation build-up in tissue. *Journal of Biomedical Optics*, *26*(9), 098003. https://doi.org/10.1117/1.JBO.26.9.098003

D'Orazio, J., et al. (2013). UV Radiation and the Skin. *International Journal of Molecular Sciences*, *14*(6), Article https://doi.org/10.3390/ijms140612222

Ellis-Davies, G. C. R. (2007). Caged compounds: Photorelease technology for control of cellular chemistry and physiology. *Nature Methods*, *4*(8), Article 8. https://doi.org/10.1038/nmeth1072

Elrick, R. H., & Parker, R. P. (1968). The use of Cerenkov radiation in the measurement of β-emitting radionuclides. *The International Journal of Applied Radiation and Isotopes*, *19*(3), 263–271. https://doi.org/10.1016/0020-708X(68)90023-9

Glaser, A. K., et al. (2013). Projection imaging of photon beams by the Čerenkov effect. *Medical Physics*, *40*(1), 012101. https://doi.org/10.1118/1.4770286

Grootendorst, M. R., et al. (2017). Intraoperative Assessment of Tumor Resection Margins in Breast-Conserving Surgery Using 18F-FDG Cerenkov Luminescence Imaging: A First-in-Human Feasibility Study. *Journal of Nuclear Medicine*, *58*(6), 891–898. https://doi.org/10.2967/jnumed.116.181032

Guo, H., et al. (2017). Weight Multispectral Reconstruction Strategy for Enhanced Reconstruction Accuracy and Stability With Cerenkov Luminescence Tomography. *IEEE Transactions on Medical Imaging*, *36*(6), 1337–1346. https://doi.org/10.1109/TMI.2017.2658661

Guo, Y., et al. (2020). Photodynamic Therapy Excited by Cerenkov Radiation from Cesium-137 Irradiator: In Vitro Studies. *Clinical Oncology and Research*, *2020*(6), 1–5. https://doi.org/10.31487/j.COR.2020.06.07

Hachadorian, R., et al. (2018). Correcting Cherenkov light attenuation in tissue using spatial frequency domain imaging for quantitative surface dosimetry during whole breast radiation therapy. *Journal of Biomedical Optics*, *24*(7), 071609. https://doi.org/10.1117/1.JBO.24.7.071609

Hachadorian, R., et al. (2021). Verification of field match lines in whole breast radiation therapy using Cherenkov imaging. *Radiotherapy and Oncology*, *160*, 90–96. https://doi.org/10.1016/j.radonc.2021.04.013

Hartl, B. A., et al. (2016). Activating Photodynamic Therapy in vitro with Cerenkov Radiation Generated from Yttrium-90. *Journal of Environmental Pathology, Toxicology and Oncology : Official Organ of the International Society for Environmental Toxicology and Cancer*, *35*(2), 185–192. https://doi.org/10.1615/JEnvironPatholToxicolOncol.2016016903

Helo, Y., et al. (2014). Imaging Cerenkov emission as a quality assurance tool in electron radiotherapy. *Physics in Medicine & Biology*, *59*(8), 1963. https://doi.org/10.1088/0031-9155/59/8/1963

Hernot, S., et al. (2019). Latest developments in molecular tracers for fluorescence image-guided cancer surgery. *The Lancet Oncology*, *20*(7), e354–e367. https://doi.org/10.1016/S1470-2045(19)30317-1

Heuvel, J. O., et al. (2022). Cerenkov Luminescence Imaging in Prostate Cancer: Not the Only Light That Shines. *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine*, *63*(1), 29–35. https://doi.org/10.2967/jnumed.120.260034

Vol. 2023 (2) 214 – 225
ISSN 2688-3651 [online]

Hu, H., et al. (2015). Feasibility study of novel endoscopic Cerenkov luminescence imaging system in detecting and quantifying gastrointestinal disease: First human results. *European Radiology*, *25*(6), 1814–1822. https://doi.org/10.1007/s00330-014-3574-2

Hu, Z., et al. (2010). Experimental Cerenkov luminescence tomography of the mouse model with SPECT imaging validation. *Optics Express*, *18*(24), 24441–24450. https://doi.org/10.1364/OE.18.024441

Hui, R. (2020). Chapter 4—Photodetectors. In R. Hui (Ed.), *Introduction to Fiber-Optic Communications* (pp. 125–154). Academic Press. https://doi.org/10.1016/B978-0-12-805345-4.00004-4

Jadvar, H. (2016). Is There Use for FDG-PET in Prostate Cancer? *Seminars in Nuclear Medicine*, *46*(6), 502–506. https://doi.org/10.1053/j.semnuclmed.2016.07.004

Jarvis, L. A., et al. (2014). Cherenkov Video Imaging Allows for the First Visualization of Radiation Therapy in Real Time. *International Journal of Radiation Oncology*Biology*Physics*, *89*(3), 615–622. https://doi.org/10.1016/j.ijrobp.2014.01.046

Jelley, J. V. (1955). Cerenkov radiation and its applications. *British Journal of Applied Physics*, *6*(7), 227. https://doi.org/10.1088/0508-3443/6/7/301

Khan, R., et al. (2021). Refractive index of biological tissues: Review, measurement techniques, and applications. *Photodiagnosis and Photodynamic Therapy*, *33*, 102192. https://doi.org/10.1016/j.pdpdt.2021.102192

Krebs, M., et al. (2021). Photopharmacological Applications for Cherenkov Radiation Generated by Clinically Used Radionuclides. *International Journal of Molecular Sciences*, *22*(16), Article 16. https://doi.org/10.3390/ijms22169010

Lesser, M. (2014). 3—Charge coupled device (CCD) image sensors. In D. Durini (Ed.), *High Performance Silicon Imaging* (pp. 78–97). Woodhead Publishing. https://doi.org/10.1533/9780857097521.1.78

Lewis, D. Y., et al. (2018). [18F]fluoroethyltyrosine-induced Cerenkov Luminescence Improves Image-Guided Surgical Resection of Glioma. *Theranostics*, *8*(14), 3991–4002. https://doi.org/10.7150/thno.23709

Li, C., Mitchell, G. S., & Cherry, S. R. (2010). Cerenkov Luminescence Tomography for Small Animal Imaging. *Optics Letters*, *35*(7), 1109–1111.

Ma, X., Wang, J., & Cheng, Z. (2014). Cerenkov radiation: A multi-functional approach for biological sciences. *Frontiers in Physics*, *2*. https://www.frontiersin.org/articles/10.3389/fphy.2014.00004

Madru, R., et al.. (2013). 68Ga-labeled superparamagnetic iron oxide nanoparticles (SPIONs) for multi-modality PET/MR/Cherenkov luminescence imaging of sentinel lymph nodes. *American Journal of Nuclear Medicine and Molecular Imaging*, *4*(1), 60–69.

Masuda, T., et al. (2018). Measurement of nuclear reaction cross sections by using Cherenkov radiation toward high-precision proton therapy. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018- 20906-z

Matsumura, Y., & Ananthaswamy, H. N. (2004). Toxic effects of ultraviolet radiation on the skin. *Toxicology and Applied Pharmacology*, *195*(3), 298–308. https://doi.org/10.1016/j.taap.2003.08.019

Mayer, G., & Heckel, A. (2006). Biologically Active Molecules with a "Light Switch." *Angewandte Chemie International Edition*, *45*(30), 4900–4921. https://doi.org/10.1002/anie.200600387

Mc Larney, B., Skubal, M., & Grimm, J. (2021). A Review of Recent and Emerging Approaches for the Clinical Application of Cerenkov Luminescence Imaging. *Frontiers in Physics*, *9*. https://www.frontiersin.org/articles/10.3389/fphy.2021.684196

Vol. 2023 (2) 214 – 225
ISSN 2688-3651 [online]

Miao, T., et al. (2019). Cherenkov imaging for linac beam shape analysis as a remote electronic quality assessment verification tool. *Medical Physics*, *46*(2), 811–821. https://doi.org/10.1002/mp.13303

Naji, M., & AL-Nahhas, A. (2012). ^{68}Ga-labelled peptides in the management of neuroectodermal tumours. *European Journal of Nuclear Medicine and Molecular Imaging*, *39 Suppl 1*, S61-67. https://doi.org/10.1007/s00259-011-1990-y

Newman, F., et al. (2008). Visual sensations during megavoltage radiotherapy to the orbit attributable to Cherenkov radiation. *Medical Physics*, *35*(1), 77–80. https://doi.org/10.1118/1.2815358

Heuvel, J., et al. (2020). 68Ga-PSMA Cerenkov luminescence imaging in primary prostate cancer: First-in-man series. *European Journal of Nuclear Medicine and Molecular Imaging*, *47*(11), 2624–2632. https://doi.org/10.1007/s00259-020-04783-1

Pogue, B. W., et al. (2018). Maps of in vivo oxygen pressure with submillimetre resolution and nanomolar sensitivity enabled by Cherenkov-excited luminescence scanned imaging. *Nature Biomedical Engineering*, *2*(4), Article 4. https://doi.org/10.1038/s41551-018-0220-3

Qian, R., et al. (2022). Minimizing adverse effects of Cerenkov radiation induced photodynamic therapy with transformable photosensitizer-loaded nanovesicles. *Journal of Nanobiotechnology*, *20*(1), 203. https://doi.org/10.1186/s12951-022-01401-0

Ran, C., et al. (2012). In Vivo Photoactivation Without "Light": Use of Cherenkov Radiation to Overcome the Penetration Limit of Light. *Molecular Imaging and Biology*, *14*(2), 156–162. https://doi.org/10.1007/s11307-011-0489-z

Ritter, Z., et al. (2021). In situ lymphoma imaging in a spontaneous mouse model using the Cerenkov Luminescence of F-18 and Ga-67 isotopes. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-03505- 3

Robertson, R., et al. (2009). Optical imaging of Cerenkov light generation from positron-emitting radiotracers. *Physics in Medicine and Biology*, *54*(16), N355-365. https://doi.org/10.1088/0031-9155/54/16/N01

Rowe, S. P., & Pomper, M. G. (2022). Molecular imaging in oncology: Current impact and future directions. *CA: A Cancer Journal for Clinicians*, *72*(4), 333–352. https://doi.org/10.3322/caac.21713

Shu, D.-Y., et al. (2018). Analysis on the emission and potential application of Cherenkov radiation in boron neutron capture therapy: A Monte Carlo simulation study. *Applied Radiation and Isotopes*, *137*, 219–224. https://doi.org/10.1016/j.apradiso.2018.04.012

Spinelli, A. E., et al. (2013). First human Cerenkography. *Journal of Biomedical Optics*, *18*(2), 020502. https://doi.org/10.1117/1.JBO.18.2.020502

Tanha, K., Pashazadeh, A. M., & Pogue, B. W. (2015). Review of biomedical Čerenkov luminescence imaging applications. *Biomedical Optics Express*, *6*(8), 3053–3065. https://doi.org/10.1364/BOE.6.003053

Tian, F., et al. (2022). Proton range monitoring based on picosecond detection using a Cherenkov radiation detector: A Monte Carlo study. *Applied Radiation and Isotopes*, *180*, 110055. https://doi.org/10.1016/j.apradiso.2021.110055

Wang, L., He, X., & Yu, J. (2021). Prior Compensation Algorithm for Cerenkov Luminescence Tomography From Single-View Measurements. *Frontiers in Oncology*, *11*. https://www.frontiersin.org/articles/10.3389/fonc.2021.749889

Wilson, M. P., et al. (2020). Diagnostic accuracy of 99mTc-sestamibi SPECT/CT for detecting renal oncocytomas and other benign renal lesions: A systematic review and meta-analysis. *Abdominal Radiology*, *45*(8), 2532–2541. https://doi.org/10.1007/s00261-020-02469-8

Vol. 2023 (2) 214 – 225
ISSN 2688-3651 [online]

Yin, J., et al. (2021). 68Ga-Labeled GX1 Dimer: A Novel Probe for PET/Cerenkov Imaging Targeting Gastric Cancer. *Frontiers in Oncology*, *11*. https://www.frontiersin.org/articles/10.3389/fonc.2021.750376

Zhang, R., et al. (2013). Real-time in vivo Cherenkoscopy imaging during external beam radiation therapy. *Journal of Biomedical Optics*, *18*(11), 110504. https://doi.org/10.1117/1.JBO.18.11.110504

Zhang, R., et al. (2012). Čerenkov radiation emission and excited luminescence (CREL) sensitivity during external beam radiation therapy: Monte Carlo and tissue oxygenation phantom studies. *Biomedical Optics Express*, *3*(10), 2381–2394. https://doi.org/10.1364/BOE.3.002381

Journal of Research High School

Vol. 2023 (2) 226 – 234
ISSN 2688-3651 [online]

Predicting Election Outcomes from Facial Images of Candidates Using an Unbiased Machine Learning Model

Raymond (Deming) Lin,[1] * Jadelyn Tran[1]

[1]Webber Academy, Calgary, AB, Canada
*Corresponding Author: raymond.lin.336@gmail.com

Advisor: Tyler Giallanza, tylerg@princeton.edu

Received January 5, 2023; Revised July 4, 2023; Accepted, July 24, 2023

Abstract

In the past, studies on psychology have shown that humans can create instantaneous judgments of a stranger's personality and characteristics, just based on a picture of their face. The most successful study reached a 72.4% accuracy in predicting election outcomes. There had been machine learning studies that tried to replicate this success, but some kind of human input and therefore bias were often present. This project aimed to create a bias-free and independent machine learning model that only uses the image of political candidates to predict their success. With no other information than a candidate's face, the model achieved a 70.43% accuracy predicting election results. Not only did the different approaches in this experiment give a quantitative way to compare different types of human thinking, but it can also be used as a benchmark for future research that further investigates the relationship between facial traits, human judgments, and machine learning.

Keywords: Election prediction, Machine learning, Classifier, Transfer learning, Facial images, Physiognomy

1. Introduction

Once every few years, the topic of elections floods the media. Typically, voters always try to judge candidates for who they are and how their platform would benefit them. However, no matter how hard individuals try to combat their varying levels of prejudice, electoral decisions are still heavily influenced by the appearance of the candidates.

Studies had shown that individuals can infer various traits about a stranger just from their face (Olivola & Todorov, 2010; Olivola & Todorov, 2009; Todorov, 2018), and a lot of that subconscious interpretation of information heavily affects electoral decision making (Ballew & Todorov, 2007). There had been numerous studies (Olivola & Todorov, 2010; Olivola & Todorov, 2009; Todorov, 2018) that tried to test how well personality and other traits can be detected by humans by just looking at faces. In addition to simplistic traits like attractiveness and emotions, short exposures to pictures of strangers can lead to predictions of more complex features such as level of social class (Todorov, 2018) and political competence (Ballew & Todorov, 2007).

For example, a study of facial judgements and election outcomes was conducted by Alexander Todorov and Charles C. Ballew (2007). They presented sixty-four test subjects with pairs of senators who were running against each other and asked the subjects to predict who was more likely to win the election. Test subjects were only allowed to judge on the candidates that they had no familiarity with, and they successfully predicted 72.4% of the Senate races in 2006. Interestingly, when Todorov and Ballew (2007) asked participants to "deliberate and make a good judgment", their predictive accuracy decreased. Participants were also shown to perform better when constrained to 250 milliseconds and thus forced to make a rapid, unreflective judgment. Not only was this study a demonstration of how much physical appearance influences electoral results, it also suggests that the more accurate judgments are often subconscious, as the more the participants consciously tried to analyze, the worse their results were.

Vol. 2023 (2) 226 – 234
ISSN 2688-3651 [online]

Inspired by the positive results of the human experiments, researchers tried to replicate these results by building machine learning models to predict election outcomes from images of faces (Joo et al., 2015; Todorov et al., 2005; Ventura et al., n.d.). Many of these models had near-human accuracy. For example, an extensive study (Joo et al., 2015) was able to reach a 67.9% accuracy in predicting the US governor races by combining classic computer vision methods with a Support Vector Machine (SVM). Specifically, the researchers trained the SVM on two types of data: personality/demographic characteristics, and physical attributes. To determine personality/demographic characteristics, the researchers asked human test subjects to evaluate the comparative characteristics of two candidates at a time (e.g., which candidate looks wealthier, which candidate looks older), and compiled those results into a series of perceived characteristics. To determine the physical attributes, they also split each of the training images into regions and extracted physical attributes, such as if they were smiling or if they were wearing glasses.

Despite the impressive accuracy of this method, some degree of further work was needed to fully explore this problem. For example, the researchers themselves named the categories of attributes that they want to extract, which could have led to certain physical features left ignored or under-represented. They also did not mention the logic behind determining the categories of perceived characteristics that they want to evaluate, and the logic of why those attributes were the most significant. Furthermore, a significant portion of the input data used for the training came from human participants to begin with; the model likely would not have achieved such a high accuracy without human contribution. In summary, previous experiments (Joo et al., 2015) likely relied on assumptions that could have introduced biases into how the algorithm could have made its prediction.

This project aimed to create a series of machine learning models that would provide a deeper understanding of both how artificial intelligence can be used to emulate subconscious human analysis during elections, and how AI can be used as an additional tool for election prediction and as a new perspective into subconscious human psychology. The following is the approach:

1. To create a new machine learning model that can generate a prediction of the relative electoral success of a politician by only using an image of their face. Unlike previous studies, this new model would use no human test subjects to provide the input data, and to not manually label certain features of the face for the model to focus on. By creating a machine learning model that does not rely on any form of human analysis, this could eliminate the bias that previous studies in this field could have introduced.
2. To create two different approaches to solve this problem, with both machine learning models having the same lack of reliance on human test subjects. The first approach purely uses the pixels of the images of candidates as the input to the machine learning model. The other approach simplifies the image down to facial characteristics perceived by a machine learning model, and to use that to generate a prediction. These two approaches could be interpreted as representations of two different ways of human thinking when asked to predict stranger's traits from their faces. These could act as a quantifiable comparison between human's conscious and subconscious judgments. Testing two different approaches would also provide an opportunity to try different methodologies to search for a higher accuracy.
3. To see how much facial appearance influenced election outcomes compared to other electoral factors that are commonly known to influence election results, such as the incumbency status and the amount of the campaign budget for the candidate. Models that predict electoral success of politicians based on these other factors will also be developed. By comparing those models to the model that analyzes facial appearance, it could lead to a better understanding of how much of a role does facial appearance of a politician plays compared to these other more well-established factors.

2. Materials and Methods

2.1 Problem Formulation

The aim of this project was to train a machine learning model to predict the outcome of an election from a facial image of a candidate. Most previous studies on this subject all paired politicians during their experimental or data collection process, training the model to predict which of the two candidates will win the election. This pairing was

Vol. 2023 (2) 226 – 234
ISSN 2688-3651 [online]

often either derived from the two candidates running against each other (Ballew & Todorov, 2007), or the two candidates with a similar perceived age (Joo et al., 2015). This logic might have seemed sound, as one would think that to evaluate the competence of a politician, there had to be a relative comparison. However, this new experiment was designed to challenge that preconception by not using any pairing in the process. All the data that were used were individually selected from an array and individually evaluated, with "winner" or "loser" as the only label. This dramatically reduce the amount of data needed to train the model due to the quadratic nature of pairwise predictions.

2.2 Data Collection

Political Candidates Dataset

Specifically for this experiment, a brand-new dataset for the facial images of politicians was created. We found public datasets with names of the winning and losing politicians from the US House of Representatives elections from 2000 to 2020 (total of 9 elections) and Canadian House of Commons elections from 2000 to 2021 (total of 8 elections). In total, there were around 18,000 candidates from the Canadian Parliament and 6,880 candidates from the US House of Representatives. This data formed a list of names for the politicians, and the information about what elections they won and lost also was used for labeling the dataset.

After the names of the candidates were found, BeautifulSoup was used to extract the HTML information from either google searches, specifically the knowledge panel, or from Wikipedia. 7068 facial images were found, one per candidate, per election.

Out of those images, an automatic filter via Deepface (Serengil, 2022) was ran to detect the face of the person in the images. This is important as Deepface is used later again to crop out parts of the image and only leave the face of the candidate. The images that did not have a detectable face were discarded, and 5,511 images remained. A round of manual filtering was also conducted, eliminating images that, for example, had unusual lightning or multiple faces; 4,501 images remained. For the candidates in this dataset, their images were standardized as much as possible by using an official campaign photo, where most candidates are smiling, facing the camera directly, and are in good lightning conditions. For many of the politicians that were removed, a different, usable photo were often manually found, increasing this dataset to a total of 4,859 images. Among these images, there were candidates that lost/won multiple times, and only one of those images were kept in the dataset. Candidates that have both won and lost had their images eliminated entirely.

The number of "winners" and "losers" had to be the same while training the program. During training, 2,018 unique images of candidates (1 image per candidate) were used, 1,009 for each category, with some "winner" images left unused (randomly selected during each train/test split). Using Deepface, these images were converted to grayscale and cropped to 48 by 48 pixels centered in their face, with the background and clothing cropped out.

Incumbency and Campaign Financing Data

Although the primary focus of this study is on facial images, this project also sought to explore how much incumbency and campaign financing information influences the predictions made by the models. BeautifulSoup was used to collect data from Ballotpedia, and using the information found by web scraping the HTML code, the information of 609 US candidates were recorded. The candidates that had this information were trained again (later fully explained under Models - Method 3), this time with incumbency and financial spending as additional input variables.

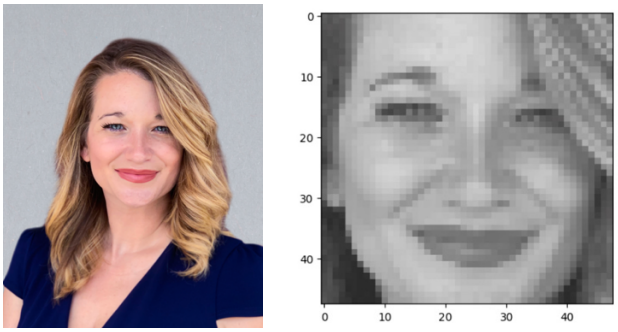


Figure 1a & 1b. This is an example of a candidate in the "Loser" category (Aliscia Andrews, Virginia 10th District). These are the original image that was collected and the processed image used for training, respectively.

Examples

Vol. 2023 (2) 226 – 234
ISSN 2688-3651 [online]

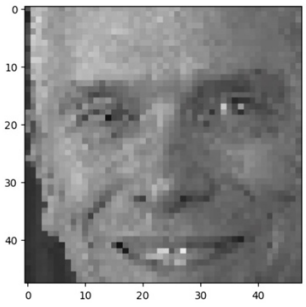


Figure 2a & 2b. This is an example of a candidate in the "Winner" category (Bill Johnson, Ohio 6th District). These are the original image that was collected and the processed image used for training, respectively.

2.3. Models

To fully understand this problem, four methods were designed. The method of input and the training process were the manipulated variables in order to obtain a more comprehensive suite of results. The four methods utilized a classifier for images, a classifier for images with an autoencoder, a logistic regression with Deepface.analyze (Serengil, 2022), and a classifier with Deepface.represent respectively. All coding was developed with Python and Google Colab ("Google", 2017). Specifically, PyTorch ("PyTorch", n.d.) was used for all of the programming with the machine learning models. Deepface (Serengil, 2022), a python library with extensive functions for facial recognition and image analysis, was used as well.

Method 1 – Pixel-Based Classifier

Using PyTorch ("PyTorch", n.d.), a neural network was developed. It had linear layers and it classified the given inputs into either the "winner" or the "loser" category. This classifier was trained only using normalized arrays of the 48x48 pixel, gray scale images, with everything but the politician's face cropped out of the photo. The model was trained with BCELoss as the loss function and a learning rate of 0.0001, until it reached a point of diminishing returns (around 750-1000 epochs).

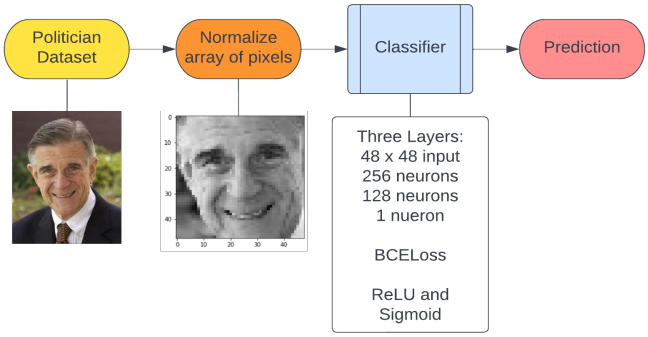


Figure 3. This is the general structure of Method 1. The classifier uses pixels of facial images to determine a result.

In the classifier, there are three linear layers in total. The first layer accepts 2,304 inputs (48 by 48 pixels), and has 256 outputs. The second layer decreases that to 128 outputs, and the third layer decreases it to 1 (a binary unit indicating if the candidate is a winner or a loser). ReLU was the activation function of the first two layers, and Sigmoid for the last layer.

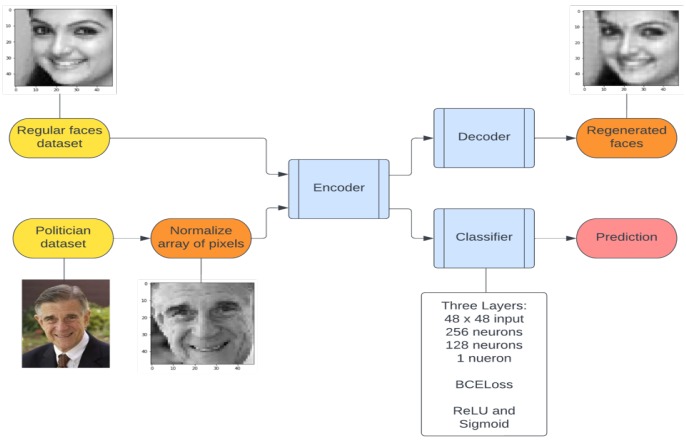


Figure 4. This is the general structure of Method 2a. The difference between Method 1 and 2a was that Method 2a used an autoencoder to implement transfer learning, and its encoder was used with the classifier to improve its feature extraction abilities.

Method 2a – Pixel-Based Classifier with Autoencoder

Due to the limited size of the training dataset, Method 2a sought to improve the performance of the model by implementing transfer learning to the classifier with the use of an autoencoder. First, an autoencoder was created to be trained on an independent

Vol. 2023 (2) 226 – 234
ISSN 2688-3651 [online]

dataset (Arora, 2020) of regular faces, with the autoencoder trained to compress and then reconstruct the facial images with as much detail as possible. By doing so, the autoencoder learned what part of a human face is essential for reconstructing the facial image, while discarding other irreverent details. Then, the autoencoder was used with the classifier in Method 1, to hopefully increase the efficiency of the model. This was because the autoencoder could assist the classifier to only focus on parts of the face that are of significance, narrowing down the facial information that influences election results.

This autoencoder is a Convolutional Neural Network that has three Conv2d layers for encoding, and three ConvTranpose2d layers for decoding. The activation function was three ReLU functions for encoding, and two ReLU and one Sigmoid function for decoding. The weights of the autoencoder were saved and then applied to the classifier from Method 1.

Method 2b – Pixel-Based Classifier with Deepface Representations

To implement another method of transfer learning, a pre-trained model was used to extract features from the candidates' images. The returned values of Deepface.represent (Serengil, 2022) are the input variable for this method. Deepface.represent was meant to be used for facial recognition by representing faces with vector embeddings. These embeddings were then used as the input for a classifier. This classifier is similar to the classifier used in Method 1 and 2a; the difference is that since the input variables were changed, the expected input of the first layer was changed as well.

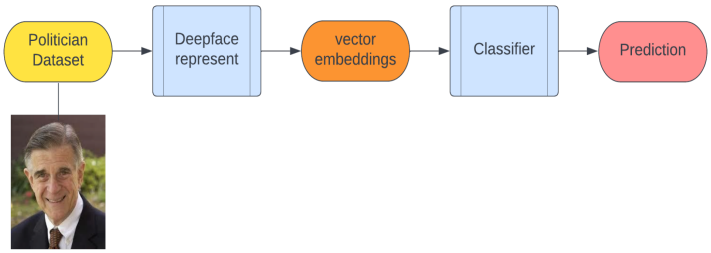


Figure 5. This is the general structure of Method 2b. Just like Method 2a, this was another implementation of transfer learning using Deepface represent.

Method 3 – Demographic-Informed Classifier

As a comparison to the prior 2 methods, which predicted election outcomes from facial images directly, for this method a model was trained on demographic/facial expression information. The Python library Deepface, and specifically the Deepface.analyze function was used, to generate these features. The function received the images of the politicians, and it returned a series of perceived values of the politicians, such as their age, composition of race, likelihood of certain emotions, gender, etc. These values were used as input variables for a Logistical Regression model from SkLearn ("Learn: Machine", n.d.). This model was also retrained with the incumbency status and the amount spent in campaign financing as additional input variables.

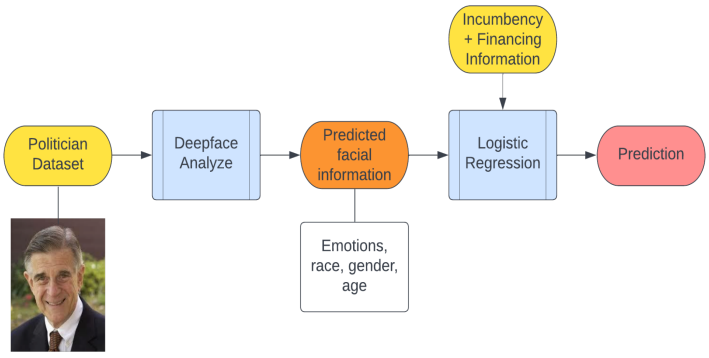


Figure 6: This is the general structure of Method 3. The perceived traits from Deepface.analyze were used as input for a logistic regression model that predicted a result

Deepface contained pre-trained models that can analyze attributes of facial images. Deepface.analyze is a function that can return multiple attributes from a facial image. All of the returned information were utilized, which included likelihood of certain emotions (angry, neutral, happy, etc.), likelihood of race (White, Latino, Asian…), age and gender.

Vol. 2023 (2) 226 – 234
ISSN 2688-3651 [online]

3. Results

Evaluations of the accuracy of the model with both the training and testing datasets were recorded. The Receiver Operating Characteristics curves (ROC curves) was plotted and the Area Under the Curve (AUC) was calculated. The following are the results.

Table 1: Accuracy and AUC value of respective methods

	Method 1	Method 2a	Method 2b	Method 3
Training Accuracy (%)	94.20	95.36	61.16	60.45
Testing Accuracy (%)	70.43	66.96	16.57	59.95
AUC	0.77	0.74	0.37	0.61

3.1 Results for Method 1, 2a, 2b

ROC curves are used here to display the performance of the models as it is a common way to evaluate a binary machine learning classifier. True positive rate (y-axis) is the proportion of positive cases that are correctly evaluated, while the false positive rate (FPR) is the proportion of positive cases that are incorrectly evaluated. The ROC curve is used for its ability to display the effectiveness of the model at a range of thresholds, or cutoff values that determines if the output of the machine learning model should be classified as positive or negative (in this case, Winner or Loser). The AUC value, the area under the blue line, is a simple numerical representation of the information of the ROC curve. It is a more wholistic evaluation compared to an accuracy percentage as it represents the effectiveness of the model in a variety of thresholds.

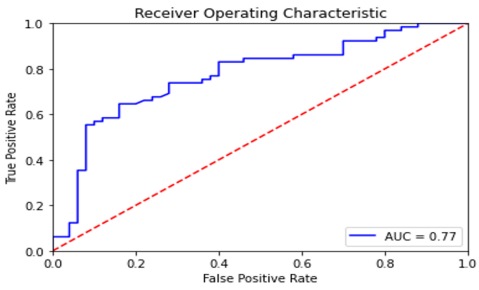
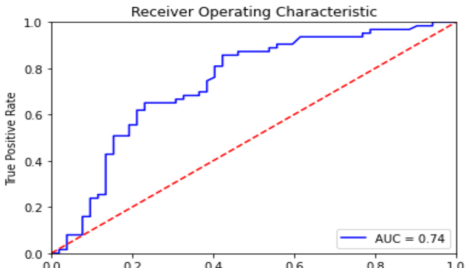


Figure 7: ROC graph for Method 1

The worst performance for a model would be the dotted red line, with an AUC value of 0.5, which would indicate that the model would have the same proportion of correct and incorrect predictions, or around just a 50% accuracy. The most ideal ROC curve would have the blue line closest to the top left of the graph, or an AUC value close to 1, indicating that at whatever threshold, the model would be able to maintain a high true positive rate, or a high rate of a correct positive evaluation.



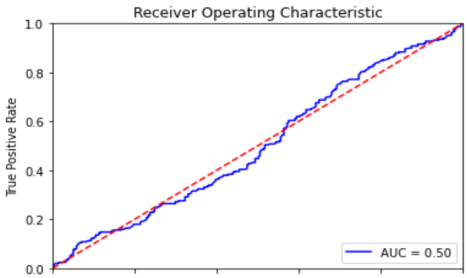


Figure 8a & 8b: ROC graph for Method 2a and Method 2b respectively

Table 2: Confusion Matrix for Method 1

	Actually Winners	Actually Losers
Predicted Winners	44	6
Predicted Losers	28	37

Table 3: Confusion Matrix for Method 2a

	Actually Winners	Actually Losers
Predicted Winners	36	16
Predicted Losers	22	41

From these results, it's clear that Method 1, with its 70.43% accuracy was the best approach. Method 2a was extremely close, with a 3.47% worse testing accuracy. It correctly predicted 4 candidates less than Method 1. However,

Vol. 2023 (2) 226 – 234
ISSN 2688-3651 [online]

Method 2a provided a better accuracy for predicting Losers. Method 2b, however, was very unsuccessful. It only achieved a 16% testing accuracy, which was worse than simply guessing with a 50% chance.

3.2 Results for Method 3

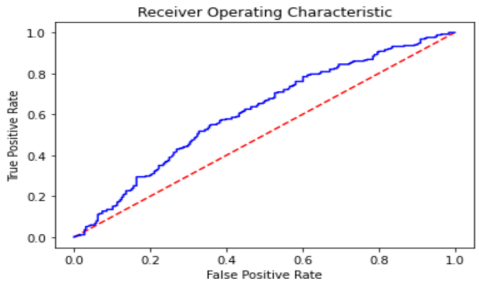


Figure 9: ROC curves for Method 3

Method 3 was relatively successful, achieving a 59.95% testing accuracy, with an AUC value of 0.61. It was not close to the success with Method 1, but its accuracy was significantly higher than 50%, the accuracy of making predictions with random guesses.

As stated before, Method 3 was retested by adding incumbency and amount spent on the campaign as additional parameters for the input of the linear regression model.

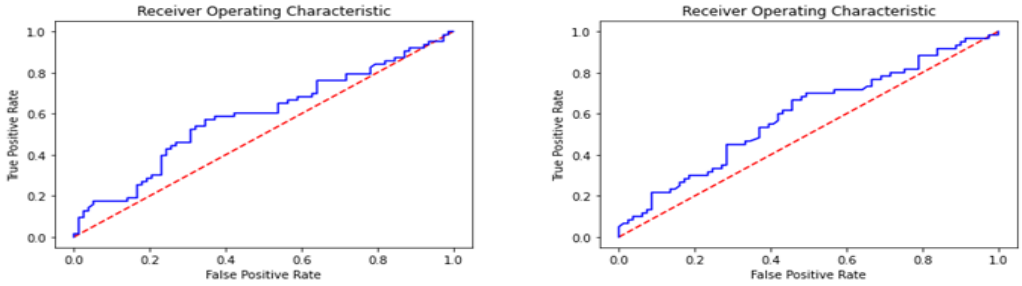


Figure 10A & 10B: ROC curves for variations of Method 3. These were results with the inclusion / exclusion of incumbency information. (330 politicians)

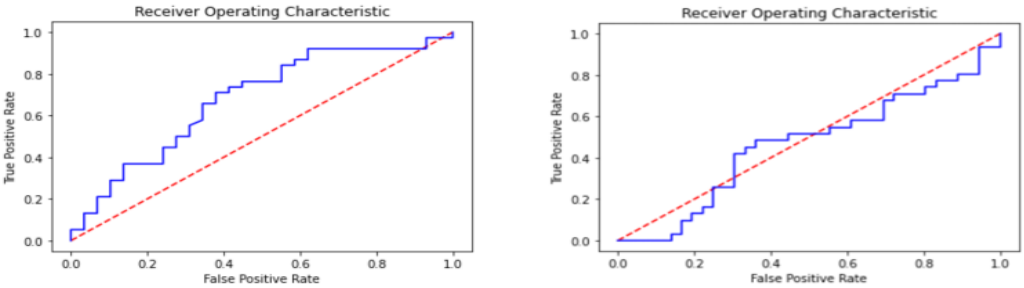


Figure 10C & 10D: ROC curves for variations of Method 3. These were results with the inclusion / exclusion of campaign financing information. (133 politicians)

The incumbency and financing information were only available for some candidates (330 for incumbency and 133 for financing). Figure 10B and 10D were baseline testing that had no information besides a candidate's face, while figure 10A and 10C were the results after incumbency / financial information is added to each group. Both with incumbency and financing, it demonstrated that the addition of this information greatly helped with the results in the respective groups.

The training with financing information as the only additional information resulted in the best ROC curve, with the highest AUC value out of the four (0.677). The training with incumbency information was close behind in those terms, but has a higher accuracy (highest of the four) of 58.79%.

4. Discussion

The model from Method 1, while only using the pixels of the images of a candidate's face, was able to reach 70.43% accuracy. It had no information on the political affiliation of the candidate, and no information of the political

Vol. 2023 (2) 226 – 234
ISSN 2688-3651 [online]

inclination of the region they are running in. This model was able to accomplish this without any other information, completely disregarding the competency, incumbency, or campaign spending of the candidate. In contrast to prior work (Joo et al., 2015), these new models had no human contribution as it only relied on an image of the politician.

These results are quite significant, as it suggests that artificial intelligence on its own, with no human influence, can emulate the appearance-based bias that effects election outcomes. Since this model had no other information to rely on, it is also a demonstration on how the facial appearance of a candidate is a heavy contributing factor in the minds of a voter. It is true that in real life, voters could be influenced by a variety of superficial aspects beyond appearance. Voters might also watch a video of a candidate or see them in person, which results in a lot more information on the candidate's appearance than just an image. However, by just using a picture, the machine learning model was still able to achieve a significantly greater-than-chance accuracy, and future improvements of this project that can analyze video samples, for example, might even further improve the results.

To make up for the relatively small size of the dataset, this project also designed two methods of transfer learning (Method 2a & 2b). They used generic data of faces train facial feature extraction, which then can be applied onto the election prediction model. Neither approach was more successful than Method 1 and Method 3. This suggests that the factors that are important for judging faces for elections are somewhat unique, and not the exact same as facial information that that are preserved in an autoencoder. A conclusion that could be reached from this is that on a dataset specifically targeted for a purpose, a large amount of data may not be necessary to achieve a significant result.

In previous human psychological studies on this topic, test subjects often only had a second or less to make a prediction. For most humans, that is not enough time to consciously analyze the candidate's faces and infer traits. Studies have shown that the shorter the time taken, the better the prediction (Ballew & Todorov, 2007). The explanation given was that humans' subconscious judgments or their "intuitions" are incredibly good at reflecting their internal judgments and biases, such as the preference when looking at an electoral candidate. In these cases, humans' subconscious made even better predictions than conscious analysis.

There is still debate about exactly how humans can do this, and the true nature of these judgments are yet to be fully understood. Are these judgments values that can be easily represented by numbers and categories of perceived emotions or traits, or are they reflections of nebulous, underlying factors that cannot be simply understood with a few words? The results from this project could provide a quantifiable way to compare these two possibilities.

Method 1 acts as a representation of human judgments that include underlying factors, while Method 3 only uses quantifiable and simplified traits to form a prediction. Method 3 reached 59.41% accuracy, showing that to a certain extent, simple and quantifiable traits such as emotions, race and age can still be incredibly useful information to analyze how voters would view a political candidate. However, Method 1 achieving the highest accuracy of 70.43% shows that there are factors unaccounted for in Method 3, and some useful traits cannot be successfully simplified or represented with understandable and basic categories.

In this project, Method 1 and Method 3 could be seen as representations of two ways of human thinking. Method 1 (by using all the pixels of a facial image) represents intuition: the subconscious, unreflective judgments that cannot be represented with a few words or values. Method 3 (by simplifying facial information into emotions, age, gender) represents consciousness: when humans intentionally try to categorize information into specific and familiar groups. This machine learning model may be a quantifiable assessment between the two methods of human judgment. With further integration into existing systems of political forecasting, the model's ability to assess voter bias can be crucial to improving election predictions. Furthermore, this project can be useful for any researchers that want to build upon these findings, use it as a benchmark, and further explore the nuances of human judgments and machine learning. Future studies based on different theories and principles can also use this model to compare with to determine which features are truly important for predicting election outcomes.

5. Conclusion

This study established a machine learning model that reached a 70.43% accuracy and an AUC value of 0.77 while predicting election outcomes by only analyzing the facial images of candidates. This was accomplished by gathering an unprecedented dataset with over two thousand facial images of political candidates, and creating four different

Vol. 2023 (2) 226 – 234
ISSN 2688-3651 [online]

methods that unlike previous studies, did not rely on the involvement of human test subjects. Not only did this project demonstrate how a politician's facial appearance has a consistent correlation to his or her political success, it also displayed how artificial intelligence can replicate voter's subconscious bias during elections, and therefore possibly other instantaneous and unreflective human judgments.

Acknowledgment

Majors thanks to Tyler Giallanza for his guidance and mentorship.

References

Arora, N. (2021). Age, Gender and Ethnicity (Face Data) csv. Kaggle.

Ballew, C. C., & Todorov, A. (2007, November 13). Predicting political elections from rapid and unreflective face judgments. Retrieved July 3, 2023, from https://www.pnas.org/doi/10.1073/pnas.0705435104

Google Colaboratory. (2017). computer software. Retrieved January 2, 2023.

Joo, J., Steen, F. F., & Zhu, S.-C. (2015, December). Automated Facial Trait Judgment and Election Outcome Prediction: Social Dimensions of Face. International Conference on Computer Vision.

Olivola, C. Y., & Todorov, A. (2010, January 23). Elected in 100 milliseconds: Appearance-Based Trait Inferences and Voting.

Olivola, C., & Todorov, A. (2009, May 5). The Look of a Winner. Retrieved January 2, 2023, from https://www.scientificamerican.com/article/the-look-of-a-winner/

PyTorch. (n.d.). Version (1.13.1). *From Research To Production*. Retrieved January 2, 2023, from https://pytorch.org/.

scikit-learn. (n.d.). Version (1.2.0). *scikit-learn Machine Learning in Python*. Retrieved January 2, 2023, from https://scikit-learn.org/stable/.

Serengil, S. I. (2022, May 10). DeepfaceVersion (0.0.75). *deepface 0.0.75*. PyPI. Retrieved January 2, 2023, from https://pypi.org/project/deepface/.

Todorov, A. (2018, May 14). *Can we read a person's character from facial images?* Scientific American Blog Network. Retrieved January 2, 2023, from https://blogs.scientificamerican.com/observations/can-we-read-a-persons-character-from-facial-images/

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005, June 10). Inferences of Competence from Faces Predict Election Outcomes. Washington D.C.; American Association for the Advancement of Science.

Ventura, C., Masip, D., & Lapedriza, A. (n.d.). Interpreting CNN Models for Apparent Personality Trait Regression. Institute of Electrical and Electronics Engineers.

Journal of Research High School

Vol. 2023 (2) 235 – 246
ISSN 2688-3651 [online]

An In Vitro Study on the Phytotoxic Effects of Glyphosate on *Spinacia oleracea*

Khaled Khraisheh[1] *

[1]Asamiah International School, Amman, Jordan
*Corresponding Author: khaled.khraisheh@gmail.com

Advisor: Nadia Jaradat, nadia.jaradat@gmail.com

Received January 8, 2023; Revised July 17, 2023; Accepted, August 14, 2023

Abstract

Glyphosate is the most common non-selective herbicide to date. Evidence shows increasing glyphosate misuse and mishandling, prompting noticeable glyphosate residues on many non-targeted plants. This study investigates, in vitro, the phytotoxicity of 0.00 % (control), 0.05 %, 0.10 % and 0.20 % glyphosate v/v% concentrations on *Spinacia oleracea*'s photosynthetic rate (1/min) and total chlorophyll content (µg chlorophyll/g tissue) physico-chemical biomarkers. A large leaf tissue was subjected to each glyphosate concentration for 2.0 min, where 300 mg samples were taken pre- and post-glyphosate treatment for chlorophyll extraction and spectrophotometric quantification. Moreover, excess treated tissue were made down to 15 small identical disks for the indirect measurement of the photosynthetic rate via photosyn- thetic flotation technique. Our results indicate immediate adverse phytotoxic effects on both photosyn- thetic biomarkers. Short-term exposure to the lowest glyphosate concentration (0.05 %) had considerably decreased the photosynthetic rate from the control, reaching to the maximum of 71.4 % for the highest glyphosate concentration (0.20 %). Each glyphosate concentration resulted in a statistically significan ($p < 0.01$) lower total chlorophyll content than their initial values. All absolute total chlorophyll content percentage changes for all glyphosate concentrations were significantly different ($p < 0.05$) from the con- trol, yet 0.05 % and 0.10 % glyphosate concentrations' results show no significance in their difference at $\alpha = 0.05$. Observations revealed the possible 5-enolpyruvylshikimate-3-phosphate synthase competitive inhibition and the decrease in magnesium ion bioavailability for chlorophyll biosynthesis, calling for more caution when handling this chemical.

Keywords: Glyphosate, Spinacia oleracea, Competitive Inhibition, Photosynthesis, Chlorophyll, Chelating Agent, Shikimic Acid

1. Introduction

Glyphosate is a well-known non-selective herbicide that is effective at "managing incisive and noxious weeds" and used in various settings including "agricultural, residential and commercial settings" (Gomes et al., 2014; US EPA, 2014). However, in the hands of the uninformed, it may harm humans, plants, food and the environment in general if it is overused. Moreover, the main method of applying this chemical is by spraying (US EPA, 2014), which increases the chances of its transferal to non-targeted plants making them susceptible to its phytotoxic effects by the wind and/or water contamination (Gomes et al., 2014). With these different possible hazardous outcomes, this study examines the in vitro effects of increasing glyphosate v/v% concentration (0.00 %, 0.05 %, 0.10 % and 0.20 %) on the photosynthetic rate (# of floating disks/min) and Total Chlorophyll Content (*TCC*) (µg chlorophyll/g tissue) of *Spinacia oleracea* after a brief 2.0 min exposure time.

Photosynthesis is an important biochemical process that most plants undergo. In it, plants convert inorganic substances (water and carbon dioxide) to organic ones (glucose) through a series of steps; some require the presence of light (light-dependent reactions) while others do not (light-independent reactions) ("Overview of Photosynthesis – Biology 2e", n.d.). It might seem like both reactions occur individually; however, they are highly dependent on one another. For example, the Calvin cycle - a cycle of light-independent reactions that occurs in the stroma of the chloroplast - requires Adenosine Triphosphate (ATP) and Reduced Nicotinamide Adenine Dinucleotide Phosphate

Vol. 2023 (2) 235 – 246
ISSN 2688-3651 [online]

(NADPH) to reduce 3-phosphoglyceric acid (3-PGA) into Glyceraldehyde-3-Phosphate (G3P) and to regenerate Ribulose-1,5-Bisphosphate (RuBP) in order to have a continuous cycle yielding glucose (Molnar & Gair, 2015) - a great energy source.

In order to attain these molecules (ATP and NADPH), light-dependent reactions must occur (summarized in Figure 1) (Arnon, 1971). This takes place in the thylakoid membrane where:

(a) A Photosynthetically Active Radiation (PAR) (between 400 nm and 700 nm) strikes a pigment molecule (e.g., chlorophylls) in Photosystem II (PSII) exciting its electron and initiating a series of electron movements until it donates to the primary electron acceptor, Pheophytin (Pheo), in its P680 reaction center (Allakhverdiev et al., 2010). To replace this missing electron, an electron is taken from a water molecule as it breaks down to two hydrogen cations, two electrons and one oxygen atom by photolysis as seen below (Arnon, 1971). This whole process is called photoactivation.

$$H_2O \xrightarrow{h\nu} 2H^+ + 2e^- + \frac{1}{2}O_2$$

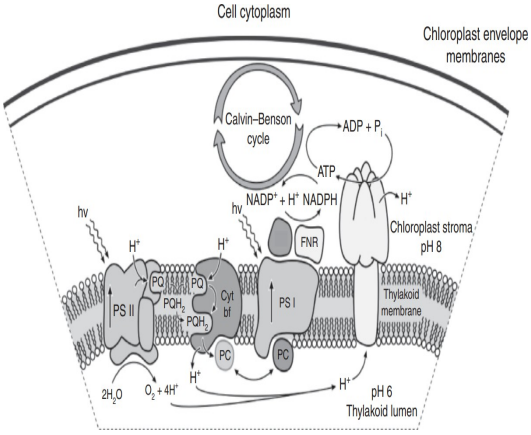


Figure 1. A diagram showing the cross section of a chloroplast and a thylakoid membrane adopted from (Masojídek et al., 2013). It also shows the steps that occur in the light-dependent reaction including photoactivation, photophosphorylation and the reduction of $NADP^+$.

(b) Pheo is then oxidized as its electron flows through multiple membrane proteins making up Electron-Transport-Chain (ETC) that actively pumps H^+ into thylakoid space with the energy released by each successive movement of electron through them. This creates a higher $[H^+]$ in the thylakoid space than in the stroma, causing diffusion of them, down the electrochemical gradient, through ATP synthase, and generating energy by chemiosmosis to phosphorylate Adenosine Diphosphate (ADP) to ATP. This whole process is called photophosphorylation.

(c) The electron from PSII will act as a source for the missing electron in Photosystem I (PSI); as the electron of its pigment molecules will also get excited and ultimately donated to its primary electron acceptor, chlorophyll A0 ("Light-dependent reactions (photosynthesis reaction) (article)", n.d.). A0 will then get oxidized and its electron will either flow back through the first ETC (if light is not the limiting factor) in a process called cyclic photophosphorylation (Arnon, 1971), or donated to the second ETC that results in reduction of Nicotinamide Adenine Dinucleotide Phosphate ($NADP^+$) to NADPH by the addition of two electrons and a H^+ by the last enzyme, Ferredoxin-$NADP^+$ reductase (FNR) (Arnon, 1971; Mulo, 2011).

As stated, glyphosate (N-(phosphonomethyl)glycine) is a non-selective herbicide that seems to indirectly effect the light-dependent reaction outlined earlier (Gomes et al., 2014). It has been long believed that the main reason behind plant death after being exposed to glyphosate is due to its inhibitory effect on 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase (Gomes et al., 2014). EPSP synthase is an important enzyme in the shikimate pathway that is responsible for catalyzing the conversion of Shikimate-3-Phosphate (S3P) to the sixth intermediate, EPSP, by addition of Phosphoenolpyruvate (PEP) to its molecule and the release of Inorganic phosphate (P_i) as seen in Figure 2 (de Souza & Sant'Anna, 2008). However, some scholars believe that their death is for the secondary indirect negative effects it has on several important components of the light-dependent reaction (Gomes et al., 2014). For example, the result of inhibiting EPSP synthase is the decrease in the biosynthesis of quinones (Gomes et al., 2014) - a set of e^-/H^+ carriers (El-Najjar et al., 2011; Nowicka & Kruk, 2010). This is because isoprenoid quinones are biosynthesized through metabolic pathways that originally start with aromatic amino acids that are the end products of the shikimate pathway (Nowicka & Kruk, 2010; Tzin et al., 2012). Thus, the biosynthesis of Plastoquinone (PQ) - an important electron carrier between PSII and Cytochrome-b_6f-complex (Cytb_6f) in the light-dependent reaction derived from

Vol. 2023 (2) 235 – 246
ISSN 2688-3651 [online]

tyrosine (Nowicka & Kruk, 2010) - will be inhibited, causing electrons to accumulate, and thus, decreases the need of electrons from the photolysis of water, and in general, decreases the photosynthetic rate of the plant.

Not only does this competitive inhibitor (competes with PEP for EPSP synthase's active site) limit the pro- duction of PQs, it also negatively affect chlorophyll biosynthesis (Gomes et al., 2014). Chlorophyll, as mentioned earlier, plays an important role in the first stages of the light dependent reaction. It consists of a porphyrin ring with magnesium as the central metal atom of its structure (Sapkota, 2020). This central metal atom is largely responsible for giving chlorophyll its color and role in photosynthesis (Farhat et al., 2016).

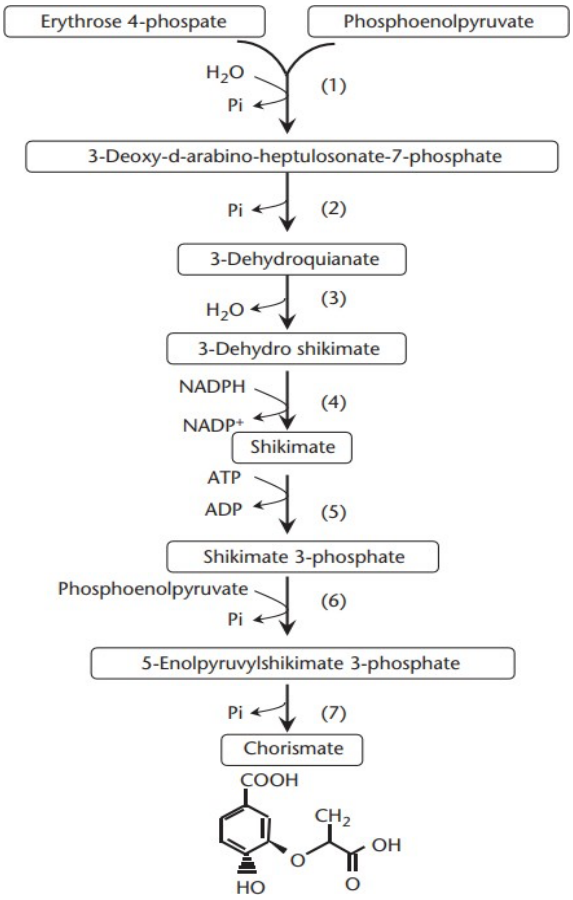


Figure 2. A flowchart adopted from (Tzin et al., 2012) that shows all seven intermediates of the shikimate metabolic pathway. EPSP synthase is responsible for the sixth (6) intermediate.

Magnesium ion is added to Protoporphyrin IX (PPIX) by the magnesium-chelatase in the early stages of chlorophyll biosynthesis (Figure 3) (Farhat et al., 2016), however, without magnesium, this obviously will not occur. Glyphosate is known as a "potent chelator for minerals" and magnesium is one of those minerals that it binds to ("Glyphosate, a chelating agent—relevant for ecological risk assessment?", n.d.). Therefore, exposing plants to glyphosate may result in magnesium deficiency, inhibiting the biosynthesis of chlorophyll and leading to lower photosynthetic rates.

As a result, it is hypothesized that as Glyphosate Concentration ([GP]) increases, *Spinacia oleracea*'s photosynthetic rate (# of floating disks/min) and *TCC* (µg chlorophyll/g tissue) will decrease due to (a) the in-crease probability of glyphosate binding to EPSP synthase instead of PEP, which ultimately leads to the decrease in the availability of aromatic amino acids needed for the biosynthesis of PQs that are responsible for the movement of electrons in the light-dependent reaction; (b) the increase probability of forming magnesium-glyphosate chelates that decreases the availability of magnesium needed for the biosynthesis of chlorophylls essential for photoactivation. To investigate these effects, the methodology of this experiment was based on photosynthetic flotation technique and spectrophotometry to measure the photosynthetic rate and TCC of the plant, Spinacia oleracea, respectively.

Photosynthetic flotation (or floating leaf disk) experiment is an indirect way of measuring the photosynthetic rate of mainly terrestrial plants such as *Spinacia oleracea*. It exploits the decrease in the buoyancy of the small leaf disks as oxygen generation from photosynthesis' light-dependent reaction (the photolysis of water) accumulates in the spaces found in its spongy mesophyll layer. These spaces were initially filled with a carbon-containing solution of a specific concentration by "force". Doing it by "force" involves creating a pressure difference "sucking" oxygen out of the leaf disks and then releasing the pressure to "push" the sodium bicarbonate ($NaHCO_3$) solution through its stomata and into the spongy mesophyll layer causing the decrease in the disk's buoyancy to the point of sinking in water. Thus, the higher the number of floating disks are at a given time, the faster the rate of photosynthesis it indirectly indicates as the oxygen replaces $NaHCO_3$ in these spaces.

Vol. 2023 (2) 235 – 246
ISSN 2688-3651 [online]

Spectrophotometry is a technique used to measure the light absorbance and/or transmittance through a solute containing liquid of unknown concentration ("Spectrophotometry", n.d.). This is done using a spectrophotometer that passes a beam of light of chosen wavelength through a liquid sample (held inside a cuvette) and onto a photocell (a light detector) ("Spectrophotometry", n.d.). Therefore, as solute concentration increases, more of the specific wavelength beam of light is absorbed giving us lower light transmittance value but higher absorbance value.

In this case, a spectrophotometer will be used on a chlorophyll-containing solution after extraction using 90 % acetone following a slightly modified (Su et al., 2010) procedure. Absorbance readings at 645 nm and 663 nm is needed to calculate the *TCC* (µg chlorophyll/g tissue) using ("3. METHODOLOGY", n.d.)'s equation (Equation 1).

$$TCC = \frac{V(20.2(A_{645})+8.02(A_{663}))}{M} \quad (1)$$

Where,
TCC : the Total Chlorophyll Content in µg chlorophyll/g tissue
V : final volume of the extract in mL
Ax : the absorbance value at Xnm
M : the mass of the leaf tissue in g

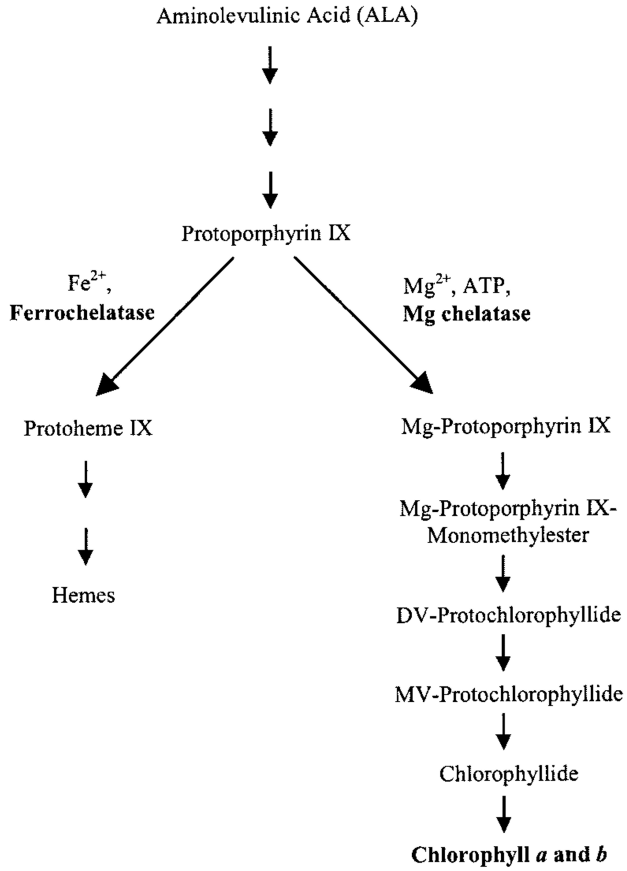


Figure 3. A flowchart adopted from (Rissler et al., 2002) that shows the metabolic pathway in which chlorophylls are synthesized.

2. Materials and Methods

2.1 Variables

Independent Variables

Glyphosate v/v% concentrations (0.00 %, 0.05 %, 0.10 % and 0.20 %) and the specific time intervals in which measurements were taken. For the photosynthetic flotation, counting the number of floating leaf disks were done after every 30.00 s for 56.0 min. For the *TCC*, sampling were taken before and after exposure to the [GP] in question for 2.0 min.

Dependent Variables

The photosynthetic rate (# of floating disks/min) of *Spinacia oleracea* measured by the rate in which the leaf disks float using the photosynthetic flotation technique and the *TCC* (µg chlorophyll/g tissue) measured spectrophotometrically.

Controlled Variables

The controlled variables of this experiment are listed in Table 1.

Vol. 2023 (2) 235 – 246
ISSN 2688-3651 [online]

Table 1. The controlled variables in different parts of the experiment (Photosynthetic Flotation (PF), Chlorophyll Quantification (CQ) and both), how they are controlled and why

Parts	Controlled Variables	How and Why Variable is Controlled
Both	Type of plants	Different types of plants have different concentrations of photosynthetic pigments, and thus, different photosynthetic rates. Therefore, Spinacia oleracea was only used throughout the experiment.
	Glyphosate exposure time	Giving a trial longer glyphosate exposure time than the others will damage its leaf more and ultimately give false results. Thus, all trials will have the leaves exposed to the glyphosate v/v% concentration in question for 2.0 min.
PF	Size and number of leaf disks	The larger the disk is, the more photosynthesising cells it has, the faster the photosynthetic rate. However, it also has a greater mass than the smaller disks so it may take longer to float. Furthermore, more disks will increase the dependence on the dissolved carbon in the solution. Thus, all disks are 5.0mm in diameter and limited to 15 in number.
	Volume and concentration of $NaHCO_3$	Dissolved $NaHCO_3$ will act as a source of carbon needed for photosynthesis. Since carbon dioxide is a limiting factor of photosynthesis, increasing or decreasing the volume and/or concentration of $NaHCO_3$ will effect the results until a certain point. Thus, excess 40.0mL of 0.1% (w/v%) $NaHCO_3$ is used for all trials to prevent it being a limiting reagent.
	Net pressure on leaf disks	The deeper the disks are, the higher the pressure (from the above liquid and the atmospheric pressure) acting upon the disks. This will decrease the size of the air bubbles making it harder for the disks to float. Thus, the volume of the liquid (0.1% (w/v%) $NaHCO_3$ in this case) is always 40.0mL and the beaker is the same for all trials.
	Light intensity	Light intensity is one of the limiting factors of photosynthesis which will increase or decrease its rate. Thus, the distance away from the light source (3.00 cm) and its power (23 W) are constant throughout the experiment.
	Total time of the experiment	Extending the total time of the experiment for one trial than the others will allow more disks to float giving false results. Thus, the total time of all trials is 56.0 min.
CQ	Mass of the leaf samples	Taking different masses of leaf samples for chlorophyll quantification will give false results if not accounted for. Thus, all leaf samples (both initial and final) in all trials are 300 mg in mass.
	Homogenization time and acetone concentration	Giving a trial less homogenization time may indicate lower total chlorophyll concentration. Moreover, the higher the concentration of acetone, the more chlorophyll is dissolved leading to higher results (Su et al., 2010). Thus, homogenization time and acetone concentration is fixed to 1.0 min and 90 %, respectively, for all trials.
	Final extract volume	Different final extract volume will give different absorbance values, and thus, false/unrepresentative results. Thus, the final extract volume was fixed to 10.0mL for all trials.
	Differential centrifugation speed and time	Centrifugation is needed to filter out organelles from the extract. This requires a specific speed and time dependent on that speed. Thus, the centrifugation speed and time were fixed to 3000 rpm and 10.0 min, respectively, for all trials.
	Absorbance wavelength	Equation 1 requires absorbance value of the extract at 645 nm and 663 nm. Changing the wavelength will give different values depending on chlorophyll's absorption spectrum. Thus, the wavelengths are fixed on the required ones for all trials.

Monitored Variables

The temperature in which the photosynthetic flotation experiment was taken place was monitored to be between 25.00 °C and 31.00 °C inclusive.

Vol. 2023 (2) 235 – 246
ISSN 2688-3651 [online]

2.2 Experimental Materials

Healthy *Spinacia oleracea* leaves where purchased from a local retailer the same day as the experiment was performed. A 23 W OSRAM DULUX Mini Twist was used as a light source and Eppendorf™Centrifuge Model 5702 was used for centrifugation. A 752n Lab Optical Instrument Single Beam UV Vis Spectrophotometer was the spectrophotometer of choice.

2.3 Photosynthetic Flotation and Sampling

300 mg of a large *Spinacia oleracea* leaf was taken as an initial chlorophyll sample and placed inside a small labeled dark container. The rest of the leaf was submerged under a 100.0 mL solution of the glyphosate v/v% concentration in question for 2.0 min. It was then dried and superficially washed with distilled water to remove any leftover solution on it. Another 300 mg of the leaf was removed as a final chlorophyll sample and placed in its designated container. Both 300 mg samples are needed for chlorophyll extraction and quantification (subsection 2.4).

What was left from the leaf was cut, with veins avoided, into 15 small circular disks of diameter 5.0 mm. Disks were then placed inside a 25.0 mL needle-less syringe containing 10.0 mL 0.1 % (w/v%) $NaHCO_3$ solution. The syringe was vigorously shaken and a combination of pulling and pushing the plunger were used while the opening tip was closed. Occasionally, the tip was opened to release the trapped air that had been forced out of the leaf tissue. This process ended once the were-floating 15 disks all sank. The whole solution with the 15 sunken disks were placed inside a small beaker and made up to 40.0 mL using the same $NaHCO_3$ solution used earlier.

The beaker was positioned 3.00 cm away from a 23 W light source and measurements started immediately. Counting the number of floating leaves was done after every 30.00 s. The temperature was monitored every 2.0 min. The solution was mixed gently after every 10.0 min to move any stuck/adhered disks. All measurements where conducted for 56.0 min. This was repeated four other times to have a total of five trials at each glyphosate v/v% concentration.

2.4 Chlorophyll Extraction and Quantification

The initial 300 mg leaf sample was further cut down into smaller pieces. It was then homogenized for 1.0 min with10.0 mL of 90 % acetone using a mortar and pestle. The homogenate was filtered and poured into a labeled graduated centrifuge test tube. The homogenate was then made up to 10.0 mL using the same acetone used earlier. This was repeated on the other leaf sample to give us a total of two labeled centrifuge tubes. These were then centrifuged at 3000 rpm for 10.0 min and the supernatant of each were taken for absorbance measurement at 645 nm and 663 nm. All steps were superficially summarized in Figure 4 for *one* trial of *one* glyphosate increment.

2.5 Risk Assessment

Safety Issues

No plastic equipment must be used when dealing with an acetone containing liquid; as acetone (especially when concentrated) can "damage the plastic's surface, softening it, smearing it, or even dissolving the plastic [itself]" ("How Acetone Affects Certain Plastics", 2020). Moreover, acetone is a "flammable liquid [and vapour], [serious] eye irritant" and "may cause drowsiness or dizziness" upon inhalation ("SAFETY DATA SHEET Acetone", 2015). Thus, the use of acetone must be in a fully ventilated room away from sparks or any heat emitting instruments. Furthermore, protective clothing (e.g., masks, lab coats, etc.) must be worn.

Similarly, proper protective eye-wear, clothing and gloves must be used when dealing with glyphosate ("MATERIAL SAFETY DATA SHEET Glyphosate 5.4", 2007). Glyphosate may cause "slight eye [and skin] irritations" which may worsen the more concentrated the product is. It is "harmful if inhaled" and slightly toxic "if small amount is swallowed" ("MATERIAL SAFETY DATA SHEET Glyphosate 5.4", 2007). Thus, a properly ventilated room is also advised ("MATERIAL SAFETY DATA SHEET Glyphosate 5.4", 2007).

Vol. 2023 (2) 235 – 246
ISSN 2688-3651 [online]

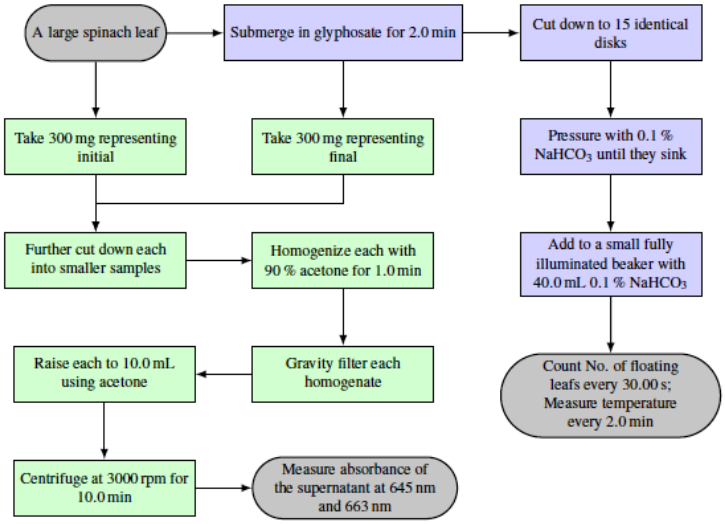


Figure 4. A flowchart outlining the key steps in the methodology used for one trial of one glyphosate increment. This includes both photosynthetic flotation and sampling (blue) as well as chlorophyll extraction and quantification (green).

Operating a centrifuge must be under the supervision or done by an expert. The centrifuge must be filled with an even number of centrifuge tubes all of similar mass and directly opposite each other. An unbalanced centrifuge at high speeds may lead to damages, injuries or even deaths. Thus, eye protection must be used when near a running centrifuge.

Environmental Issues

While acetone is "non toxic to aquatic organisms", "readily biodegradable in aerobic systems" and "poorly absorbed onto soils or sediments", it should "[not] enter drains and watercourses" ("SAFETY DATA SHEET Acetone", 2015). This also applies to glyphosate as it "[must] not contaminate water" as it "can result in oxygen depletion due to decomposition of dead plants" causing nearby fish to suffocate ("MATERIAL SAFETY DATA SHEET Glyphosate 5.4", 2007). Moreover, no excessive material consumption was used. The experiment was designed to minimize waste as much as possible while still allowing it to proceed.

2.6 Statistical Analysis

Where suitable, t-Test was conducted on mean experimental values at 95 % and 99 % confidence levels. Moreover, one-way ANOVA test was use with Tukey HSD post-hoc test at 95 % confidence level. H_0 = mean values of the variable(s) in question are *not* statistically significant at $\alpha = 0.05$ and/or $\alpha = 0.01$ where differences may have been the result of random chance; H_a = mean values of the variable(s) in question are statistically significant at $\alpha = 0.05$ and/or $\alpha = 0.01$ where differences may *not* be the cause of random chance.

Table 2. Paired two-tailed t-Test on the Initial and Final *TCC* of each glyphosate concentration (%)

Glyphosate Concentration (%)	*t*	*p*	$p < 0.05$?	$p < 0.01$?
0.00	2.90	0.0441	yes	no
0.05	17.3	6.53×10^{-5}	yes	yes
0.10	5.73	0.00460	yes	yes
0.20	16.7	7.55×10^{-5}	yes	yes

The statistical tests used are found in Table 2 and Table 3 below. Note that with $k = 4$ and $df_{within} = 16$, $Q_{0.05} = 4.046$ is the critical value at $\alpha = 0.05$ for Tukey HSD test. If $Q > Q_{0.05}$, then it is statistically significant at $\alpha = 0.05$. Similarly, $p < 0.05$ and $p < 0.01$ show significance at $\alpha = 0.05$ and $\alpha = 0.01$ respectively.

Table 3. Tukey HSD post-hoc test on the absolute % changes in *TCC*

A(%)	&	*B*(%)	*diff*	*n*(*A*)	*n*(*B*)	Standard Error	*Q*	$Q > Q_{0.05}$?
0.00	&	0.05	17.2	5	5	2.45	7.00	yes
0.00	&	0.10	26.3	5	5	2.45	10.7	yes
0.00	&	0.20	38.2	5	5	2.45	15.6	yes
0.05	&	0.10	9.12	5	5	2.45	3.71	no
0.05	&	0.20	21.0	5	5	2.45	8.55	yes
0.10	&	0.20	11.9	5	5	2.45	4.84	yes

Note—Done after insuring that H_0 is rejected in one-way ANOVA ($p = 6.9302 \times 10^{-8}$) at $\alpha = 0.05$; $Q_{0.05} = 4.046$; *A* and *B* are glyphosate concentrations.

Vol. 2023 (2) 235 – 246
ISSN 2688-3651 [online]

3. Results and Discussion

The results of this experiment suggests that glyphosate has a prominent effect on both the light-dependent reactions (oxygen gas generation) and *TCC* of the terrestrial plant *Spinacia oleracea*. These effects get more pronounced as the concentration of the substance increases.

Take for instance the mean number of floating leaf disks from the photosynthetic flotation experiment (Figure 5). In line with the hypothesis, the mean number of floating leaf disks in most time intervals are higher in that of the control than the ones treated with glyphosate. Increasing at an exponential rate, and on average, 92 % of the total 15 leaf disks were floating by the end of the experiment (56.0 min) in the control. On the other hand, exposure to as low as 0.05 % [GP] gave a value ((8.2 ±1.3)) surprisingly 41 % lower than the control ((13.8 ±1.1)) at that time interval. This decrease increased to 59 % for 0.10 % [GP] ((5.6 ±1.3)) and 72 % for 0.20 % [GP] ((3.8 ± 2.5)).

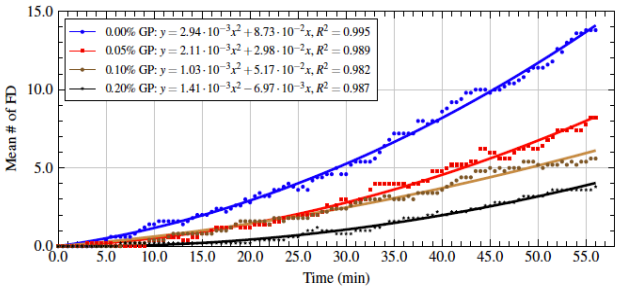


Figure 5. The average (n = 5) number of floating leaf disks (FD) after each 30.00 s time intervals. Polynomial functions were fitted to each glyphosate (GP) concentration. The standard deviations were omitted for better visualisation.

Moreover, the mean rate of change in number of floating leaf disks (a.k.a. the photosynthetic rate) between zero and 56.0 min show a decreasing trend with increasing [GP] (Table 4). This seems to be in an exponential decay fashion (R^2 = 0.987) with a clear indication of a plateau starting from, and unexpectedly speaking, 0.20 % [GP] (Figure 6). Yet, as expected, this concentration shows the largest percentage change (71 %) from the photosynthetic rate of the control (0.252/min) than the others have. However, the difference between adjacent rates is the highest (0.104/min) between 0.05 % [GP] and the control, indicating an immediate and strong adverse effects.

The general trend of the above results are as expected and demonstrated in several published studies. With exposure to glyphosate, the production of aromatic amino acids decreases as it inhibits EPSP synthase from catalysing the sixth intermediate of the shikimate pathway (de Souza & Sant'Anna, 2008; El-Najjar et al., 2011; Gomes et al., 2014; Nowicka & Kruk, 2010; Tzin et al., 2012). As a linear metabolic pathway, interrupting any of its intermediates would stop it completely (Tzin et al., 2012). Doing so decreases the biosynthesis of PQs needed for electron movement in the light-dependent reaction (El-Najjar et al., 2011; Nowicka & Kruk, 2010; Tzin et al., 2012). Causing electrons to accumulate, the photolysis of water decreases as indicated by the decrease in the rate of oxygen generation. Our results show a significant decrease in this area indirectly measured by the rate of floating disks after application with glyphosate. As this is a competitive inhibition, increasing [GP] increases the frequency of successful collisions with EPSP synthase's active site (Gomes et al., 2014; Nowicka & Kruk, 2010). This explains why our results show decreasing rates with increasing inhibitor concentration, consistent with present studies.

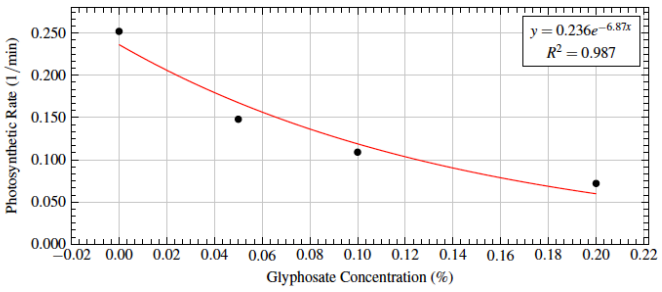


Figure 6. The relationship between the photosynthetic rate (indirectly measured via the mean rate of change in number of floating leaf disk) and glyphosate concentration. A best fit line (-) is presented. Data from Table 4.

However, plateauing at such a low concentration is somewhat unexpected given that glyphosate is commercially sold at around 50 % concentration. Yet publications show similar effects on different plants by using different techniques of measurement such as the direct measurement of the activity of PSII, the electron transport rate through it and/or shikimic acid content (Huang et al., 2012; Meloni et al., 2022). While our study show a decrease in photosynthetic rate post-exposure to glyphosate, some studies did not (Cedergreen & Olesen, 2010; Khan et al., 2020). One of which is (Khan et al., 2020) whose results

Vol. 2023 (2) 235 – 246
ISSN 2688-3651 [online]

seems to indicate the exact opposite effects. In all of the [GP]s, post- photosynthetic rate was higher than the ones pre-exposed to glyphosate up until higher concentrations; where effects are similar to ours. Both (Khan et al., 2020)

and (Cedergreen & Olesen, 2010) attempted to explain such findings by the "increased efficiency of CO_2 fixation" upon exposure to low-doses of [GP]s, but such mechanism requires further research as they suggested.

Yet nonetheless, our results indicate a clear, strong and prominent effect of glyphosate on the photosynthetic rate of *Spinacia oleracea* at low-doses and limited exposure time. Differences in time of sampling as well as the type, size and age of the plant may have led to such differences compared to (Khan et al., 2020) and (Cedergreen & Olesen, 2010). Moreover, their "low-doses" in their in vivo experiments may be considered "high-doses" in our in vitro investigation with leaf tissue used as small as 5.0 mm in diameter. Lastly, our indirect measurement of the photosynthetic rate could be a factor aiding in this disparity. This is also a limitation of our experiment preventing us from determining the exact causes of our quantitative observations. These limitations also extend to the lack of statistical tests and high standard deviation values, which may be caused by slight temperature differences between insufficient amount of trials, adherence of leaf disks to the bottom of the beaker, leaf disks held by the surface tension of water and respiration that consumes oxygen and decreases the buoyancy of the disks. All limitations may have contributed to the fluctuation in number of disks at any given time interval and between trials. If not bounded to the lack of apparatus, oxygen and carbon dioxide sensors could be used and more trials could be conducted to allow for statistical tests to be done and decreases the mentioned limitations specifically for the photosynthetic flotation experiment.

Table 4. The mean rate of change in the number of floating leaf disks between zero and 56 min for each glyphosate concentration (%)

Glyphosate Concentration (%)	Mean Rate of Change in [0, 56] (1⁄min)*
0.00	0.252
0.05	0.148
0.10	0.109
0.20	0.072

Note—There are no standard deviation values and statistical tests due to the way rates were obtained. * The photo- synthetic rate.

Table 5. The mean Initial and Final Total Chlorophyll Content (*TCC*, in µg⁄g) and the mean absolute % change of each glyphosate concentration

Glyphosate Concentration (%)	Mean *TC* Initial	*C* (µg⁄g) Final	Mean Absolute % Change
0.00	448.7 ± 7.2	445.1 ± 5.6	0.8 ± 0.6a
0.05	441.2 ± 14.5	362.0 ± 21.5	18.0 ± 2.6b
0.10	428.8 ± 27.9	311.2 ± 30.4	27.1 ± 9.2b
0.20	438.3 ± 9.5	267.6 ± 25.4	39.0 ± 5.3c

Note—Numbers represent Mean ± Standard Deviation for a sample set of data (n = 5). Letters indicate significant difference at $p < 0.05$ by one-way ANOVA with Tukey HSD post-hoc test. Statistical tests results in Table 3.

In regards to the absolute percentage change in the *TCC* of *Spinacia oleracea* (Figure 7), clear differences can be concluded between [GP]s. However, and contrary to our expectations, qualitative findings seem to indicate little to no color changes. Yet, quantitatively speaking, and similar to the photosynthetic rates, exposure to the lowest [GP] gave an astonishing 18.0 % change in *TCC* that is significant at α = 0.05 (Table 5). Higher concentrations gave higher percentage changes all of which are statistically significant at α = 0.05. However, and with a high standard deviation value (9.2 %), the absolute percentage change at 0.10 % [GP] is not statistically significant from its lower adjacent concentration at the same confidence level. The use of plastic cuvettes could have led to such a problem; as acetone was able to deform the cuvette's structure leading to the fracturing of the beam of light and falsely analyzed as absorbance. A cuvette of other material should be used.

As a chelator of minerals, glyphosate decreases the bioavailability of magnesium ions needed for chlorophyll synthesis by forming magnesium-glyphosate chelates that prevents magnesium-chelatase from catalyzing the addition of magnesium to PPIX (Farhat et al., 2016; Gomes et al., 2014; Guo et al., 2015; Huang et al., 2012). With higher [GP]s, the probability of forming the chelates and their quantity increases as indirectly indicated by our results (Figure 7) and (Huang et al., 2012). (Huang et al., 2012)'s results took days, which show that the damage done by glyphosate is rather progressive than immediate. Nonetheless, we can 95 % confidently conclude that the addition of glyphosate is the factor behind the decrease in *TCC*, which is more significant at higher [GP]s. This may contribute to the decrease of

Vol. 2023 (2) 235 – 246
ISSN 2688-3651 [online]

photosynthetic rates mentioned previously; as such pigments are needed in the initial stages of the light-dependent reaction (photoactivation), without which, the photolysis of water decreases and so does oxygen generation.

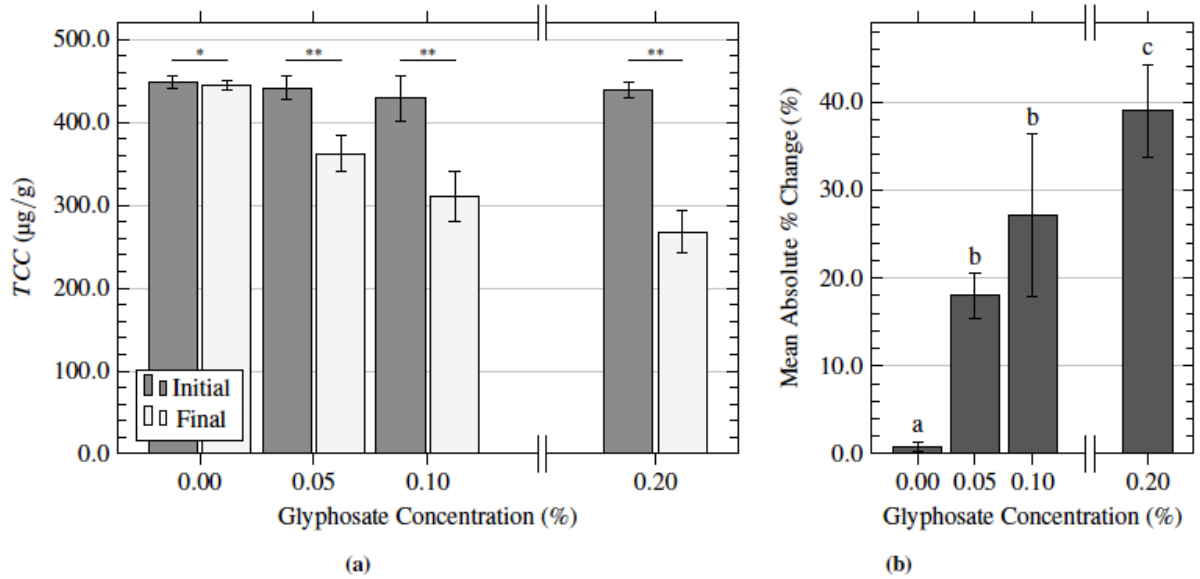


Figure 7. (a) The initial and final Total Chlorophyll Content (TCC) after exposure to each glyphosate concentration. Paired two-tailed t-Test was used to indicate significant difference at $p < 0.05$ (*) and $p < 0.01$ (**) (Table 2). (b) The average absolute % change in TCC after exposure to each glyphosate concentration. Letters indicate significant difference at $p < 0.05$ by one-way ANOVA with Tukey HSD Test as a post-hoc test (Table 3). Error bars in both (a) and (b) are of the standard deviation values with $n = 5$. Data from Table 5.

4. Conclusion

This study was able to determine the in vitro effects of increasing [GP] on both the photosynthetic rate and *TCC* of *Spinacia oleracea*. With prominent and strong effects even upon exposure to the lowest [GP] (0.05 %), it was clear and statistically significant that glyphosate is responsible for the changes caused to both dependent variables. Increasing [GP] also attributed to the decreases found in both dependent variables, which supports our hypothesis and present detailed studies. Such causes may be explained by the inhibitory effect of this chemical on EPSP synthase, the nature of this inhibition and the decrease in magnesium ion bioavailability needed for chlorophyll synthesis. With some dissimilarities in photosynthetic rates (inverse effects) and qualitative evidence (no notable color changes) with other publications, it may be merely the cause of our experimental limitations, differences in plant, exposure time, sampling time and the overall way our experiment was designed. Yet, with nature in mind, this study was able to generate conclusions in a way that minimizes harm to the environment and the ecosystem, and consistent, in most cases, with cited articles. The drawn conclusions could be of great interest to the users of this herbicide, increasing caution when spraying near similar non-targeted plants. While our study investigated the immediate effects of this chemical on two aspects of this organism, hormonal changes was not in mind and further research should be held investigating the reversibility of mentioned effects and the factors that decreases the mortality rate upon accidental exposure to this chemical.

References

METHODOLOGY. (n.d.). *King Saud University*, 54. https://fac.ksu.edu.sa/sites/default/files/methodology.pdf

Allakhverdiev, S. I., Tomo, T., Shimada, Y., Et al. (2010). Redox potential of pheophytin a in photosystem II of two cyanobacteria having the different special pair chlorophylls [Publisher: National Academy of Sciences Section: Biological Sciences]. *PNAS*, *107*(8), 3924–3929. https://doi.org/10.1073/pnas.0913460107

Vol. 2023 (2) 235 – 246
ISSN 2688-3651 [online]

Arnon, D. I. (1971). The Light Reactions of Photosynthesis [Publisher: National Academy of Sciences Section: National Academy of Sciences Annual Meeting (April 1971): Symposium]. *PNAS*, *68*(11), 2883–2892. https://doi.org/10.1073/pnas.68.11.2883

Cedergreen, N., & Olesen, C. F. (2010). Can glyphosate stimulate photosynthesis? *Pesticide Biochemistry and Physiology*, *96*(3), 140–148. https://doi.org/10.1016/j.pestbp.2009.11.002

de Souza, A. X., & Sant'Anna, C. M. R. (2008). 5-Enolpyruvylshikimate-3-phosphate synthase: Determination of the protonation state of active site residues by the semiempirical method. *Bioorg Chem*, *36*(3), 113–120. https://doi.org/10.1016/j.bioorg.2007.12.007

El-Najjar, N., Gali-Muhtasib, H., Ketola, R. A., Et al. (2011). The chemical and biological activities of quinones: Overview and implications in analytical detection. *Phytochem Rev*, *10*(3), 353. https://doi.org/10.1007/s111 01-011-9209-1

Farhat, N., Khouni, A., Zorrig, W., Et al. (2016). Effects of magnesium deficiency on photosynthesis and carbohydrate partitioning. *Acta Physiologiae Plantarum*, *38*. https://doi.org/10.1007/s11738-016-2165-z

Glyphosate, a chelating agent—relevant for ecological risk assessment? (n.d.). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5823954/

Gomes, M. P., Smedbol, É., Chalifour, A., Et al. (2014). Alteration of plant physiology by glyphosate and its by-product aminomethylphosphonic acid: An overview. *Journal of experimental botany*. https://doi.org/10.1093/jxb/eru269

Guo, B., Guo, Y., Hong, H., Et al. (2015). Co-expression of G2-EPSPS and glyphosate acetyltransferase GAT genes conferring high tolerance to glyphosate in soybean. *Frontiers in Plant Science*, *6*, 847. https://doi.org/10.338 9/fpls.2015.00847

How Acetone Affects Certain Plastics [Section: News and Updates]. (2020). https://www.millerplastics.com/how-acetone-affects-certain-plastics/

Huang, J., Silva, E. N., Shen, Z., Et al. (2012). Effects of glyphosate on photosynthesis, chlorophyll fluorescence and physicochemical properties of cogongrass (Imperata cylindrical L.). *undefined*. https://www.semanticscholar.org/paper/Effects- of- glyphosate- on- photosynthesis%2C- and- of- Huang- Silva/9ef9b5f50a40ff33d630406acf4c02b429998425

Khan, S., Zhou, J. L., Ren, L., Et al. (2020). Effects of glyphosate on germination, photosynthesis and chloroplast morphology in tomato. *Chemosphere*, *258*, 127350. https://doi.org/10.1016/j.chemosphere.2020.127350

Light-dependent reactions (photosynthesis reaction) (article). (n.d.). https://www.khanacademy.org/science/ap-biology/cellular-energetics/photosynthesis/a/light-dependent-reactions

Masojídek, J., Koblizek, M., & Torzillo, G. (2013). Photosynthesis in Microalgae. *Handbook of Microalgal Culture: Biotechnology and Applied Phycology*, 20–39. https://doi.org/10.1002/9781118567166.ch2

MATERIAL SAFETY DATA SHEET Glyphosate 5.4. (2007). *Alligare, LLC*. https://www.lakerestoration.com/pdf/glyphosatemsds.pdf

Meloni, D., Nieva, M., & Martínez, C. (2022). EFFECTS OF GLYPHOSATE ON GERMINATION AND PHOTOSYNTHESIS IN Prosopis alba G.: A BIOCHEMICAL APPROACH. *BIOFIX Scientific Journal*, *7*, 9. https://doi.org/10.5380/biofix.v7i1.81308

Molnar, C., & Gair, J. (2015). 5.3: The Calvin Cycle. In *Concepts of Biology - 1st Canadian Edition*. BCcampus. https://opentextbc.ca/biology/chapter/5-3-the-calvin-cycle/

Vol. 2023 (2) 235 – 246
ISSN 2688-3651 [online]

Mulo, P. (2011). Chloroplast-targeted ferredoxin-NADP+ oxidoreductase (FNR): Structure, function and location. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1807*(8), 927–934. https://doi.org/10.1016/j.bbabio.20 10.10.001

Nowicka, B., & Kruk, J. (2010). Occurrence, biosynthesis and function of isoprenoid quinones. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1797*(9), 1587–1605. https://doi.org/10.1016/j.bbabio.2010.06.007

Overview of Photosynthesis – Biology 2e. (n.d.). https:// opentextbc. ca/ biology2eopenstax/ chapter/ overview- of-photosynthesis/

Rissler, H., Collakova, E., DellaPenna, D., Et al. (2002). Chlorophyll Biosynthesis. Expression of a Second Chl I Gene of Magnesium Chelatase in Arabidopsis Supports Only Limited Chlorophyll Synthesis. *Plant physiology*, *128*, 770–9. https://doi.org/10.1104/pp.010625

SAFETY DATA SHEET Acetone. (2015). *Thornton & Ross Ltd*, 6. https://www.nhsggc.org.uk/media/236208/msds-acetone.pdf

Sapkota, A. (2020). Chlorophyll- Definition, Structure, Types, Biosynthesis, Uses. https://microbenotes.com/chlorophyll/

Spectrophotometry [Section: Spectrophotometry]. (n.d.). https://tipbiosystems.com/blog/spectrophotometry/

Su, S., Zhou, Y., Qin, J. G., Et al. (2010). Optimization of the Method for Chlorophyll Extraction in Aquatic Plants [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02705060.2010.9664402]. *Journal of Freshwa- ter Ecology*, *25*(4), 531–538. https://doi.org/10.1080/02705060.2010.9664402

Tzin, V., Galili, G., & Aharoni, A. (2012). Shikimate Pathway and Aromatic Amino Acid Biosynthesis [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470015902.a0001315.pub2]. In *eLS*. American Cancer Society. https://doi.org/10.1002/9780470015902.a0001315.pub2

US EPA, O. (2014). Glyphosate. https://www.epa.gov/ingredients-used-pesticide-products/glyphosate

Journal of Research High School

Vol. 2023 (2) 247 – 252
ISSN 2688-3651 [online]

Healthcare Workers Experiencing PTSD during the COVID-19 Pandemic

Xinyue Going[1] *

[1]Northwood High School, Irvine, CA, USA
*Corresponding Author: xinyuegong58@gmail.com

Advisor: Daniel Loh, write2loh@gmail.com

Received March 23, 2023; Revised August 17, 2023; Accepted, September 26, 2023

Abstract

Under the pressure of the COVID- 19 pandemic, mental illnesses such as post-traumatic stress disorder (PTSD) became more prevalent, especially among healthcare workers (HCW). This article examines the impacts of PTSD on HCWs during COVID- 19, analyzing the causes and risk factors, long term impacts, as well as existing coping methods. Specific stress factors from the pandemic include an increased workload and the fear of being infected. Other factors such as age and gender can also increase the risk of inflicting PTSD. Even though studies have reported that most PTSD symptoms persist for a lifetime, there are some helpful coping mechanisms already in use to help alleviate the symptoms. Through analyzing the causes, impacts, and coping methods of PTSD, it is probable that better treatments for this condition can be developed to help HCWs and many others who are struggling with similar situations.

Keywords: Mental Health, PTSD, Healthcare Workers, COVID-19, Coping

1. Introduction

In December 2019, the new and highly-contagious SARS-CoV2 or Coronavirus 2019 (COVID- 19) first emerged in Wuhan, China, and quickly spread throughout the world (Carmassi et al., 2020). The outbreak was officially deemed a global pandemic on March 11, 2020, by the World Health Organization (WHO). The pandemic disrupted societies, economies, and healthcare systems worldwide (Froessl and Abdeen, 2021). By August 2020, the death toll in the United States exceeded 187,000, creating an immediate demand for an effective vaccine and more healthcare workers (HCW) (Bender et al., 2021). Because they were expected to step up at the most dangerous time of the COVID pandemic, HCWs were directly exposed to the virus and an increased workload.

Kelsey Ryan, a 28-year-old critical care nurse in Michigan, was summoned to the frontline with her colleagues to face countless patients suffering from the deadly virus. Due to a lack of treatment and technology, Ryan witnessed the loss of hundreds of lives in just two months and became a victim of post-traumatic stress disorder (PTSD). PTSD affects people who have experienced traumatic events such as military action, natural disasters, sexual violence, or serious illness or injury. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), grouped symptoms of PTSD into four categories: intrusive memories, avoidance, negative mood, and hyperarousal (Johnson et al., 2020). “I still have nightmares every night," Ryan said after she woke up each night, choking and sweating (Wayland, 2020). Unfortunately, she is not alone. Many other HCWs suffered the same symptoms during the worst of the pandemic, and the continuous demand for work drained them of energy to recover from their mental struggles. Initial studies performed during the first wave of the pandemic indicate that more than 50% of the HCWs involved have experienced symptoms of depression, anxiety, and PTSD (Benham et al., 2020).

This paper examines the effects of the COVID- 19 pandemic on the mental health of HCWs in the United States, especially focusing on the increased cases of PTSD. The National Comorbidity Survey Replication, a national representative community that holds surveys of the prevalence and correlations of mental disorders in the US,

Vol. 2023 (2) 247 – 252
ISSN 2688-3651 [online]

"estimated the lifetime prevalence of PTSD among adults in the United States to be 6.8%." However, the prevalence rate of PTSD in HCWs is significantly higher, ranging from 15% to 20%, depending on the location (Benham et al., 2020). By analyzing the psychological symptoms of HCWs during COVID- 19, this article aims to contribute to the finding of effective intervention methods for these symptoms to help HCWs prepare for future health crises. To ensure the validity of anticipated findings, this paper is developed after a thorough selection of past research and literature. A diverse range of texts focusing on the objectives of this paper have been studied to take into account multiple perspectives and provide a more accurate depiction of how HCWs develop PTSD from the COVID pandemic.

2. Causes and Risk Factors

During a pandemic, HCWs are expected to respond immediately to medical emergencies, creating high psychopathology rates (Amsalem et al., 2021). This is primarily due to the many stressors of a pandemic, such as a rapid increase in workload, lack of effective treatment and technology, and witnessing a significant number of deaths. HCWs also have to work overstraining long hours while worrying about their health and the health of their loved ones (Froessl and Abdeen, 2021). The various national public health membership associations distributed surveys to their members during the COVID– 19 pandemic crisis. From the 26, 174 responses received, 54.2% reported that they were working more than 60 hours per week, 51.5% could not take time off, and 47.0% spent more than 75% of their time on COVID– 19 response activities. As a result, they were exposed to a higher risk of PTSD (See Figure 1), and 36.8% of all respondents reported symptoms of PTSD (Bryant-Genevier et al., 2021). In addition, restrictive safeguards during the crisis exacerbated the situation by "limiting physical contact, social activities, religious services, and other traditional means of emotional connection," creating another driving factor for mental health concerns (Bender et al., 2021).

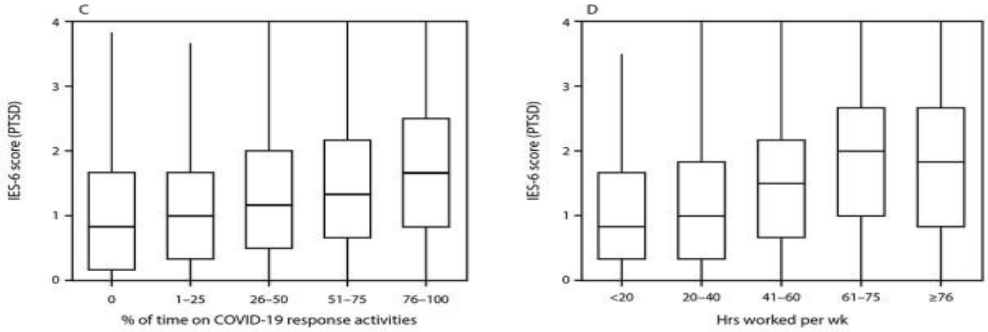


Figure 1. Distribution of 6-item Impact of Event Scale scores for post-traumatic stress disorder among state, tribal, local, and territorial public health worker respondents, by percentage of work time spent directly on COVID- 19 response activities for the majority of 2020 (panel C), and hours worked in a typical week since March 2020 (panels D) (Bryant-Genevier et al., 2021)
*IES-6 = 6-item Impact of Event Scale; PTSD = post-traumatic stress disorder

Although the COVID- 19 outbreak caused a general increase in PTSD among HCWs, independent risk factors can influence the prevalence and severity of psychological symptoms. These factors include age, gender, occupation, and specialization. Being young, female, and a nurse were significant risk factors for mental health issues like PTSD. For instance, a survey of 1, 132 HCW participants revealed that women had over double the odds of developing PTSD compared to men (Hennein et al., 2020). Another study with 657 HCWs in New York City indicated that among those who screened positive for PTSD, 64% were nurses, which was significantly higher than the 40% of attending physicians with PTSD (Shechter et al., 2020). Moreover, during the global pandemic, many governments carried out lockdowns, which shifted schooling, parenting, and household duties to parents, and these responsibilities ultimately landed on women (Froessl and Abdeen, 2021). As a result, the pandemic indirectly imposed additional contributing factors for women to get PTSD.

Vol. 2023 (2) 247 – 252
ISSN 2688-3651 [online]

3. Long Term and Health Impacts

Besides the direct symptoms of PTSD, the disorder can lead to detrimental long-term impacts on the lives of HCWs, creating lasting memories and stress from the pandemic. In a recent study of 350 HCWs, 35% reported symptoms suggesting probable PTSD, with 28% reporting symptoms that lasted 30 days and 24% 90 days (See Table 1) (Amsalem et al., 2021). These HCWs will regularly be jarred into a world of painful flashbacks. As a result, they often feel disconnected from families and friends. They have lower morale, lower productivity, and prolonged absences, which might also negatively affect healthcare organizations during emergencies like the pandemic (Bryant-Genevier et al., 2021).

Table 1. Longitudinal Presentation of the Percentage of Positive Cases for Self-Report Anxiety (GAD-7 ≥ 5), Depression (PHQ-9 ≥ 5), Suicidal Ideation (PHQ9, Item 9), and PTSD (PC-PTSD≥3) (Amsalem et al., 2021)

Items	Baseline n = 350	Day 30 n = 280	Day 90 n = 267	p-value
	n (%)	n (%)	n (%)	
Severity level of anxiety (GAD)				
Mild	100 (29)	76 (27)	72 (27)	
Moderate	66 (19)	44 (16)	46 (17)	
Severe	49 (14)	26 (9)	27 (10)	
Total	215 (62)	146 (52)	145 (54)	.007
Severity level of depression (PHQ)				
Mild	88 (25)	68 (24)	68 (26)	
Moderate	60 (17)	42 (15)	39 (15)	
Moderately severe	34 (10)	27 (10)	27 (10)	
Severe	20 (6)	9 (3)	12 (5)	
Total	202 (58)	146 (52)	148 (55)	.091
Suicidal ideation	65 (19)	38 (14)	49 (18)	.149
PC-PTSD	121 (35)	79 (28)	65 (24)	.000

Additionally, the immense workloads and pressure from the pandemic increased burnout among HCWs, causing physical and emotional exhaustion, depersonalization, and a decreased sense of personal accomplishment (Restauri and Sheridan, 2020). There is also an association between PTSD and burnout, as 98% of those meeting the diagnostic criteria for PTSD were positive for burnout syndrome (Mealer et al., 2009). Burnouts often create a lack of purpose and motivation in HCWs, causing the formation of suicidal thoughts and substance abuse. For example, a study of 1354 respondents showed that burnout increased the odds of suicidal ideation by 85% (Menon et al., 2020). However, the healthcare industry cannot risk losing a single worker during an urgent crisis.

In addition, another study showed that 8% of plastic surgeons with burnout syndromes were positive for alcohol abuse, and 5% were positive for substance abuse (Khansa and Janis, 2019). Through the use of alcohol and drugs, though, HCWs are only seeking temporary relief from mental stress at the cost of physical health damage. Some of the health impacts of PTSD include lower sleep quality and the potential development of obesity (DeLucia et al., 2019). HCWs with PTSD constantly live in a miserable state, and they might not even be able to distinguish delusions from reality. This dramatically exacerbates their quality of life and ruins their careers. However, HCWs are in the healthcare system because people rely on them to respond during emergencies and save lives, making it necessary for HCWs to cope with psychological difficulties and develop resilience.

4. Coping with PTSD

While PTSD can have chronic symptoms and is very hard to recover from, studies have found helpful coping mechanisms that can be implemented effectively to help prevent and minimize the effects of the disorder in HCWs. For instance, six specific psychological intervention programs are identified across 12 studies: psychological first aid (PFA); eye movement desensitization and reprocessing (EMDR); anticipate, plan, and deter (APD); resilience at work (RAW); resilience and coping for the healthcare community (RCHC) and trauma risk management (TRiM) (Hooper et al., 2021). Although further research is needed to test the effectiveness of these programs, initial evaluations have found that each program reduced the psychological impact of PTSD to a certain extent. APD and RCHC might be the most suitable for HCWs as they are newly designed and intended for the HCW population (Hooper et al., 2021).

Vol. 2023 (2) 247 – 252
ISSN 2688-3651 [online]

APD is a model that helps HCWs learn to anticipate, plan, and deter potential problems, which emotionally prepares them for the specific stressors they will be facing. A qualitative study that implemented APD in two responder teams during Africa's 2014–2015 Ebola epidemic found that APD effectively reduced psychological problems in 90% of the high-risk HCWs (Schreiber et al., 2019). Likewise, RCHC is a 3 to 5- hour interactive workshop facilitated by trained social workers, counselors, and psychologists who hold either a master's degree or a Ph.D. It covers many topics, including common types of stress, reactions to severe stress and trauma, and individual and collective strategies to cope with stress and traumatic events (Yuma et al., 2019). Even though more research is needed, APD and RCHC are helpful tools to educate HCWs to deal with stress, allowing them to develop resilience in difficult circumstances and maintain a positive mindset at work.

Other more convenient and immediately accessible interventions have also proven effective. Some individual strategies are “micro-practices," which require only a few seconds and are readily available for healthcare individuals to manage stress. These include mindfulness practices, breathing exercises, and limiting exposure to social media (Restauri and Sheridan, 2020). System-based interventions can also be used, in which hospitals can enhance working conditions and medical technologies, expand staffing size, and implement flexible schedules with sufficient breaks to provide a safe and comfortable working environment and reduce overwork (Bryant-Genevier et al., 2021). While HCWs can receive facilitated help to cope with PTSD, communal coping is another mechanism that allows people to connect emotionally with others and support each other in a shared, stressful situation. For example, a qualitative study on intensive care unit nurses, who are known to experience high levels of work-related post-traumatic stress, indicates the importance of a positive social network to develop resilience because it acts as a primary source of support and encouragement during challenging times (Bender et al., 2021). In a following study, a sample of 89 former HCWs in the US provided insights on how they perceive the quality, structure, and strategies for emotional connection, demonstrating how emotional connection has been a crucial part of their coping routines during intense work times (Bender et al., 2021).

5. Discussion

The COVID- 19 pandemic created a global outbreak of anxiety and hysteria as people faced a deadly virus. There was a high demand for HCWs who stood at the frontlines and put themselves at risk of contracting this deadly virus; by doing so, they also put themselves at risk for major mental health disorders like PTSD. They experienced a dramatic increase in workload and constantly faced a fear of infecting themselves or their loved ones. Although this study has focused on HCWs in the United States, there seems to be consistent data and similar findings in research from other countries. For example, similar risk factors may exist globally, influencing the severity of psychological symptoms, and similar coping mechanisms can be applied to HCWs worldwide. However, due to the recency of the pandemic, there might be limitations in sample selections and research methods during data collection, as well as insufficient data to cover all aspects of the pandemic and its effects on mental health. Therefore, more studies about PTSD on HCWs should be conducted in the next few years to answer more questions. Overall, this study emphasizes how the pandemic has led to mental health issues like PTSD, its long-term impacts, and the importance of coping strategies for HCWs to ensure accurate medical decision-making and continued professional performance.

6. Conclusion

The COVID- 19 pandemic has left permanent scars in the minds of many HCWs. The mountainous workload and the innumerable lives on their shoulders contributed to increased psychological difficulties for HCWs. Also, they constantly feared death and worried about their families and friends. However, studies have shown that women are more susceptible to mental problems like PTSD, increasing the prevalence rate among nurses because it is a more female-dominated specialization. The pandemic significantly increased the number of responsibilities placed on women as they have to take care of both their critically-ill patients and their families at home. Burnouts caused by stress and workloads in HCWs can also lead to suicide risks and physical health problems. Fortunately, many HCWs are able to cope with their symptoms through emotional connection and some coping mechanisms that are proven

Vol. 2023 (2) 247 – 252
ISSN 2688-3651 [online]

effective for PTSD. Nurses like Kelsey Ryan are recovering, but the debilitating COVID- 19 experience will stick with them forever.

In this paper, the HCWs examined are those who weren't directly infected by the COVID- 19 virus itself. Still, they suffered all the same from the extreme stress and pressure of taking care of the patients and witnessing countless deaths. Even though the analysis of how the pandemic increased the risk of PTSD in HCWs is valuable information about PTSD and other related mental illnesses, that's not enough to improve on existing coping mechanisms. Therefore, it is suggested that more research should be done on behaviors of HCWs who either already had PTSD from past experiences or directly suffered from the COVID- 19 virus. Data from studies conducted on various aspects of the topic will allow comparisons to be made between those who have slightly different circumstances. This will eventually help in the development of better treatments for a wide range of victims in the future. By then, not only will HCWs be benefitted, but those who work in other dangerous and highly stressful professions will also suffer less.

References

Amsalem, D., et al. (2021, November 5). Psychiatric symptoms and moral injury among us healthcare workers in the COVID- 19 ERA - BMC psychiatry. BioMed Central. Retrieved January 5, 2023, from https://0-bmcpsychiatry-biomedcentral-com.brum.beds.ac.uk/articles/10.1186/s12888-021-03565-9

Bender, A. E., et al. (2021). "Making Sure We Are All Okay": Healthcare Workers' Strategies for Emotional Connectedness During the COVID- 19 Pandemic. Europe PMC. Retrieved January 5, 2023, from https://europepmc.org/article/MED/33456094

Benham, T. L., et al. (2020, October 13). Preparing for the second surge: Preventing posttraumatic stress disorder and building resilience for health care workers in the face of covid- 19: Disaster medicine and public health preparedness. Cambridge Core. Retrieved January 5, 2023, from https://www.cambridge.org/core/journals/disaster-medicine-and-public-health-preparedness/article/preparing-for-the-second-surge-preventing-posttraumatic-stress-disorder-and-building-resilience-for-health-care-workers-in-the-face- of-covid19/2020C7548677260B51F8C1311D397B0F

Bryant-Genevier, J., et al. (2021, July 1). Symptoms of depression, anxiety, post-traumatic stress disorder, and suicidal ideation among state, tribal, local, and territorial public health workers during the COVID- 19 pandemic - United States, March–April 2021. Centers for Disease Control and Prevention. Retrieved January 5, 2023, from https://www.cdc.gov/mmwr/volumes/70/wr/mm7026e1.htm

Carmassi, C., et al. (2020, October). PTSD symptoms in healthcare workers facing the three coronavirus outbreaks: What can we expect after the COVID- 19 pandemic. Psychiatry research. Retrieved January 5, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7370915/

DeLucia, J. A., et al. (2019, February 5). Prevalence of post-traumatic stress disorder in emergency physicians in the United States. Western Journal of Emergency Medicine: Integrating Emergency Care with Population Health. Retrieved January 5, 2023, from https://escholarship.org/uc/item/3b02x51n

Froessl, L. J., & Abdeen, Y. (2021, September 30). The silent pandemic: The psychological burden on frontline healthcare workers during COVID-19. Psychiatry journal. Retrieved January 5, 2023, from https://pubmed.ncbi.nlm.nih.gov/34631873/

Hennein, R., Mew, E. J., & Lowe, S. R. (2020). Socio-ecological predictors of mental health outcomes among healthcare workers during the COVID-19 pandemic in the United States. PLOS ONE. Retrieved January 7, 2023, from https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0246602

Hooper, J. J., et al. (2021, May 1). Addressing the psychological impact of covid-19 on Healthcare Workers: Learning from a systematic review of early interventions for Frontline Responders. BMJ Open. Retrieved January 5, 2023, from https://bmjopen.bmj.com/content/11/5/e044134

Vol. 2023 (2) 247 – 252
ISSN 2688-3651 [online]

Khansa, I., & Janis, J. E. (2019). A growing epidemic: Plastic Surgeons and Burnout – a literature review ... Special Topic. Retrieved January 11, 2023, from https://www.researchgate.net/publication/333721073_A_Growing_Epidemic_Plastic_Surgeons_and_Burnout_-_A_Literature_Review

Mealer, M., et al. (14 November 2009). The prevalence and impact of post traumatic stress disorder and burnout syndrome in nurses. Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1002/da.20631. Retrieved: February 4, 2023

Menon, N. K., et al. (2020, December 9). Association of Physician Burnout with suicidal ideation and medical errors. JAMA Network Open. Retrieved January 11, 2023, from https://jamanetwork.com/journals/jamanetworkopen/article-abstract/2773831

Restauri, N., & Sheridan, A. D. (2020, May 27). Burnout and posttraumatic stress disorder in the coronavirus disease 2019 (covid-19) pandemic: Intersection, impact, and interventions. Journal of the American College of Radiology. Retrieved January 5, 2023, from https://www.sciencedirect.com/science/article/pii/S1546144020305469

Schreiber, M., et al. (2019). Maximizing the Resilience of Healthcare Workers in Multi-hazard Events: Lessons from the 2014–2015 Ebola Response in Africa. Academic.oup.com. Retrieved January 12, 2023, from https://academic.oup.com/milmed/article/184/Supplement_1/114/5418686

Shechter, A., et al. (2020, June 16). Psychological distress, coping behaviors, and preferences for support among New York healthcare workers during the COVID-19 pandemic. General Hospital Psychiatry. Retrieved January 7, 2023, from https://www.sciencedirect.com/science/article/abs/pii/S0163834320300839

Wayland, M. (2020, June 27). 'I still have nightmares every night' - health workers struggle with PTSD symptoms as coronavirus takes toll. CNBC. Retrieved January 7, 2023, from https://www.cnbc.com/2020/06/27/coronavirus-takes-emotional-toll-health-workers-suffer-ptsd.html

Yuma, P., et al. (2019). Resilience and coping for the Healthcare Community: A post-disaster group work intervention for healthcare and Social Service Providers. DigitalCommons@TMC. Retrieved January 12, 2023, from https://digitalcommons.library.tmc.edu/jfs/vol19/iss1/8/

Journal of Research High School

Vol. 2023 (2) 253 – 261
ISSN 2688-3651 [online]

Effects of Financial Literacy on Well-being and Financial Perception among High School Students

Justin Shin[1] *

[1]Ames High School, Ames, IA, USA
*Corresponding Author: justinshin1117@gmail.com

Advisor: Nikole Schieffer, niki.schieffer@ames.k12.ia.us

Received May 11, 2023; Revised August 31, 2023; Accepted, September 26, 2023

Abstract

High school students face financial obligations and activities in today's financial environment. Understanding the determinants of financial perception and consumption attitudes is important for improving financial well-being among high school students. The commonly accepted relationship between financial literacy and financial well-being is often examined among the adult population but rarely examined among high school students. The study aimed to assess the magnitude of the influence of financial literacy on financial perception and consumption attitudes among high school students. In addition, the author also investigated the role of lifestyle depletion on financial perception and consumption attitudes to identify the magnitude of this influence. High school students in the U.S. were given an online survey with questions about financial literacy, lifestyle depletion, financial perception, and consumption attitudes. Data were analyzed using multiple regression tests. The findings indicate that teenagers with higher financial literacy perceive their financial situation more positively and are more likely to have saving-oriented consumption attitudes. On the other hand, teenagers with higher levels of lifestyle depletion perceive their financial situation more negatively and are more likely to endorse saving-oriented consumption attitudes. This study contributes to the literature by showing the differential roles of financial literacy and lifestyle depletion on financial perception and consumption attitudes. The discussion includes the importance of financial literacy education and the role of lifestyle depletion in the financial well-being of high school students.

Keywords: Financial literacy, Financial perception, Depletion, Consumption attitudes, Multiple hierarchical regression

1. Introduction

In the coming years, current teenagers can shape the future of economic markets. Currently, teens have increasing purchasing power and access to financial information and markets (Duffett, 2017). When they become more involved in financial activities and decisions, it is critical to understand their consumption attitudes and financial perceptions, which could affect their financial decisions and behaviors.

In this study, the author examines two important factors contributing to high school students' financial perception and consumption attitudes: financial literacy and lifestyle depletion. Previous literature pointed out the importance of financial literacy from the perspective of knowledge and resources among college students (Fan & Chatterjee, 2019). Recent studies also recognize the importance of current psychological resources and demands that would influence teenagers' financial attitudes and intentions. When people lean towards a saving tendency and attitude, they tend to be more cautious with their spending on products or services. Conversely, if individuals hold positive views about indulgent spending, their financial choices and behavior are more likely to lean towards indulgent spending rather than saving.

Vol. 2023 (2) 253 – 261
ISSN 2688-3651 [online]

1.1 Financial Literacy

Financial literacy is defined as the ability to understand and analyze financial options, make financial plans, and respond appropriately to financial events (Fox et al., 2005; Remund, 2010). Financial literacy is often a reliable predictor of personal wealth and savings. It is known to positively impact a person's financial management. For high school students, financial literacy is related to saving income from their jobs, managing allowances, and making necessary purchases. The link between financial literacy and consumption behavior has been well-established in the literature, with studies showing that individuals who are more financially literate tend to make better financial decisions and have more positive financial outcomes. (Fariana et al., 2021)

On the other hand, a lack of financial literacy can have negative consequences such as frequent borrowing, impulsive purchases, and managing personal resources poorly, which could lead to financial difficulties and personal debt (Farinella, Bland, & Franco, 2017). Most studies examining financial literacy have investigated how it affects effective financial management, making informed decisions, and increasing savings among adults, including college students (Remund, 2010). As more high school students start to get involved in financial behaviors, it would be essential to understand to what extent financial literacy is associated with economic perception, especially among high school students.

Another critical aspect of consumption behavior among high school students is impulsive buying and unexpected purchases to satisfy their immediate needs or wants. Examining hedonistic consumption behavior or attitudes among teenagers is especially relevant as they form a sizeable target market and financial strength (Dayan et al., 2010). Recent studies have shown that the surge in social media marketing has influenced teenagers' consumption attitudes and behavior (Duffett, 2017). That is, teenagers are now exposed to products and goods that were not accessible to them. As studies indicated that financial literacy could help individuals make informed and planned purchases for young adults (Anisa et al., 2020; Hamid, & Loke, 2021), it would be important to examine whether financial literacy predicts their consumption orientation, such as saving-oriented and hedonic-oriented consumption attitudes among high school students.

1.2 Lifestyle Depletion

Another important factor affecting financial perception and consumption attitudes is lifestyle depletion. A growing body of research suggests that lifestyle depletion can also significantly impact consumption attitudes and financial well-being (Montoya & Scott, 2013). Lifestyle depletion refers to the depletion of an individual's physical and mental resources, which can occur when they are faced with prolonged stress or other challenging life circumstances (Baumeister & Vohs, 2007).

Thus, this research also explores the role of lifestyle depletion on their financial perception. Previous studies have examined this lifestyle depletion among adults with multiple responsibilities and obligations that exhaust their self-control (Baumeister & Vohs, 2007). Researchers found that consumers often have limited self-regulatory resources to exert self-control (Strack &Deutsch, 2006), and these resources are likely to be diminished when they experience higher levels of depletion when dealing with life stress. Research has shown that people with higher depletion are less likely to self-regulate and more likely to have negative expectations about the future (Sharma, Tiwari, Zutshi, & Singh, 2023). Lifestyle depletion is a commonly experienced phenomenon among high school students and is often linked to role-related stress. As many teenagers may experience role overload due to occupying multiple roles, such as student, helping parents, and household responsibilities, these high levels of depletion may influence their financial perceptions about the future.

This psychological stress can exhaust self-control and poor decision-making, including impulsive or irrational financial choices, higher depletion may make less optimal financial decisions such as impulsive buying for fun or immediate rewards or saving less money (Baumeister & Vohs, & Tice, 2007). Since lifestyle depletion affects self-regulation influencing financial intentions and motivations, I expect lifestyle would shape consumption attitudes among high school students. For example, students with high burdens and depletion may feel less likely to exercise self-control, resulting in hedonic spending and less saving. Scholars who examined college students' debt found that

Vol. 2023 (2) 253 – 261
ISSN 2688-3651 [online]

life stressors often led them to lose control of their debt decision, resulting in negative psychological well-being (Peltier, Dahl, & Schibrowsky, 2016). By the same token, similar patterns are expected to be found among high school students.

1.3 Significance of this Research

Although researchers have extensively examined the impact of financial literacy and lifestyle depletion and its effects on young adults or the general population, few studies have explored the potential impact of depletion and financial literacy on teenagers' financial perception and consumption attitudes. High school is a period of significant transition, during which young people begin to make important decisions about their future, including their finances. Teenagers are engaging in markets by working, spending, and saving. Given the importance of both financial literacy and lifestyle depletion on consumption behavior and financial well-being, examining the role of these factors in affecting their financial perception and consumption attitudes would provide valuable insights into promoting healthier financial habits and financial well-being.

In summary, examining the role of financial literacy and lifestyle depletion among high school students is crucial to support their financial well-being and future success. In this regard, the current investigation aims to the impact of financial literacy and lifestyle depletion among adolescents on their financial perceptions and consumption attitudes. The study will help us to identify potential interventions or strategies to improve this population's financial literacy and decision-making skills.

1.4 Hypothesis

The current study examines whether financial literacy and lifestyle depletion are associated with high school students' financial perception and consumption behavior. Based on the literature, the following hypotheses were derived.

H1. I expect that there will be a positive relationship between financial literacy and financial perception among high school students.
H2. I expect that there will be a negative relationship between lifestyle depletion and financial perception among high school students.
H3. I expect financial literacy to be positively associated with saving-oriented consumption attitudes (H3a), and negatively with hedonic consumption attitudes (H3b).
H4. I expect that lifestyle depletion will be negatively associated with saving-oriented consumption attitudes (H4a) and positively associated with hedonic consumption attitudes (H4b).

Additionally, I control various demographic factors such as race, gender, health status, and work status.

2. Materials and Methods

2.1 Sample

The data is collected from a Qualtrics survey. After questions were developed and adapted from previous studies, the survey was distributed to high schools across the United States. The author used social networks and the national high school registry to recruit participants. To be eligible, participants had to be a) currently attending high school in the States, b) willing to answer financial-related questions including over 100 survey questions. A total of 380 high school students participated in the study via Qualtrics survey. Participants could skip questions or stop the survey if they felt uncomfortable. The primary data collection period was from July 2021 and September 2021. The study was conducted anonymously, and participants' identities were not disclosed.

Vol. 2023 (2) 253 – 261
ISSN 2688-3651 [online]

2.2 Measure

Perceptions of Financial Future.

Participants were asked to rate their perceptions of current financial situations from 'Very Bad' (1) to 'Very Good' (5), so that higher scores mean optimistic views about their financial situation.

Attitudes towards Consumption.

The measurement of attitudes toward consumption was based on previous studies (Aro & Wilska, 2014). Participants were given nine statements related to their attitudes towards consumption and asked to rate each statement on a five-point Likert scale ranging from 1 ("Completely disagree") to 5 ("Completely agree"). Researchers found three factors based on this scale, saving-oriented, deprived, and hedonistic consumption. However, deprived and saving-oriented items are similar in content, so I used saving-oriented and hedonistic consumption for this study. Mean scores were calculated for each attitude based on these domains.

Financial fitness for life.

The Council for Economic Education created the FFFL-HS test to test the academic achievement of high school students in personal finance and economics based on the FFFL curriculum (Walstad & Rebeck, 2005). The test comprises 50 questions, with each theme consisting of 10 items. The sum score was computed so that higher scores indicate higher financial fitness.

Lifestyle-Based Depletion.

Depletion among high school students was measured with a modified version of depletion from the previous study by Montoya and Scott (2013). This measure taps into role-related stress among high school students. Based on the research, researchers have added four more relevant items. The modified scale includes 12 items that capture relevant lifestyle depletion among high school students. The mean score was computed so that higher scores indicate higher levels of lifestyle-based depletion.

Demographics

The high school students who participated in the study were asked to complete a short demographic survey that included questions about their gender, self-assessed health status (rated on a scale from 1=very bad to 5=very good), race (coded as 1=White, 2=Black, 3=Hispanic, 4=Asian, or 5=Other), number of siblings, and their mother's educational level (rated on a scale from 1=no school completed to 8=Master's, professional or doctoral degree).

Data Analysis

All statistical analyses were performed using Statistical Package for Social Sciences software (SPSS), version 26. Before the main analysis, descriptive analyses were conducted to describe the demographic characteristics of high school students who participated in the online survey. Second, the author used a linear regression model to examine the effect of financial literacy and lifestyle depletion on high school students' perception of financial situations and three types of consumption attitudes (Hypothesis 1-4). Dependent variables were financial perception and hedonic and saving-oriented consumption attitudes. Independent variables were financial literacy and lifestyle depletion.

To examine hypotheses, the author ran multiple linear regression models to test the effects of financial literacy and lifestyle depletion on financial perceptions and consumption attitudes. Gender, race, self-rated health, mother's education, and work status were included as covariates to adjust confounding variables. A linear regression model is a common method to uncover whether a relationship between independent and dependent variables exists. The following equation describes the linear models that the author tested in this study. The ordinary least squares (OLS) method was used for estimating the intercept and coefficients of independent variables.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \epsilon$$

Y= Dependent variables, β_0= the intercept, β_1, β_2= coefficients of independent variables, X_1,X_2= Independent variables (i.e., financial literacy, lifestyle depletion), and ϵ= error term.

Vol. 2023 (2) 253 – 261
ISSN 2688-3651 [online]

Adjusted R-squared is a way to measure how well a model explains the variation in data, considering both the number of predictors and the number of samples used in the analysis. It goes up only when adding a new variable makes the model's predictions better, considering how many predicting variables are examined in the regression analysis.

Before the regression analysis, a correlation matrix of all variables was calculated to assess multicollinearity. All correlation coefficients were below .30, so there were no concerns. In particular, the correlation between two independent variables was .14, which was not significant. Each hypothesis was tested using a two-tailed analysis at the α=.05 level of statistical significance. The author also chose statistical significance at the level of 0.05, often denoted as "p < 0.05." This is a common way researchers use to decide if their findings are meaningful or just due to chance. If something is statistically significant at the 0.05 level, it suggests that there's a reasonable basis to believe that the observed effect or relationship the current study found is likely real and not just random chance.

3. Results

Table 1 presents 380 respondents of the survey's demographics, including their age, self-identified gender, race, self-rated health (1=not good, 5=very good), work status (0=working, 1=not working), and mother's education (years of education). Respondents included 164 female and 194 male high school students. Twenty-two students did not wish to answer or skip the gender questions. Regarding race, most participants (*N*=206, 54.2%) were White. For Hypothesis 1, Race category (1=White, 2=Black, 3=Hispanic, 4=Asian, 5=Other) was recoded into two categories (0=White, 1=Minority). Regarding financial literacy scores, the average score was relatively higher than other national studies (Jang et al., 2014). In addition, the lifestyle depletion score was rather high compared to previous studies (Montoya & Scott, 2013), which found an average score of 2.87.

Table 1. Univariate descriptions of study variables (*N*=380)

Variable	%	Mean	*SD*	Range
Gender (Female)	45.8			
Race (White)	54.2			
Work status	51.1			
Self-rated Health		3.75	.89	1-5
Mother's education		14.4	2.16	1-18
Financial Literacy		62.2	4.37	0-65
Lifestyle Depletion		3.12	.98	1-5
Financial Perceptions		3.24	1.10	1-5
Consumption Attitudes		16.55	5.46	4-28
Hedonic		9.64	2.34	4-15
Saving Oriented				

*Note. *S.D.* stands for standardized deviation.

To examine Hypothesis 1-4, I conducted three multiple regressions to determine whether financial literacy and lifestyle depletion (independent variables) predicted financial perception and consumption attitudes (hedonic and saving-oriented) in separate models. The tested models and predictors are seen in Table 2, where each column represents different regression results. In all the regression models, possible demographic covariates, including gender, race, health, sibling size, mother's education, and work status, were included.

Table 2. Regression table of predicting financial perception and consumption attitudes

Variables	Financial Perception	Hedonic	Saving-oriented
	β	*β*	*β*
Gender	-.08	-.05	.030
Race	-.11*	-.07	-.060
Self-rated health	.15*	.11*	.24**
Mother's education	.013*	.03	.04
Work status	.020	-.085	.036
Financial Literacy	.15*	-.12*	.10*
Lifestyle Depletion	-.11*	-.01	.171**
Adjusted R^2	.12	.130	.076

Note. *β* represents standardized regression coefficients. Gender is defined as male (0) and female (1). Work status is defined as working (0) and not working (1). Sibling size and mother's education in years are defined as continuous. Race/ethnicity is dummy coded as White (0) and minority as (1). *p<.05, **p<.01.

The first hypothesis on financial perception was supported. As seen in Table 2, financial literacy was positively associated with perceptions about current finances (*β* =.15, p<.05). On the other hand, lifestyle deletion was negatively associated with financial perceptions (*β* =-.12, p<.05). In terms of demographic variables, race (minority) was negatively associated with financial perception

Vol. 2023 (2) 253 – 261
ISSN 2688-3651 [online]

whereas mother's education (higher education in years) was positively related to financial perception. That is, minority students were more likely to have negative perceptions about their financial situations. On the other hand, students whose mother's education level was high reported positive perceptions about their financial situations.

The second hypothesis was partially supported. Higher financial literacy was negatively associated with hedonic consumption attitudes (β =-.12, p<.05). Lifestyle depletion did not predict hedonic consumption attitudes. Regarding demographic variables, self-rated health was positively associated with hedonic consumption attitudes (β =.11, p<.05).

The third hypothesis was also supported. Financial literacy was a significant predictor of saving-oriented consumption attitudes (β =.10, p<.05), and lifestyle depletion was a predictor of saving-oriented consumption attitudes (β =-.17, p<.05) in the opposite direction. Students with higher financial literacy scores and lifestyle depletion scores reported higher saving-oriented consumption attitudes. Regarding demographic variables, only self-rated health was associated with saving-oriented consumption attitudes (β=-.24, p<.01), which suggests that current health status is also related to their saving intentions.

4. Discussion

The results from this study demonstrate the role of financial literacy and lifestyle depletion on the financial well-being of high school students. In particular, this study examined whether financial knowledge and life stress burdens are associated with high school students' financial perception and consumption attitudes. The general findings confirm previous studies by highlighting the roles financial literacy and lifestyle depletion play in high school students' perceived and anticipatory financial intentions. Current results are generally consistent with previous research that suggests that financial literacy is a useful skill and knowledge. In contrast, lifestyle depletion may play a bit complicated role in affecting the perceived financial situation and saving intentions.

First, finding the association between financial literacy and a positive financial perception among high school students is consistent with previous research showing that a higher level of financial literacy helps individuals to make informed financial decisions and develop a realistic view of their finances (Lusardi et al., 2010). Thus, it is possible that high school students with higher levels of financial literacy have the skills and knowledge to manage their budget and participate in economic activities, which enable them to see their financial circumstances in a positive direction. Indeed, previous studies have shown that students taking personal finance courses have a greater knowledge of finance and demonstrate responsible financial behavior, such as saving more money and using credit cards responsibly (Gutter, Copur, and Garrison 2010; Mandell & Klein, 2009).

Second, positive association between financial literacy and saving-oriented consumption attitudes are consistent with previous research on the effect of financial literacy on saving behavior and reduced spending among young adults (Farinella et al., 2017). The current findings suggest that the influence of financial literacy on saving attitudes may begin early in high school, emphasizing the importance of financial literacy in the high school curriculum. It also indicates that high school students who acquire knowledge and skills to make financial decisions may create a budget that will maximize their savings.

Although previous studies have shown that depletion may be positively associated with hedonic consumption attitudes in young adults (Cui et al., 2021; Dey& Srivastava, 2017), current finding shows that high school students with higher lifestyle depletion may not be motivated to pursue immediate rewards or seek pleasure related purchase. In addition, this different finding can be explained from the context of the survey period (summer of 2021), where COVID-19 was still prevalent. Students may have felt less compelled to pursue hedonic purchases during the global health crisis.

Findings about lifestyle depletion and financial perception show that teenagers who feel depleted view their financial situations less positively. This negative linkage is consistent with the results that Baumeister et al. (2007) reported, who revealed a significant negative relationship between depletion and financial perceptions among the general population. They stated that people who experience higher lifestyle depletion are less likely to exert self-control, tend to perceive their financial situations as tight and make poor financial decisions. The logic is that depleted people may not be motivated to find or utilize the resources to manage their finances effectively, which could result in negative perceptions about their finances. High school students with higher levels of depletion may also view their

Vol. 2023 (2) 253 – 261
ISSN 2688-3651 [online]

financial circumstances negatively due to their current situation. Also, as stated above, students may have been influenced by COVID-19 when there were negative views on economic situations worldwide (Glowacz & Schmits, 2020).

Interestingly, I found that the lifestyle depletion score was relatively higher than what has been reported in previous studies (Montoya & Scott, 2013). Given that the survey was distributed in the summer of 2021, during COVID-19, when teenagers reported increased mental health issues and depletion due to prolonged stress, it is possible that unique stressors related to COVID-19 may have affected their reporting higher depletion (Guessoum et al., 2020). This higher depletion could have influenced their perception of finances and saving-oriented attitudes.

One notable thing about the demographic variables I found was that self-rated health was significantly associated with financial perception and consumption attitudes. Students in good health perceive their finances positively, more likely to report hedonic consumption attitudes and saving attitudes simultaneously. Given that there are few studies on these topics, future studies should consider examining the influence of health on financial well-being among high school students. Also, the financial perception was positively associated with the mother's education but negatively associated with race (minority) and sibling size (large), indicating that the perceived financial situation may reflect their personal financial circumstances.

Overall, understanding how teenagers form their financial perceptions and consumption attitudes can provide valuable insights into promoting healthier financial habits and financial well-being. The current study highlights the factors contributing to highly depleted teens' financial intentions and suggests ways to promote more positive financial well-being. To improve financial behavior or decision-making, it is recommended to implement financial education in schools and communities and encourage families to provide teenagers with financial guidance (Maison & Maison, 2019). These changes can help teens develop better financial habits and create a more financially stable future.

4.1 Implication for public policy and Intervention

The study has important implications for educators and families. Studies have shown that students taking personal finance courses have a greater knowledge of finance and demonstrate responsible financial behavior, such as saving more money and using credit cards responsibly (Gutter et al., 2010). As the study findings suggest a positive effect of financial literacy on high school students, interventions focusing on financial literacy could benefit them by providing information and skills to manage their finances effectively. In addition, the study results show that intervention would be important for high school students who feel depleted due to life stress or responsibilities. Higher levels of depletion appear to shape negative perspectives on their finances but also help them consider savings' benefits. Despite the benefits of saving-oriented consumption attitudes, negative perspectives on finances may deter them from pursuing active financial activities such as investment or monitoring their budget. Thus, programs addressing lifestyle depletion can help students be better equipped to manage their stress and help depleted students maintain balanced views on their finances.

There are several limitations to this study. First, data collection was based on snowballing and convenience sampling. Thus, the generalization of the findings should be cautioned. Second, the research design was cross-sectional, meaning causality cannot be determined between variables. It is also possible that perceptions about finance and consumption attitudes may be influenced by the financial circumstances of high school students; however, I did not collect data on family household income since high school students may need access to this information. Future studies could incorporate this data. Finally, the main survey was conducted during COVID-19, a highly stressful period for most high school students. Dealing with extreme circumstances may have left them feeling more stressed out and viewing their financial circumstances more negatively than in the pre-COVID-19 time. Future studies could consider following up with these students and see how post-COVID-19 may change their views on finances and consumption attitudes.

5. Conclusion

This paper examines the effect of financial literacy and depletion on financial perception and consumption

Vol. 2023 (2) 253 – 261
ISSN 2688-3651 [online]

attitudes among high school students. Financial literacy predicted better financial perception and more saving-oriented consumption attitudes as expected. On the contrary, lifestyle depletion predicted a negative perception of financial situations and positively related to saving-oriented consumption attitudes. The potential influence of COVID-19 and students' financial circumstances was discussed in interpreting the results. Further study of this topic would benefit from household income or the financial burden that high students experience to disentangle the effect of financial literacy from students' financial circumstances. Given the significant role of financial literacy and lifestyle depletion, it would be important to consider providing valuable education for high school students who would be the main consumers in the future market.

Acknowledgment

Thank Mrs. Schieffer for time and effort that you used to help guide me in my research topic.

References

Anisa, N. A., et al. (2020). Financial literacy on impulsive buying behavior in Y generation. *Quantitative Economics and Management Studies, 1(1),* 70-75.

Aro, R., & Wilska, T. A. (2014). Standard of living, consumption norms, and perceived necessities. International *Journal of Sociology and Social Policy, 34(9/10), 710–728*. DOI:10.1108/IJSSP-06-2013-0064

Baumeister, R. F., & Vohs, K. D. (2007). Self-Regulation, ego depletion, and motivation. *Social and personality psychology compass, 1(1),* 115-128.

Baumeister, R. F., et al. (2008). Free will in consumer behavior: Self-control, ego depletion, and choice. *Journal of Consumer Psychology, 18(1)*, 4-13.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. Current directions in psychological science, 16(6), 351-355.

Cui, P., et al. (2021). The dark side of the pursuit of happiness comes from the pursuit of hedonia: The mediation of materialism and the moderation of self-control. *Current Psychology*, 1-11.

Dey, D. K., & Srivastava, A. (2017). Impulse buying intentions of young consumers from a hedonic shopping perspective. *Journal of Indian Business Research, 9(4),* 266-282.

Duffett, R. G. (2017). Influence of social media marketing communications on young consumers' attitudes. *Young Consumers, 18(1),* 19-39.

Fan, L., & Chatterjee, S. (2019). Financial socialization, financial education, and student loan debt. *Journal of Family and Economic Issues, 40(1*), 74-85.

Fariana, R. E., Surindra, B., & Arifin, Z. (2021). The influence of financial literacy, lifestyle and self-control on the consumption behavior of economic education student. International Journal of Research and Review, 8(8), 496-503.

Fox, J., Bartholomae, S., & Lee, J. (2005). Building the case for financial education. *The Journal of Consumer Affairs, 39(1), 195-214.* http://dx.doi.org/10.1111/j.1745-6606.2005.00009.x

Glowacz, F., & Schmits, E. (2020). Psychological distress during the COVID-19 lockdown: The young adults most at risk. *Psychiatry Research, 293*, 113486. https://doi.org/10.1016/j.psychres.2020.113486

Guessoum, S. B., et al. (2020). Adolescent psychiatric disorders during the COVID-19 pandemic and lockdown. *Psychiatry research, 291,* 113264. DOI: 10.1016/j.psychres.2020.113264

Vol. 2023 (2) 253 – 261
ISSN 2688-3651 [online]

Gutter, M. S., Garrison, S., & Copur, Z. (2010). Social learning opportunities and the financial behaviors of college students. *Family and Consumer Sciences Research Journal, 38(4),* 387-404.

Hamid, F. S., & Loke, Y. J. (2021). Financial literacy, money management skill and credit card repayments. *International Journal of Consumer Studies, 45(2)*, 235-247.

Jang, K., Hahn, J., & Park, H. J. (2014). Comparison of financial literacy between Korean and U.S. high school students. *International Review of Economics Education, 16*, 22-38.

Kindle, P. A. (2010). Student perceptions of financial literacy: Relevance to practice. *Journal of Social Service Research, 36(5), 470-481*. http://dx.doi.org/10.1080/01488376.2010.510951

Lusardi, A., Mitchell, O. S., & Curto, V. (2010). Financial literacy among the young. *Journal of consumer affairs, 44(2), 358-380.*

Mahapatra, M. S., & Mishra, R. K. (2019). Role of self-control and money attitude in personal financial planning. *The Indian Economic Journal, 67(3-4)*, 362-366.

Maison, D., & Maison, D. (2019). The psychological perspective in financial behaviour. *The Psychology of Financial Consumer Behavior,* 1-49.

Mandell, L., & Klein, L. S. (2009). The impact of financial literacy education on subsequent financial behavior. *Journal of Financial Counseling and Planning, 20*(1).

Montoya, D. Y., & Scott, M. L. (2013). The effect of lifestyle-based depletion on teen consumer behavior. Journal of *Public Policy & Marketing, 32(1)*, 82-96.

Peltier, J. W., Dahl, A. J., & Schibrowsky, J. E. (2016). Sequential loss of self-control: Exploring the antecedents and consequences of student credit card debt. *Journal of financial services marketing, 21*, 167-181.

Remund, D. L. (2010). Financial literacy explicated: The case for a clearer definition in an increasingly complex economy. *Journal of consumer affairs, 44(2)*, 276-295.

Sharma, G. D., et al. (Eds.). (2023). The psychological underpinnings of financial well-being (Vol. 16648714). Frontiers Media S.A.

Strack, F., & Deutsch, R. (2006). Reflective and impulsive determinants of consumer behavior. *Journal of Consumer Psychology, 16(3)*, 205-216.

Walstad, W. B., & Rebeck, K. (2017). The test of financial literacy: development and measurement characteristics. *The Journal of Economic Education, 48(2)*, 113-122.

Journal of Research High School

Vol. 2023 (2) 262 – 271
ISSN 2688-3651 [online]

Adolescent Brain Development

Anika Kumar [1] *

[1]Woodinville High School, Woodinville, WA, USA
*Corresponding Author: me@anikakumar.net

Advisor: Lee Lazar, leelazar17@g.ucla.edu

Received May 24, 2023; Revised September 5, 2023; Accepted, October 2, 2023

Abstract

This paper reviews recent research on the factors that impact adolescent brain development. Numerous research studies have analyzed the biological and psychological changes that underlie this transitional period within the lifespan. Nevertheless, many remain at a level that can be difficult for the general public to understand, causing the public to associate adolescence with only danger and risk and not providing adolescents with the opportunity for exploration. Therefore, this literature review aims to compile research on the main factors that affect adolescent development in a comprehensible paper to provide a more accurate depiction of adolescence. This paper will begin with a discussion on adolescence and a background on the current adolescent research. The paper will then be divided into six sections describing the most extensively researched and critical factors influencing the adolescent brain. This includes the effects of puberty, emotional development, risk-taking, social environments, sleep, and learning and decision-making on the developing adolescent. Finally, this paper will conclude with key findings to broaden the understanding of adolescence beyond a time of 'storm and stress' to help shift the adolescent narrative and how society can best support them by giving adolescents the space to navigate through hardships and complete their goals as they define the future.

Keywords: Adolescence, Brain development, Adolescent growth, Brain growth, Neuroscience, Psychology

1. Introduction

Over the years, 100,000 research papers have been published on adolescent development, giving more light to understanding adolescence. Adolescence is the phase marked by the start of pubertal maturation between childhood and adulthood, around ages 9 to 25 (Wierengai et al., 2017). It is a critical transitional period defined mainly by the development of essential neural networks that, together with social, emotional, and hormonal changes, set the stage for adult life. Recent studies have explored adolescents' biological and psychological changes to understand this transitional period. Studies have also focused on the effects of adolescent changes on mental health, contributing to internalizing disorders like depression and externalizing disorders like substance abuse (Crone and Dahl, 2012). However, despite increasing research on adolescents, public understanding of adolescents is declining, resulting in a negative public perception as seen through the stigmatization and discrimination against adolescents in the social world and media (Lau, A. S. et al. 2016). This paper aims to provide an accessible overview of critical factors in adolescent brain development to help facilitate society to support opportunities that will give adolescents the environment to take healthy risks and explore themselves to promote a positive transition not defined by negative consequences that can help build a stable future ("Exploration & Risk Taking", n.d.).

Vol. 2023 (2) 262 – 271
ISSN 2688-3651 [online]

2. Puberty

Puberty a phase marked by the increase in sex-specific hormonal changes that give rise to physical and biological attributes. These neurobiological changes that occur at puberty onset are critical for the social, emotional, and cognitive maturation necessary for the ability to achieve reproductive success (Piekarski et al., 2017). Before puberty, postnatal (the period before childhood) brain development causes a shift within sensitive periods in which plasticity is enhanced (Piekarski et al., 2017). Researchers denote this as the sensitive period because it is the most critical time for adolescents to acquire a particular skill rapidly. However, as puberty progresses, adolescents' plasticity declines to stabilize already learned skills that can impact our academic, cognitive, and long-term memory-keeping abilities. This is due to the changes in hormone levels that may have reorganizing effects on human brain development (Wierengai et al., 2017). According to Wierengai (2017), as our abilities become more specialized, the volume in an adolescent's brain loses its ratio of gray matter to white matter. Gray matter serves to process information in the brain to white matter, which then passes messages between different areas of gray matter within the central nervous system. During the period before puberty, gray matter reaches its peak before being encroached by white matter that increases and stabilizes into adulthood, meaning that cognitive functions and behaviors are fine-tuned (Suleiman et al., 2017).

Additionally, gonadal hormones are enhanced during puberty, which is involved in how people experience romantic love, including oxytocin (role in childbirth and nursing), vasopressin (role in maintaining blood flow to vital organs), dopamine (role in how we feel pleasure), serotonin (role in regulating our emotions), and cortisol (role in helping our body respond to stress) (Suleiman et al., 2017). The development of sex hormones causes adolescents to stimulate neural signals in their brains that affect the dopamine system. This unique reward-seeking system facilitates goal- seeking behavior in adolescence. For example, an adolescent will use this system to seek out relationships for romantic love and sex. Furthermore, changes in hormones during pubertal onset are essential to understand as they can cause more activation and sensitivity in regions of adolescents' brains that may intensify emotional reactivity (mood swings), impulsivity (fast driving), and novelty-seeking behaviors (developing relationships), as well as poorer sleep regulation. (Lee et al., 2014). Social and cultural aspects also play a vital role in framing the sexuality of youth. Parents, peers, and the media can all activate adolescents' neural reward circuitry and risk-taking behavior. Though maternal relationships in adolescents' early life play a part in adolescents' brain development, adolescents tend to communicate their stress more with their peers when faced with scenarios of anxiety due to the heavy amount of time adolescents spend with their peers going to school than their parents. Thus, the buffer of stress shifts from parents to friends, making adolescents want to be more independent from their parents and be more strong-willed against them.

3. Emotional Development

Adolescence is a time marked by profound hormonal changes that can fine-tune critical neural networks that work to produce and manage emotions. Thus, adolescence is characterized by a time of extreme sensitivity in one's emotional life to both desirous and non-desirous emotional experiences. Compared to adults and children, adolescents are known to have hyper-emotionality (the state of being extremely emotional), being happier or more depressed in response to events like talking to friends (Guyer et al., 2016). Unlike adults, adolescents' positive feelings do not last as long and generally have a more challenging time regulating their emotions, causing adolescents to experience disorders like mood swings and depression. Throughout adolescence, youth slowly develop the emerging ability to exert greater self-directed control over behaviors and emotions. According to Guyer and researchers (2016), learning to manage one's emotional reactions is a principal task of adolescence because discovering new emotional experiences, like jealousy and acceptance, can cause heightened emotions that should be managed. However, experiences of early life adversity, like separation from biological parents or exposure to violence, may stump this growth of emotional development in the brain. Two categories of early life adversities that youth experience are threat (risk of being potentially harmed) and deprivation (suffering a lack of a substantial benefit) (Hein et al., 2020). These types of adversities may change the processing of an adolescent's brain by impacting the adolescent's mental health, stress responses, and emotional processing. As reported in Hein's study, adolescents without violence exposure started with

Vol. 2023 (2) 262 – 271
ISSN 2688-3651 [online]

a higher level of amygdala activation (the stimulation of the cerebral cortex into a state of general wakefulness or attention) to a threatening image in the scanner but eventually habituated to it. In contrast, youth who experienced early life adversity started with less activation but did not habituate. Overall, violence exposure and social deprivation were related to factors of anxiety and depression. Given all these changes in the neurodevelopment of emotionality, adolescence is the critical window for the emergence of internalizing (keeping your feelings or issues inside and not sharing your concerns with others) disorders like depression, anxiety, or self-harm. Girls have been shown to be at a higher risk for the emergence of internalizing disorders 2-3 times more than boys due to their earlier puberty onset and more rapid pubertal tempo (developmental change over time) due to the fact that girls tend to have close-knit friend groups (Pfeifer et al., 2020). One out of five adolescents have a mental illness that persists into adulthood (Lee et al., 2014). According to Pfeifer et al. (2020), peer relations (including peer rejection, close friendships, and romantic or sexual relationships) significantly determine well-being and mental health in childhood and adolescence. Youths increasingly report friends as an essential source of social and emotional support as they age (Fuligni, 2019). The central relationship stressor that causes mental health problems is the experience of being rejected. Increased neural sensitivity to peer rejection naturally increases across pubertal development but is exacerbated among adolescents with depression. Though researchers know little about how brain development relates to positive aspects of social connections, like close friendships, with this data, researchers can help pinpoint and decrease the symptoms of adolescents showing the emergence of internalizing disorders and help adolescents gain a stronger sense of their emotions.

4. Risk-Taking

Adolescents display heightened attraction to adventurous and thrilling experiences despite their evident risk relative to children and adults. This is because adolescence is crucial for explorative learning, risk-taking, and sensation-seeking (Braams et al., 2015). Increased risk-taking in adolescents results from dopamine systems gaining heightened sensitivity to rewards, which leads to an overactive reward system in the brain, causing adolescents to make risky decisions. According to Chien and colleagues (2015), peer presence may encourage adolescent risk- taking by sensitizing brain regions associated with reward sensitivity (Figure 2). Greater adolescent risk-taking in peer presence can be explained by the fact that youth spend more of their social life with friends than adults. Greater risk-taking in the company of peers is consistent with a group polarization (when a group makes a more drastic decision than its members would have made if acting on their own) effect of peer influence. Chien's study demonstrated that adolescents took greater risks in a simulated driving task than adults when observed by peers.

Risky behavior reflects the contribution of two brain systems when decision-making. The first is an incentive processing system that biases decision-making based on the valuation and prediction of potential rewards and punishments. The presence of peers may sensitize this incentive system to respond to cues signaling the potential rewards of risky behavior. The second is a cognitive control system that aids goal-oriented decision-making by keeping impulses in check and providing the brain with the machinery to make alternative decisions. Thus, it is likely that youth with reasonable executive control and peer groups with similar interests will be able to experiment with risky behavior without causing severe results (Chein et al., 2011).

Although negative risk-taking has been commonly associated with adolescents, positive risk-taking is just as used. Positive risks can include trying a new sport or activity or trying public speaking for the first time. Risks can be categorized as more than just negative by assessing the benefit it has to an adolescent's well-being, the rigor of the potential consequences, and the relative social appropriateness of the behavior. Positive and negative risk- taking are disparate but are influenced by some of the same underlying characteristics (e.g., participation in team sports is associated with a higher rate of substance use). According to Duell and Steinberg (2020), sensation-seeking and reward sensitivity are associated with greater positive and negative risk-taking, showing that positive and negative risk-taking were significantly and positively correlated. Though there are many correlations between the two types of risk, there are also many differentiating factors. Harmful risk-taking was associated with the psychological aspects of impulsivity, high reward sensitivity, and punishment, while those did not seem to characterize positive risk-taking. Positive risk-taking is associated with greater school engagement, whereas negative risk-taking is associated with

Vol. 2023 (2) 262 – 271
ISSN 2688-3651 [online]

lower school engagement. There are also gender differences in risk-taking, the main difference being that adolescent girls are less likely to take adverse risks than boys, as seen in (Figure 1) (Romer et al., 2017). This may have to do with early maturation seen in girls around the start of adolescence. Not only do adolescent girls experience bodily maturation faster than adolescent boys, but they also reach a greater cognitive level. Young adults who perform fatal risk-taking tendencies show brain responses with weak cognitive control. High sensation seekers may be more inclined towards risk-taking behaviors that can lead to harmful outcomes. However, they appear to learn from these experiences as they age, compared to adolescents with impulse control problems. According to Romer's study (2017), many adolescents can be described as "hyper-rational" as they rely on the risks and benefits of their behavior even more than adults do. Therefore, adolescents are prone to more risk-taking than adults and children, though it can also have many positive outcomes.

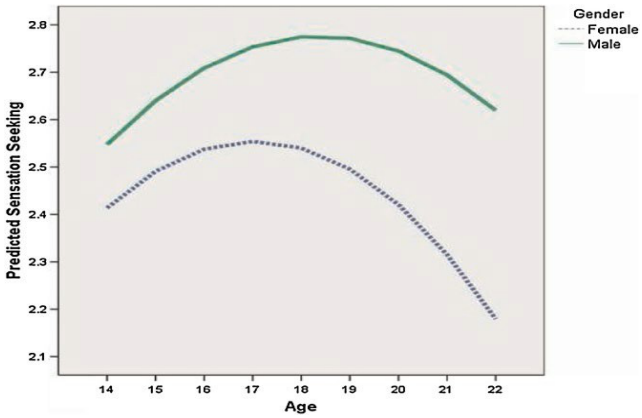


Figure 1. U.S. Survey of 1800 adolescents shows trends in sensation seeking (exhibiting heightened attraction to novel and exciting experiences despite their evident risk) from females and males in the United States. The x-axis indicates the age of participants, and the y-axis indicates the peak of sensation seeking. Image from (Romer et al., 2010).

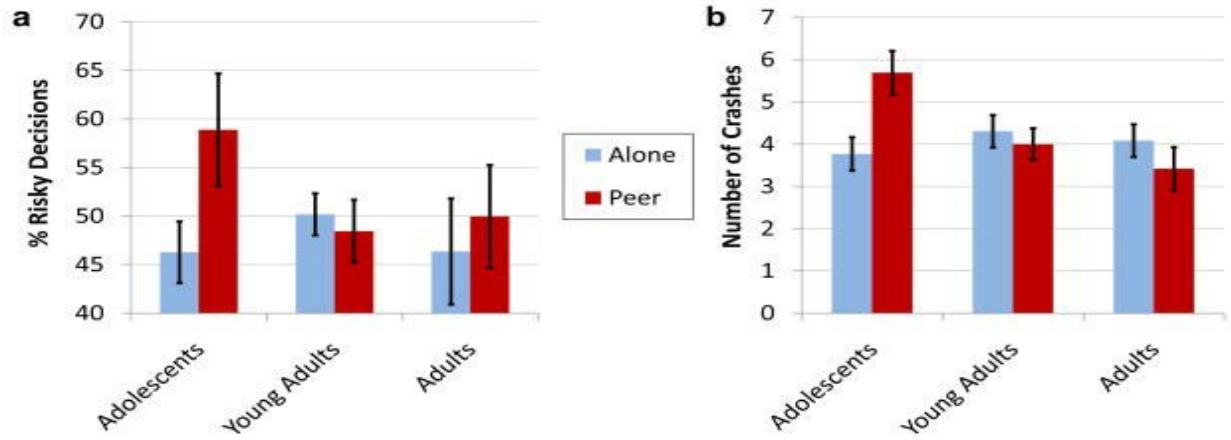


Figure 2. Car-driving scanner, in which all 40 participants from different age groups took more risk-taking decisions in the context of peers. The x-axis represents the different age groups, the y-axis in Figure 2a is the mean percent of risky behavior, and in Figure 2b is the number of crashes when playing car simulation. Error bars indicate the standard error of the mean. Image from (Chein et al., 2011).

5. Social Environment

Another major shift that occurs alongside adolescent brain development is the experience of extensive changes in adolescents' social behavior and environments. The development of the social brain, a neural network of the brain involved in social perception and cognition, allows adolescents to form more complex, hierarchical relations and have more sensitivity to the acceptance and rejection of their peers (Kilford et al., 2012). According to Kilford and colleagues (2012), adolescents can now make better sense of the world through processing signals (the understanding, analyzing, modifying, and synthesizing of signals such as sound and images) and encompassing a wide range of cognitive processes to interact with one another. This allows adolescents to be more equipped in motivational processing and decision-making. Young adults have increased activation in the posterior superior temporal sulcus (pSTS), involved in interpreting complex social gestures, and other engagement of the dorsomedial prefrontal cortex

Vol. 2023 (2) 262 – 271
ISSN 2688-3651 [online]

(dmPFC), engaged in inferring another's mental states compared to adults (Kilford et al., 2012). Thus, greater engagement of the dmPFC reflects the increased use of the social brain to understand others better, even when unnecessary. As adolescents develop an increased sensitivity to others' perceptions of themselves, they are also more prone to social anxiety than any other anxiety disorder (Kilford et al., 2012). This is because youth become increasingly self-conscious as they develop a better ability to evaluate themselves and others, causing them to believe in an imaginary audience (the phenomenon that others constantly criticize them) (van Hoorn et al., 2019). While adolescents are motivated by reward and dopamine-seeking goals such as social affiliation, many struggle with social anxiety and experience approach-avoidance due to fear of rejection or humiliation, which is associated with greater cortisol responses (Muscatell et al., 2012). Another factor that can cause social reluctance among adolescents is their carefulness when choosing their friends due to the profound influence friends may have on their performance (Kilford et al., 2012). Following the development of the social brain, adolescent decision-making often occurs in the context of peers, parents, or other critical social agents. When adolescents decide based on social context, they observe the social influence of their decision and the social outcome (van Hoorn et al., 2019). According to the van Hoorn paper (2019), the best way to make decisions is to consider the role of neural systems involved in effect (the ability to make inferences about others' emotions and feelings), cognitive (the ability to make inferences about others' thoughts and beliefs) control, and social information processing (how individuals and groups interact and establish social relationships).

As these extensive changes in adolescents' social behavior unfold alongside the development of the social brain, adolescence can also be characterized as a time when youth gain the ability to consider the needs, concerns, and perspectives of others (Fuligni, 2019). Understanding the scenarios of others better allows adolescents to feel the need to help and positively impact others. Thus, showing signs of prosocial behavior is defined as voluntary behavior intended to benefit others. Prosocial behavior can be in the form of sharing, giving, and making decisions involving outcomes for oneself and others (Güroğlu and Crone, 2014). Two main processes correlating with prosocial behavior are social cognitive perspective-taking and empathic concern (Blankenstein et al., 2020). Social cognitive perspective-taking involves understanding one's mental state by observing their social situation (Freschmann et al., 2019). While empathy focuses on understanding (cognitive empathy) and relating to one's situation (affective empathy). For an adolescent to perform prosocial behavior, they must have both of these two processes. For example, empathy can cause others distress when faced with specific scenarios (exc. a homeless man begging for money), causing them to avoid the situation rather than help. When researching prosocial behavior, researchers cannot take observations from a random group of benefactors of prosocial behavior. This is due to adolescents having favorability towards certain groups of people when making contributions (Figure 3). Researchers determine adolescents' longitudinal change in prosocial behavior by examining their behavior towards strangers, friends, and family. According to Padilla-Walker and colleagues (2018), adolescents are more likely to help

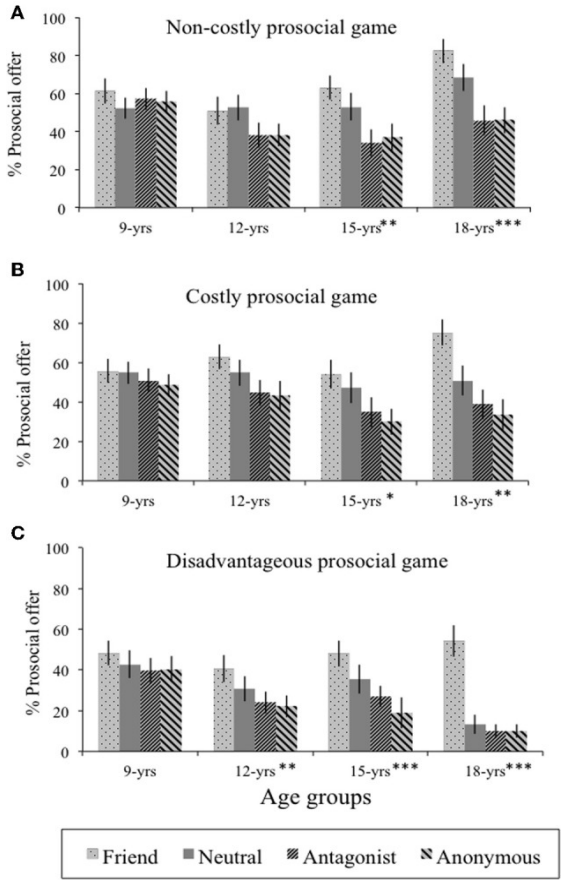


Figure 3. Assigned games to 125 participants from four age groups to evaluate favorability in prosocial behavior based on relational status. The x-axis is the different age groups, and the y-axis is the mean frequency and standard errors of prosocial offers. Each figure is distributed into A) non-costly prosocial game, B) costly prosocial game C) disadvantageous prosocial game (Güroğlu and Crone, 2014).

Vol. 2023 (2) 262 – 271
ISSN 2688-3651 [online]

someone they know (exc. friends, family) than a stranger, in large part because they are motivated to help by relationship quality and norms that are a part of relationships they have with friends and family. However, friends tend to be the common target for youth making contributions due to adolescents being surrounded by peers. When gifting, adolescents' prosocial behavior toward their family is more likely motivated by their desire to maintain the relationship rather than wholly from a moral sense of self. Prosocial behavior toward peers increased steadily among adolescents, in contrast to prosocial behavior toward the family, which was reasonably stable across adolescence and peaked. In comparison, prosocial behavior toward strangers increased across early to mid-adolescence and then flattened out during adulthood. There are sex-specific differences in prosocial behavior between girls and boys. Girls' perspective-taking increased between mid-adolescence, but boys' perspective-taking increased in late adolescence (Figure 4). Researchers suggest this may have to do with girls' faster maturation in cerebral cortical development from early adolescence to mid-adolescence, causing girls to be about two years ahead of boys in intellectual and social cognitive functioning as an adolescent. According to Van der Graaf and team (2014), boys may lack empathetic concern because of pubertal maturation that increases testosterone, inducing competitive behavior and thereby reducing empathy. As one can see, factors that affect adolescents' prosocial behaviors are dispositional (empathy, perspective-taking), relational (mother and father warmth, friend connections) behaviors, and demographics (child, gender, family structure, ethnicity).

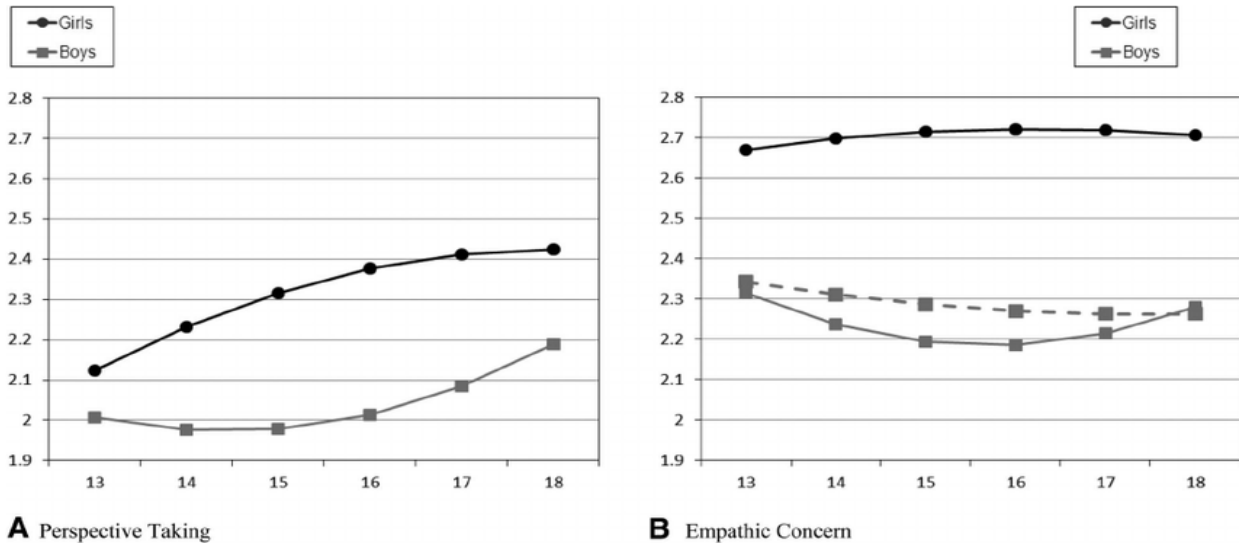


Figure 4. Development of prosocial behavioral skills between girls and boys. 497 participants were observed, with 214 of them being girls. The x-axis is the age of the participants, and the y-axis is the comparative fit index of A) perspective-taking and B) empathic concern. The solid lines represent the best-fitting models for developing A) perspective-taking and B) empathic concern. The dashed line in Figure 1B represents the boys' development of empathic concern corrected for differences in pubertal status from ages 13 to 16. (Van der Graaf et., 2014)

6. Sleep

Sleep is the body's essential cycle to rejuvenate. It is vital for memory consolidation, learning, and academic performance (Galván, 2020). The Galván et al. (2020) paper elaborates on sleep's ability to deepen cognitive sophistication, improve emotion regulation, and intensify social cognition during adolescence.

Nonetheless, sleep deprivation among adolescents is common. According to Hagenauer and Lee (2012), growing adolescents require, on average, as much as 9–10 hours of sleep per night. However, over 45% of adolescents in the United States report obtaining less than 8 hours of sleep on school nights. Inadequate sleep may amplify the neural imbalance between affective and cognitive control systems in adolescents by diminishing their ability to control their impulses while increasing their reactivity to rewarding activity (Telzer et al., 2013). Therefore, a lack of sleep can

Vol. 2023 (2) 262 – 271
ISSN 2688-3651 [online]

impair cognitive control by inhibiting activation of the dorsolateral prefrontal cortex (DLPFC), a brain region known for predicting an action's consequences (Telzer et al., 2013). According to Telzer and colleagues, adolescents who experienced inferior sleep quality have reported less self-esteem and more indifference in their behavior when making decisions and socializing. Thus, a lack of sleep can negatively impact adolescents' mental health, leading to symptoms of depression and suicidal thoughts. The root cause of adolescents' vast sleep deprivation is their tendency to stay up late. Three primary components are essential in determining the timing of sleeping and waking. They are the circadian system ("internal 24-hour clock"), homeostatic sleep pressure (drive for sleep), and other external factors (such as stress, medication, and environment) (Hagenauer and Lee, 2012). The circadian system generates daily rhythms and is entrained by external time cues (or "zeitgebers," such as sunlight) to maintain a periodicity that matches environmental rhythms, such as the 24-hour day. Due to adolescents' circadian rhythm associating with zeitgebers, it may be harder for an adolescent's body to fall asleep during the daytime when light is around than at night as the daily rhythms' main goal is to align the circadian system to the solar day. During pubertal onset, adolescents undergo biological alterations in their circadian and homeostatic systems that alter adolescents capacity to fall asleep faster (Galaván, 2020). Adolescents take longer to build up sleep pressure (the biological process that causes the brain to feel the need to sleep), which means older adolescents do not feel the need to sleep until later in the night (Figure 5).

The reduced need for sleep is one of the many reasons youth are encouraged to do activities like outdoor sports to build their sleep pressure. To help, parents can create parent-set earlier bedtimes, which not only provide a regular schedule and encourage adolescents to spend sufficient time in bed but may advance the circadian rhythm phase by reducing light exposure in the evenings and increasing light exposure in the morning (Telzer et al., 2013).

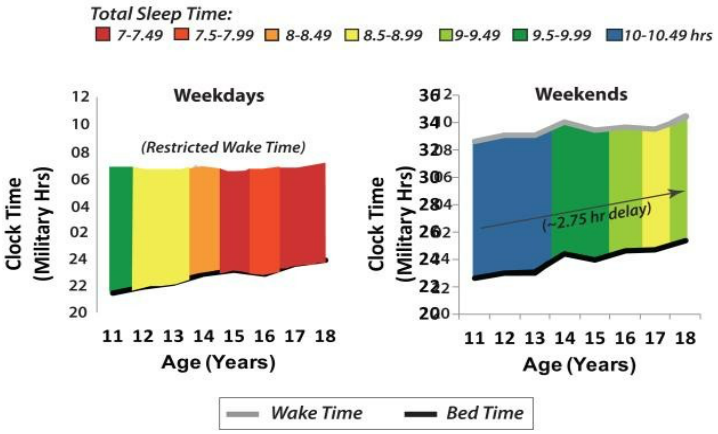


Figure 5. The trend of adolescents' timing of sleep decreases with age. The x-axis is the different countries' participants' ages in years, and the y-axis is the clock hours in military hours. The average bedtimes (black line) and wake times (gray line) are shown for both weekdays (left graph) and weekends (right graph). The total sleep time is indicated by color (key at the top of the figure). The arrow in the right graph demonstrates the change in color, indicating that adolescents were not getting the recommended sleep of 9-10 hours (Hagenauer and Lee, 2012).

7. Learning and decision-making

Adolescence is marked by a sensitive period when learning and decision-making are accelerated and emphasized. As adolescents reach adulthood, they begin to feel the burden of more responsibilities to successfully transition to complete independence as an adult. A major component in decision-making is temporal discounting, which is the process of evaluating the value of waiting for a future reward depending on the vastness of the reward and the delayed time (Sullivan-Toole et al., 2019). Temporal discounting involves three cognitive processes: valuation (the value placed on a specific outcome), cognitive control (engaging in goal-directed cognitive processes), and prospection (thinking about the future) (Anandakumar et al., 2018). With these three cognitive abilities, individuals will vary in preference for a more minor vs. significant reward and vice versa. According to Anandakumar and colleagues (2018), individual variability in temporal discounting selection could be explained by differences in functional brain organization that develop throughout adolescence, like the shifting gray matter in our brains. According to Anandakumar and colleagues (2018), the sensitivity to rewards in young adults relates to how that individual values the proposed reward or how well that individual can inhibit reflexive urges or the ability to think about the future. Anandakumar's longitudinal neuroimaging study on temporal discounting showed that there is indeed an age-related increase in the preference for waiting for the more immense rewards. On average, younger adolescents weigh delay costs heavily in their decisions, but as they grow older, they become increasingly tolerant of delays to wait for higher-

Vol. 2023 (2) 262 – 271
ISSN 2688-3651 [online]

value rewards (Sullivan-Toole et al., 2019). This suggests that age and individual differences in neural circuitry are related to an individual's preference for immediate versus delayed rewards (Anandakumar et al., 2018). Reward sensitivity is also essential in promoting learning, though a critical behavior that declines with age (Davidow et., 2016). The negative implications of reward sensitivity in adolescents have been well documented, but much less is known about the possible adaptive side for learning (Hauser et., 2015). Adolescent learning far exceeds adults' learning capabilities due to adolescence being associated with quick- changing environmental demands, which require excellent adaptive skills and high cognitive flexibility (Figure 6) (Davidow et al., 2016). According to Hauser and team (2015), youth show improved learning from reinforcement outcomes (rewarding desired behaviors and/or punishing undesired ones) and enhanced episodic memory (the ability to recall and mentally re-experience specific episodes from one's past). This suggests a solid link between reinforcement learning and episodic memory for rewarding outcomes. According to Hauser and colleagues (2015), learning from reinforcement is linked to how episodic memories are shaped and how they are seen as more good than bad. The hippocampus is a region of the brain that supports this reinforcement learning, as it is known for its role in learning and memory, particularly episodic memory, which contributes to reward-related behaviors like reinforcement learning, reward-guided motivation, and value-based decision-making. Adolescence is when potent and positive memories are formed, emphasizing the importance of learning from experiences to cement an adolescent's ability to learn from errors and make better future decisions.

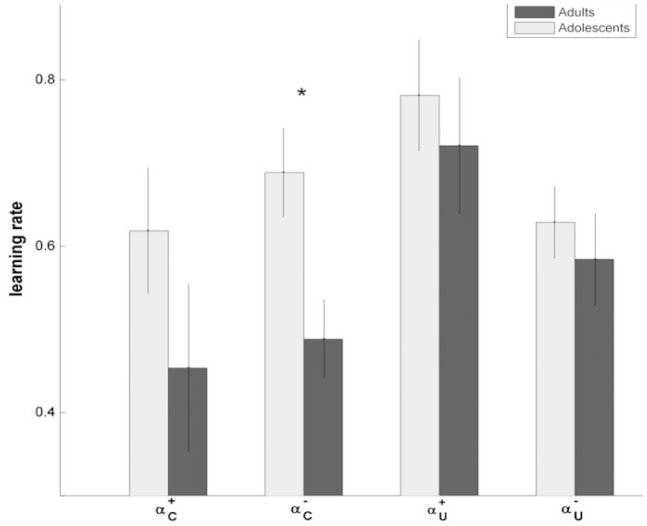


Figure 6. Differences between adolescents and adults in learning. The x-axis represents the different learning stimuli being compared between age groups of 36 participants, and the y-axis represents the rate at which the participants learned their chosen object. The asterisk indicates that $p < .05$ (data significant) to show multiple comparisons corrected (Hauser et., 2015).

8. Conclusion

This paper reviews recent research on the various factors that impact adolescent brain development. Adolescence is one of the most crucial life stages in which individuals grow and adapt to form their identities, interests, and goals. Through reviewing several studies on adolescence, it can be concluded that encouraging parents to understand the underlying reasons for their child's behavior is crucial in helping adolescents stay on the right track by providing them with a supportive and understanding environment. Additionally, adolescents who take the time to understand the social implications of their actions recognize the rationale behind their choices and make more informed decisions. Society can fundamentally transform its approach by ensuring a comprehensive understanding of adolescent developmental research to support the development of policies and programs to give adolescents the chance for healthy exploration and positive risk-taking. Further developmental neuroscience and psychology exploration could be conducted to support future youth who will impact our communities' social and economic prosperity in our nation and our world ("Exploration & Risk Taking", n.d.). Studying adolescents' physiological and neural development will challenge and change society's negative perception of adolescence and construct an optimistic trajectory for adolescents and the future.

References

Anandakumar, J., et al. (2018). Individual differences in functional brain connectivity predict temporal discounting preference in the transition to adolescence. *Developmental cognitive neuroscience*, *34*, 101-113.

Vol. 2023 (2) 262 – 271
ISSN 2688-3651 [online]

Blankenstein, N. E., et al (2020). Behavioral and neural pathways supporting the development of prosocial and risk-taking behavior across adolescence. *Child development*, *91*(3), e665-e681.

Braams, B. R., et al. (2015). Longitudinal changes in adolescent risk-taking: a comprehensive study of neural responses to rewards, pubertal development, and risk-taking behavior. *Journal of Neuroscience*, *35*(18), 7226-7238.

Casey, B. J., & Caudle, K. (2013). The teenage brain: Self control. *Current directions in psychological science*, *22*(2), 82-87.

Chein, J., et al. (2011). Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. *Developmental Science*, 14(2), F1 – F10

Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. Nature reviews neuroscience, 13(9), 636-650.

Davidow, J. Y., et al. (2016). An upside to reward sensitivity: the hippocampus supports enhanced reinforcement learning in adolescence. *Neuron*, *92*(1), 93-99.

Duell, N., & Steinberg, L. (2020). Differential correlates of positive and negative risk taking in adolescence. *Journal of youth and adolescence*, *49*(6), 1162-1178.

Ferschmann, L., et al. (2019). Prosocial behavior relates to the rate and timing of cortical thinning from adolescence to young adulthood. *Developmental cognitive neuroscience*, *40*, 100734.

Fuligni, A. J. (2019). The need to contribute during adolescence. *Perspectives on Psychological Science*, *14*(3), 331-343.

Galván, A. (2020). The need for sleep in the adolescent brain. *Trends in cognitive sciences*, *24*(1), 79-89.

Güroğlu, B., van den Bos, W., & Crone, E. A. (2014). Sharing and giving across adolescence: an experimental study examining the development of prosocial behavior. *Frontiers in Psychology*, *5*, 291.

Guyer, A. E., Silk, J. S., & Nelson, E. E. (2016). The neurobiology of the emotional adolescent: From the inside out. *Neuroscience & Biobehavioral Reviews*, *70*, 74-85.

Hagenauer, M. H., & Lee, T. M. (2012). The neuroendocrine control of the circadian system: adolescent chronotype. *Frontiers in Neuroendocrinology*, *33*(3), 211-229.

Hauser, T. U., et al. (2015). Cognitive flexibility in adolescence: neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. *Neuroimage*, *104*, 347-354.

Hein, T. C., et al. (2020). Childhood violence exposure and social deprivation are linked to adolescent threat and reward neural function. *Social cognitive and affective neuroscience*, *15*(11), 1252-1259.

Kilford, E. J., Garrett, E., & Blakemore, S. J. (2016). The development of social cognition in adolescence: An integrated perspective. Neuroscience & Biobehavioral Reviews, 70, 106-120.

Lau, A. S., et al. (2016). Adolescents' stigma attitudes toward internalizing and externalizing disorders: Cultural influences and implications for distress manifestations. Clinical psychological science : a journal of the Association for Psychological Science, 4(4), 704–717.

Lee, F. S., et al. (2014). Adolescent mental health—opportunity and obligation. *Science*, *346*(6209), 547-549.

Muscatell, K. A., et al. (2012). Social status modulates neural activity in the mentalizing network. Neuroimage, 60(3), 1771-1777.

Vol. 2023 (2) 262 – 271
ISSN 2688-3651 [online]

Padilla-Walker, L. M., Carlo, G., & Memmott-Elison, M. K. (2018). Longitudinal change in adolescents' prosocial behavior toward strangers, friends, and family. *Journal of Research on Adolescence*, *28*(3), 698-710.

Pfeifer, J. H., & Allen, N. B. (2020). Puberty initiates cascading relationships between neurodevelopmental, social, and internalizing processes across adolescence. *Biological Psychiatry*.

Piekarski, D. J., et al. (2017). Does puberty mark a transition in sensitive periods for plasticity in the associative neocortex?. *Brain research*, *1654*, 123-144.

Romer, D., Reyna, V. F., & Satterthwaite, T. D. (2017). Beyond stereotypes of adolescent risk taking: Placing the adolescent brain in developmental context. *Developmental cognitive neuroscience*, *27*, 19-34.

Suleiman, A. B., et al. (2017). Becoming a sexual being: The 'elephant in the room' of adolescent brain development. *Developmental cognitive neuroscience*, *25*, 209-220.

Sullivan-Toole, H., et al. (2019). Worth working for: The influence of effort costs on teens' choices during a novel decision making game. *Developmental cognitive neuroscience*, *37*, 100652.

Telzer, E. H., et al. (2013). The effects of poor quality sleep on brain function and risk taking in adolescence. *Neuroimage*, *71*, 275-283.

UCLA Center for the Developing Adolescent. (n.d.). Exploration & Risk Taking. https://developingadolescent.semel.ucla.edu/

Van der Graaff, J., et al. (2014). Perspective taking and empathic concern in adolescence: gender differences in developmental changes. *Developmental psychology*, *50*(3), 881.

van Hoorn, J., et al. (2019). Incorporating the social context into neurocognitive models of adolescent decision-making: A neuroimaging meta-analysis. Neuroscience & Biobehavioral Reviews, 101, 129-142.

Wierenga, L. M., et al. (2018). Unraveling age, puberty and testosterone effects on subcortical brain development across adolescence. *Psychoneuroendocrinology*, *91*, 105-114.

Journal of Research High School

Vol. 2023 (2) 272 – 279
ISSN 2688-3651 [online]

A Review: Stem Cell Therapy for Neurological Disorders

Zishi Li[1] *

[1]Palos Verdes Peninsula High School, Rolling Hills Estates, CA, USA
*Corresponding Author: zishi.li2022@gmail.com

Advisor: Dr. Hakan Coskun, Hakan.Coskun@childrens.harvard.edu

Received March 31, 2023; Revised August 30, 2023; Accepted, September 27, 2023

Abstract

The nervous system has limited self-regeneration ability, and many drugs used to treat neurological disorders often incur serious side-effects. Conventional treatment methods are often ineffective at targeting diversified pathologies of neurological diseases and breaching the blood-brain barrier. Consequently, stem cell therapy appears to be a viable option for treatment. Under right conditions, stem cells can be guided to differentiate into neurons and glial cells, providing vital regeneration of cells and tissues not possible with conventional treatment methods. Additionally, stem cell therapies are minimally invasive and less damaging to the patient's body. Stem cell transplants have been shown overall to be supportive measures of treatment and delay progression of neurodegenerative and neuromuscular diseases. Currently, mesenchymal stem cell treatments are the most widely explored as mesenchymal stem cells are relatively easy to harvest and transplant. With additional research done on optimal administration routes and microenvironmental factors that affect the efficacy of engraftment, other types of stem cells could also become viable options of treatment clinically. Yet, many difficulties must be overcome in research before stem cell therapies become safe and effective treatments. Risks associated with stem cell treatments include teratoma, tumorigenesis, and inflammation. Most significantly, transplanted stem cells incompatible with the host body could trigger immunorejection, causing grafts to fail. The review assessed levels of success achieved in research and evaluate clinical trials done on stem cell treatments of neurodegenerative disorders, neuromuscular diseases, hemorrhages, spinal cord injuries, as well as other supportive roles stem cells play in treatment.

Keywords: Cellular and molecular biology; Neurobiology; Stem cell therapy; Neurodegenerative diseases; Mesenchymal stem cells

1. Introduction

In recent years, stem cell therapy has become a new potential treatment for many previously uncurable diseases. Specifically, stem cell therapies have received significant successes in the treatment of leukemia and other types of cancers. Yet, one of the most promising potentials for stem cell therapies is in the field of neurological disorders, namely the treatment of neurodegenerative diseases such as Parkinson's and Alzheimer's disease, as well as other neurological pathologies such as spinal cord injury and multiple sclerosis. Despite many challenges in clinical studies due to poorly understood biological mechanisms of neurological disorders and stem cell differentiation post-transplant, a significant number of studies have shown that stem cell therapy is a non-invasive and moderately effective treatment of such diseases.

For patients suffering from neurological disorders, there is a wide range of stem-cell treatment options available, each harvested from a different source. Currently, the most common types of utilized stem cells are neural stem cells (NSCs), mesenchymal stem cells (MSCs), hematopoietic stem cells (HSCs), induced pluripotent stem cells (iPSCs), and embryonic stem cells (ESCs). Each option has its respective advantages and disadvantages, with MSCs transplants

Vol. 2023 (2) 272 – 279
ISSN 2688-3651 [online]

being the most widely explored currently as they could be harvested and transplanted easily. However, other options could also be considered for their effectiveness and safety. Comparing current research that has been done on different types of stem cell therapies on neurological disorders could illuminate the optimal method of treatment for each type of illnesses, as well as demonstrating areas where more research is necessary. This research has pointed out the strong advantages of specific treatments, the risks associated with certain methods, as well as the ethical concerns generally connected with the field.

2. Discussion

2.1 MSCs Overview

Currently, a majority of stem cell research done on neurological disorders involves mesenchymal stem cells (MSCs). MSCs are able to differentiate into cartilages, bones, and other skeletal tissues. Compared to other stem cell treatments, MSCs possess a number of advantages: easily harvested from the bone marrow and the umbilical cord, no requirement for genetic matching, as well as low chance of immunorejection. Taken together, MSCs transplant became an optimal option for the treatment of multiple sclerosis, spinal cord injury, and stroke. Studies have shown that with patients who received MSCs transplant have shown improvement in motor rehabilitation. Compared to standard rehabilitation therapy, MSCs therapy has shown overall better improvement in neurological functions such as pinprick, light touch, and bladder function. MSCs' neuroprotective effects make them ideal candidates for multiple sclerosis treatment. Furthermore, MSCs treatments often come with non-severe side effects such as fever, headache, and numbness, with no serious side effects such as death or tumor observed.

A large number of MSC clinical trials and research have been conducted to treat neurodegenerative diseases due to MSCs' significant paracrine effects, stimulating the body's own cells into cell and tissue regeneration. Furthermore, MSCs' immunomodulation effects have been shown to regulate neuroinflammations triggered by dysfunctional neuroresponses. In clinical trials, MSCs have been shown to secrete growth factors such as endothelial growth factors which promote angiogenesis, restoring nutrients and oxygen to the damaged brain tissues (Rahbaran, 2022). MSCs have also been shown to secrete neuroprotective factors which promote hippocampal neurogenesis. MSCs secreted exosomes have been used in exosome treatments which can reduce the effect of aß deposition and increase neuronal memory cells in Alzheimer's Disease (AD) patients (Chen, 2021). These exosomes could breach the blood-brain barrier, increasing dopamine levels and decreasing apoptosis of dopaminergic neurons in Parkinson's Disease (PD) patients (Chen, 2020). MSC exosome treatments were also shown to be able to repress glial cell functions and restore neural dysfunction in ALS mouse models.

Co-administration of MSCs with other bioactive substances were also investigated to treat neurodegenerative diseases. Administration of MSCs with resveratrol, a Sirt1 activator, allowed MSCs to engraft in the hippocampus of the AD murine model, increasing learning memory and neurogenesis (Wang, 2018).

In patients suffering from neurodegenerative diseases, routes of MSCs administration seemed to significantly affect treatment results. Intraventricularly administered MSCs have been shown to secrete microRNA which reduces nuclear kappa B expression, decreasing production of inflammatory cytokines and down-regulating cell proliferation. MicroRNA have also been shown to induce astrocyte activation and promote synaptogenesis in AD rat models (Nakano, 2020). However, in a phase I clinical trial, intracranially administered MSCs did not slow pathological development or trigger adverse side-effects in patients within a period of three months.

In animal models, MSCs injections have been shown to secrete neuroprotective factors which promote hippocampal neurogenesis, protecting neuronal tissues from oxidative stress. However, this treatment was not observed to slow disease progression within a 24-month period, and the neuroprotective effects observed in animal models were not present (Kim, 2013).

Much research has also been done on the use of MSCs to treat neuromuscular diseases such as amyotrophic lateral sclerosis (ALS) and muscular dystrophy. MSCs' paracrine effects, ability to stimulate the bodies' own cells for regeneration, make them optimal candidates for stimulating regeneration of depleted muscle tissues and neurons in patients with neuromuscular diseases. In particular, co-injection of MSCs with other substances has been investigated

Vol. 2023 (2) 272 – 279
ISSN 2688-3651 [online]

to improve engraftment, or the migration of stem cells to the desired location and start regeneration, in animal and human models. Co-administration of MSCs with cytokine reception antagonists have been shown to be able to bypass certain undesirable microenvironmental factors (Greco SJ & Rameshwar, 2008). MSC treatments' efficacy are largely varied by microenvironmental factors, as inflammatory state of recipient muscle cells vary outcome. Studies have shown that intramuscular co- injection of MSCs with macrophages can increase engraftment. One study showed that MSCs transplantation successfully restored 5% to 50% of skeletal myofibers in dogs with Golden Retriever muscular dystrophy. However, this technique failed in human trials as intra-arterial infusion led to atrial fibrillation and thalamic stroke (Sampaolesi, 2006).

Overall, human clinical trials using MSCs have yielded limited results. In nine Duchenne Muscular Dystrophy patients who received umbilical cord-MSCs transplantation, pulmonary function improved in all nine patients, yet there was no significant limb muscle strength improvement. Implantation of Cardiosphere-Derived cells and Bone Marrow-Derived cells showed no significant improvement of motor activity in patients (Aminzadeh, 2018). Optimum dosage of MSCs implantation must be determined, as there were signs that repeated transplantation of autologous MSCs could slow ALS progression (Siwek, 2020).

MSC treatments have also been used as a tool to treat hemorrhages. Hemorrhages are particularly hard to treat as they involve primary damage and secondary damage. Primary damage is the mechanical damage caused by the bursting of the blood vessel, while secondary damage is caused by the release of toxic materials such as thrombin, erythrocyte lysate, and free radicals. In research, MSC secretome showed anti-inflammatory effect in subarachnoid hemorrhage rat models by polarization of the microglia to an anti-inflammatory phenotype, reducing pro-inflammatory cytokines in the parietal complex and the hippocampus (Park, 2012). Furthermore, Adipose Tissue MSCs can stably proliferate to lower risks of apoptosis in hemorrhage rat models and increase levels of vascular endothelial growth factors and neural function score.

Some research has been done using MSCs in the treatment of spinal cord injuries. Spinal cord injuries are dangerous conditions that could cause permanent paralysis in parts of the body. These injuries are particularly difficult to treat as injuries to the central nervous system cause chemical and immune system changes, hindering attempts at regeneration.

In one research, autologous MSCs harvested from the patient's bone marrow were directly transplanted to the patients' spinal cords and re-injected using lumbar tapping. In a six-month period, six out of ten patients showed improvement in motor power of their upper limbs. Three patients showed improvements in completion of daily activities and electrophysiology and decrease in injury cavities in MRI scans. These results illustrate that MSCs could promote neuro-regeneration, stabilizing pathological developments (Han, 2021).

2.2 NSCs Overview

NSCs are multipotent stem cells that have the ability to differentiate into neurons and glial cells. When treated with different growth factors such as Fibroblast GF Growth Factors, Epidermal Growth factors, or Glial-Derived Neutrophic Growth Factors, NSCs can differentiate into neuronal cells, astrocytes, or oligodendrocytes. In current studies, NSCs are most commonly used to treat multiple sclerosis and Parkinson's Disease. In theory, they are ideal sources to replenish depleted brain cells and tissues due to neurodegenerative diseases. Yet, NSCs are often hard to harvest as they are limited to the hippocampus and usually in contact with cerebrospinal fluid. As a result, most NSCs are harvested from the fetal brain, raising ethical concerns. Currently, research is being done using induced pluripotent stem cells (iPSCs) to generate neuronal progenitor cells, which might prove to be a viable way to harvest NSCs in the future.

Most research using NSCs involve Alzheimer's Disease. AD is characterized by the degeneration of dopaminergic and cholinergic neurons or motor neurons. As AD patients often exhibit different pathological developments, cell therapy must be specifically targeted to a specific group of patients. Early phases of AD development often involve the hippocampal circuitry, making it a target for NSCs transplantation. In AD mice transplanted with NSCs, a significant increase in trophic factors (NSC secretome) was shown to decrease the level of tau phosphorylation, a key pathological characteristic of AD.

Vol. 2023 (2) 272 – 279
ISSN 2688-3651 [online]

It is key to investigate the effects of microenvironmental conditions of the central nervous system on the efficacy of NSC treatments. Environmental conditions such as the blood-brain barrier may prevent maximum engraftment of NSCs transplantations. In some cases, transplanted NSCs form cell clusters which decreases their ability to travel to the desired sites. One possible improvement to bypass undesirable factors is the application of NSCs-extracellular vehicles, nanosized particles secreted by cells that can carry soluble molecules. Research has shown that sensorimotor control was improved after EVs carried NSCs to injury sites, reducing neuronal apoptosis.

Another route of treatment targets the neuro-signaling pathways of the central nervous system. Treatment of donor cells with compounds that allow axonal extension promoted axonal outgrowth which allows extensive integration into the CNS. In AD mice, NSCs overexpressing enzyme choline acetyltransferase improved physical performance of mice population as one of the causes of AD is the reduction in acetylcholine neurotransmitter levels (Park, 2013).

NSCs' paracrine effects have also been explored to ameliorate conditions and slow pathological development of neurodegenerative disease. NSCs transplantations have been shown to decrease neuroinflammation and promote neurogenesis, synaptogenesis, and the release of neuroprotective substances (Lilja, 2015).

In the human body, endogenous NSCs are usually silenced. However, they could be activated under pathological conditions, such as the release of thrombin, erythrocyte lysate, excitatory amino acid, free radicals, or other bioactive substances caused by intracerebral hemorrhage. Subsequently, the efficacy of NSCs treatments is largely affected by factors of the central nervous system.

In patients paralyzed by spinal cord injuries, attempts have been made to manipulate NSCs to regenerate depleted neurons and cell tissues, restoring senses to the body. NSCs treated with Glial Neutrophic Growth Factors (GDNF) have been shown to differentiate to a larger extent compared to unmodified NSCs in treating patients suffering from cervical spinal cord injury.[15] Furthermore, researchers found that NSCs transplantation can decrease inflammation caused by spinal cord injuries and reduce M1 macrophage activation, achieve immunomodulation, and increase neurological performance (Cizkova, 2007).

Many studies have been done on using NSCs to promote recovery from post-spinal cord injury. In particular, one clinical trial studied differentiation of NSI-566 neural stem cells post engraftment in the treatment of spinal cord injury.

In NSI-566 studies on rat and primate models, spinal grafting has been shown to significantly improve neurological function and promote extensive axonal routing (Curtis, 2018).In a similar human study, four patients with injuries in different parts of the spine were injected with NSI-566 cells intraspinally with an observation period of 553 days, evaluating their neurophysiology, imaging, and presence of antibodies post-injection. The data has seen no significant change in quality-of-life scores in all four patients. In all four patients, diffusion tensor imaging reflected stable appearance of the injury site and the surrounding sites with no signs of decay or improvement (Cheng, 2016).

In conclusion, it is critical to understand and improve the translation between animal and human studies, increasing efficacy of certain treatments and reducing possible risks associated with other treatments (Balez, 2016).

2.3 IPSCs Overview

Induced pluripotent stem cells are relatively new in discovery. They are derived from adult somatic cells and reprogrammed back to an embryonic pluripotent state. Currently, much debate exists over the "stemness" and iPSCs, evaluating their degree of differentiation when entering the human body. Yet, one significant advantage of using iPSCs is that they do not involve the moral issues of using embryonic stem cells as they do not physically harm any human embryos.

IPSCs are ideal sources for disease modeling as they can be used to stimulate brain tissues which are hard to acquire. Neurons, oligodendrocytes, and astrocytes can all be differentiated from iPSCs. Typically, neurodegenerative diseases develop at old age, requiring elderly animal models. Yet, it is impractical and expensive to wait years for these models. One study has been able to derive iPSCs from individuals with Hutchinson-Gilford Progeria Syndrome, which possess a truncated product of the LMNA gene, triggering faster aging. These iPSCs are optimal for usage in AD or HD pathological studies.

Vol. 2023 (2) 272 – 279
ISSN 2688-3651 [online]

Studies have shown that iPSCs are able to reset the epigenome, resetting some factors to a juvenile state, reversing the aging effect of neurodegenerative pathologies. Yet, one challenge of using iPSCs to regenerate neurons in AD patients is that these newly generated neurons often exhibit the same phenotypic neuropathology such as abnormal aß levels and elevated tau phosphorylation (Krencik, 2011).

2.4 IPSCs derived Astrocytes

As the most abundant glial cell type in the central nervous system, astrocytes provide important neural support such as synaptogenesis, synaptic plasticity, metabolic support to neurons, as well as myelination of white matter. Astrocytes often exhibit different reactive phenotypes in response to various stimulus. A1 astrocytes are induced by injuries, secreting toxic molecules and leading to neurodegenerative diseases. A2 astrocytes are induced by ischemia, releasing neuroprotective cytokines, modulating neuroinflammation, and inhibiting cell apoptosis. In AD patients, astrocytes exhibited increased aß production, altered cytokine release, and dysfunctional CA2+ homeostasis.

In current studies, iPSCs have been used to differentiate into neuroepithelial cells, eventually induced into NSCs. In turn, long-time expansion of NSCs via growth factors and terminal differentiation of maturation are key steps of conversion into astrocytes (Liu, 2013). These efforts may allow extensive astrocyte studies and efficient astrocyte transplantations in neuro-deficient patients.

2.5 ESCs Overview

Embryonic stem cells are stem cells harvested from the human blastocysts with high pluripotency. Currently, a limited number of clinical trials utilized embryonic stem cells due to their risks of uncontrolled differentiation, possibly leading to tumorigenesis. Furthermore, using ESCs for treatment involves ethical concerns of destroying human embryos. What limited studies on ESCs there were yielded some result. One study showed that under certain conditions, ESCs can differentiate into ganglionic eminence-like progenitor cells, which can develop into GABAergic and cholinergic cells which can be adapted into endogenous neuronal circuits, restoring abilities of spatial memory and learning (Duncan, 2017).

3 General Comparison

MSCs, as the most widely experimented and easiest to harvest type of treatment, is significantly more reliable and efficient compared to other treatments in the treatment of neurodegenerative diseases, spinal cord injuries, and hemorrhages. NSCs have the advantage of being native to the nervous system, and under the right conditions, they could be accurately guided to achieve the desired effects of regeneration. In cases of paralysis, NSC transplantations have been shown to be effective at decreasing inflammation and promoting regrowth. With additional studies on animal models and patients, the range of application of NSCs could be widened. Yet, NSCs difficulty to acquire is a significant disadvantage. As a relatively new technology, iPSCs development is still in nascent stages, especially in the field of neurological disorders. However, it has already been shown to be an effective supportive measure that triggers multiple beneficial neurological pathways. Its accessibility and manageability excel beyond other treatments. There are relatively few treatments that involves ESC due to its ethical concerns and difficulty to harvest.

4 Ethical Considerations

Evidently, although stem cell therapies prove to be potential treatments of a variety of diseases in the future, currently, ethical concerns remain to be a significant obstacle to overcome. Firstly, many believe that using stem-cell therapy is playing-God, as the basic cell-line construction of the human body is altered. More significant ethical considerations perhaps include the creation and destruction of Embryos in the process of ESCs research. Much research involves the usage of stem cells derived from human embryos, and this process has been prohibited in many countries. In countries such as the United States, restrictions have been placed on such research as many argue,

Vol. 2023 (2) 272 – 279
ISSN 2688-3651 [online]

"deriving stem cells destroys the blastocyst, an unimplanted embryo at the sixth to eighth day of development" (Harvard Stem Cell Institute, 2021). Other concerns include payment to oocyte donors and the privacy concerns of the donors. As a result, many researchers instead create embryos for ESCs research purposes (Lo, 3009). Such research is again criticized for playing God and disrupting the natural harmony. The manipulation of all types of stem cells involve the question of informed consent in clinical trials, as the patients have to be informed of the potential risks and ethical concerns of their treatments. Ethical and legal guidelines need to be established regarding the trial of stem cells treatments and the methods of harvest.

5 Conclusion

Overall, stem cell therapies appear to be a viable option of treatment for many neurological diseases and illnesses. In particular, MSCs and NSCs transplants were shown to be promising methods of treatment for neurodegenerative diseases, neuromuscular diseases, hemorrhages, and spinal cord injuries. MSCs and NSCs have proven to be the most effective methods of treatment, and future research should focus on improving the efficacy of such treatments. Clinical studies of iPSCs and ESCs remain limited as these types of stem cells typically play supporting roles in treatment, promoting neurogenesis and revitalization of the neural network. The exact association between stem cell transplants and regeneration of neurological pathways should be studied in more details. Aside from the prospects, certain risks of stem cell therapies must be assessed before clinical trials. NSCs' cell identities must be assessed to avoid cell line switching and uncontrolled differentiation. Furthermore, karyotype analysis of pre-transplant stem cells can be performed to avoid chromosomal rearrangements and tumorigenesis. Ethical concerns should also be taken into consideration, as research should be carried out in accordance with government guidelines and with moral and ethics in mind. In conclusion, stem cell therapies are treatments with significant potential if risks and ethical concerns could be overcome.

5.1 Future Prospects

In future research, risks of stem cell therapies should be assessed, evaluating the invasiveness of transplant surgical procedures. Translation between animal studies and human trials should be studied, as many treatment methods that have been shown to be effective in animals fail to show therapeutic effects when trialed in humans. Furthermore, research on microenvironmental factors should be done to increase efficiency of treatments by increasing engraftment and allowing transplanted stem cells to travel to the desired area. In addition, to maximize engraftment and dosage of cells delivered to target areas, optimal administration routes should be determined, a route that bypasses undesirable microenvironmental conditions.

References

Aminzadeh, MA et al. Exosome-Mediated Benefits of Cell Therapy in Mouse and Human Models of Duchenne Muscular Dystrophy. *Stem Cell Reports*. 2018 Mar 13;10(3):942-955. doi: 10.1016/j.stemcr.2018.01.023. Epub 2018 Mar 1. PMID: 29478899; PMCID: PMC5918344.

Balez, R., et al. Neuroprotective effects of apigenin against inflammation, neuronal excitability and apoptosis in an induced pluripotent stem cell model of Alzheimer's disease. *Sci Rep*. 2016 Aug 12;6:31450. doi: 10.1038/srep31450. PMID: 27514990; PMCID: PMC4981845.

Chen, HX et al. Exosomes derived from mesenchymal stem cells repair a Parkinson's disease model by inducing autophagy. *Cell Death Dis*. 2020 Apr 27;11(4):288. doi: 10.1038/s41419-020-2473-5. PMID: 32341347; PMCID: PMC7184757.

Chen, Yi-An, et al. "Mesenchymal stem cell-derived exosomes ameliorate Alzheimer's disease pathology and improve cognitive deficits." *Biomedicines* 9.6 (2021): 594.

Vol. 2023 (2) 272 – 279
ISSN 2688-3651 [online]

Cheng, Z., et al. Anti- Inflammatory Mechanism of Neural Stem Cell Transplantation in Spinal Cord Injury. *Int J Mol Sci*. 2016 Aug 23;17(9):1380. doi: 10.3390/ijms17091380. PMID: 27563878; PMCID: PMC5037660.

Cizkova, D., et al. Functional recovery in rats with ischemic paraplegia after spinal grafting of human spinal stem cells. *Neuroscience*. 2007 Jun 29;147(2):546-60. doi: 10.1016/j.neuroscience.2007.02.065. Epub 2007 May 23. PMID: 17524565; PMCID: PMC3417127.

Curtis, E., et al. A First-in- Human, Phase I Study of Neural Stem Cell Transplantation for Chronic Spinal Cord Injury. *Cell Stem Cell*. 2018 Jun 1;22(6):941-950.e6. doi: 10.1016/j.stem.2018.05.014. PMID: 29859175.

Duncan, T., and Valenzuela, M. Alzheimer's disease, dementia, and stem cell therapy. *Stem Cell Res Ther*. 2017 May 12;8(1):111. doi: 10.1186/s13287-017-0567-5. PMID: 28494803; PMCID: PMC5427593.

"Examining the Ethics of Embryonic Stem Cell Research." *Harvard Stem Cell Institute* (HSCI), hsci.harvard.edu/examining-ethics-embryonic-stem-cell-research. Accessed 20 Aug. 2023.

Greco, SJ, and Rameshwar P. Microenvironmental considerations in the application of human mesenchymal stem cells in regenerative therapies. *Biologic*s. 2008 Dec;2(4):699-705. doi: 10.2147/btt.s2765. PMID: 19707450; PMCID: PMC2727904.

Han, M., et al. Mesenchymal stem cell-derived extracellular vesicles promote microglial M2 polarization after subarachnoid hemorrhage in rats and involve the AMPK/NF-κB signaling pathway. *Biomed Pharmacother*. 2021 Jan;133:111048. doi: 10.1016/j.biopha.2020.111048. Epub 2020 Nov 28. PMID: 33378955.

Khazaei, M., et al. GDNF rescues the fate of neural progenitor grafts by attenuating Notch signals in the injured spinal cord in rodents. *Sci Transl Med*. 2020 Jan 8;12(525):eaau3538. doi: 10.1126/scitranslmed.aau3538. PMID: 31915299.

Kim, KS et al. Long-term immunomodulatory effect of amniotic stem cells in an Alzheimer's disease model. *Neurobiol Aging*. 2013 Oct;34(10):2408-20. doi: 10.1016/j.neurobiolaging.2013.03.029. Epub 2013 Apr 24. PMID: 23623603.

Krencik, R., Zhang SC. Directed differentiation of functional astroglial subtypes from human pluripotent stem cells. *Nat Protoc*. 2011 Oct 13;6(11):1710-7. doi: 10.1038/nprot.2011.405. PMID: 22011653; PMCID: PMC3198813.

Lilja, A., et al. Neural Stem Cell Transplant-Induced Effect on Neurogenesis and Cognition in Alzheimer Tg2576 Mice Is Inhibited by Concomitant Treatment with Amyloid-Lowering or Cholinergic α7 Nicotinic Receptor Drugs. *Neural Plast*. 2015;2015:370432. doi: 10.1155/2015/370432. Epub 2015 Jul 15. PMID: 26257960; PMCID: PMC4518185.

Liu, Y., et al. Medial ganglionic eminence-like cells derived from human embryonic stem cells correct learning and memory deficits. *Nat Biotechnol*. 2013 May;31(5):440-7. doi: 10.1038/nbt.2565. Epub 2013 Apr 21. PMID: 23604284; PMCID: PMC3711863.

Lo, B., Parham L. Ethical issues in stem cell research. *Endocr* Rev. 2009 May;30(3):204-13. doi: 10.1210/er.2008-0031. Epub 2009 Apr 14. PMID: 19366754; PMCID: PMC2726839.

Nakano, M., et al. Bone marrow-derived mesenchymal stem cells improve cognitive impairment in an Alzheimer's disease model by increasing the expression of microRNA-146a in hippocampus. *Sci Rep*. 2020 Jul 1;10(1):10772. doi: 10.1038/s41598-020-67460-1. PMID: 32612165; PMCID: PMC7330036.

Park, D., et al. Improvement of cognitive function and physical activity of aging mice by human neural stem cells over-expressing choline acetyltransferase. *Neurobiol Aging*. 2013 Nov;34(11):2639-46. doi: 10.1016/j.neurobiolaging.2013.04.026. Epub 2013 May 31. PMID: 23731954.

Vol. 2023 (2) 272 – 279
ISSN 2688-3651 [online]

Park, JH et al. Long-term results of spinal cord injury therapy using mesenchymal stem cells derived from bone marrow in humans. *Neurosurgery.* 2012 May;70(5):1238-47; discussion 1247. doi: 10.1227/NEU.0b013e31824387f9. PMID: 22127044.

Rahbaran, M., et al. Therapeutic utility of mesenchymal stromal cell (MSC)-based approaches in chronic neurodegeneration: a glimpse into underlying mechanisms, current status, and prospects. *Cell Mol Biol Lett* 27, 56 (2022).

Sampaolesi, M., et al. Mesoangioblast stem cells ameliorate muscle function in dystrophic dogs. Nature. 2006 Nov 30;444(7119):574-9. doi: 10.1038/nature05282. Epub 2006 Nov 15. Erratum in: *Nature.* 2013 Feb 28;494(7438):506. Erratum in: Nature. 2014 Mar 13;507(7491):262. PMID: 17108972.

Siwek, T., et al. Repeat Administration of Bone Marrow-Derived Mesenchymal Stem Cells for Treatment of Amyotrophic Lateral Sclerosis. *Med Sci Monit.* 2020 Dec 10;26:e927484. doi: 10.12659/MSM.927484. PMID: 33301428; PMCID: PMC7737405.

Wang, X., et al. Resveratrol promotes hUC-MSCs engraftment and neural repair in a mouse model of Alzheimer's disease. *Behav Brain Res.* 2018 Feb 26;339:297-304. doi: 10.1016/j.bbr.2017.10.032. Epub 2017 Nov 2. PMID: 29102593; PMCID: PMC5729114.

Journal of Research High School

Vol. 2023 (2) 280 – 285
ISSN 2688-3651 [online]

Closed Loop Wearable Device for Parkinson's Tremor Monitor and Suppression

Aryan Ganesh

[1]Christopher High School, Gilroy, CA
*Corresponding Author: aryanganesh.aiml@gmail.com

Advisor: Gyan Krishna Sreekar, gyankrishna06@gmail.com

Received March 21, 2023; Revised August 21, 2023; Accepted, September 21, 2023

Abstract

Parkinson's disease does not have a permanent cure, medication is one of the primary methods used to manage tremors. However, the process of managing medication dosages can involve a lot of trial-and-error approaches, and tremors can still appear at different times of the day, requiring further dosage adjustments. Monitoring tremor patterns in Parkinson's patients is very important for accurate dosage planning. In this study, a simple, compact, and affordable sensor based closed-loop wearable device was developed to monitor and control tremor for assisting Parkinson's patients. The system detected tremors using an accelerometer and displays tremor data on a LED-display. This tremor data was read by a micro-controller prior and then preprocessed using moving average of the tremor intensity. The tremor count was then displayed on the LED-display and a dashboard using Adafruit cloud platform. The device was tested on a Parkinson's patient to measure tremor parameters such as tremor acceleration and tremor count. The system was tested fifteen times on a Parkinson's patient and was able to detect the tremors successfully every time it was tested. This system also includes a tremor suppression capability by applying a counterforce whenever the tremor count went past threshold count. In order to control the tremors, the threshold count is determined to be 27 counts. However, the repeatability of the amount of counterforce and tremor suppression needs to be improved.

Keywords: Parkinson's, Wearable device, Tremor monitor

1. Introduction

Parkinson's disease (PD) is a complex neurological disorder that involves both motor and non-motor symptoms resulting from the progressive loss of neurons in the brain. Unfortunately, there is currently no known cure for this degenerative disease, and only symptomatic treatment is available. Despite extensive research, the underlying cause of PD remains unknown (Dexter et al., 2013; Jose et al., 2010; Hughes et al., 2002). Gender appears to play a significant role in the frequency and severity of PD symptoms, although there are some inconsistent findings in the literature. Women with PD tend to present with a milder phenotype than men. Tremors are a common symptom of PD, often beginning in a limb and gradually spreading throughout the body (Pahuja et al., 2016; Heller et al., 2014).

Tremor refers to an involuntary and rhythmic oscillation of a body part, which can result in jerking movements or convulsions throughout the body. A common example of this is the hand tremor, which can make it challenging to grasp objects such as a glass of water or a pen for writing (Chinta et al., 2005; Park et al., 2009). There are two main categories of tremor: resting tremor and action tremor. Resting tremor occurs when the muscles are at rest and there is no intentional movement, while action tremor involves shaking of a body part during a voluntary act of moving that body part (Agarwal and Biagioni, 2022).

Although Parkinson's disease does not have a permanent cure, medication and surgery are two methods used to manage tremors. However, the process of managing medication dosages can involve a lot of trial-and-error

Vol. 2023 (2) 280 – 285
ISSN 2688-3651 [online]

approaches, and tremors can still appear at different times of the day, requiring further dosage adjustments. Additionally, caregivers of patients who have undergone Deep Brain Stimulation surgery may rely on their intuition when re-tuning the brain pacemaker, which is an unscientific approach. To address these challenges, there is a need to scientifically measure the tremor pattern so that more accurate dosage planning, and brain pacemaker re-tuning can be performed. This would help to improve the overall management of Parkinson's disease and enhance the quality of life for those affected by the condition.

Traditionally, tremor has been evaluated through physical examinations during clinical appointments. However, this approach can be subjective and may not capture the full spectrum of the symptom in a patient's daily life (Sigcha et al., 2021). In recent years, studies have been conducted on building systems for tremor suppression. However, these studies have typically used surface electromyography (EMG) burst signals from muscles, which can be prone to errors due to electrode positioning, changes in skin conductance, and crosstalk from other muscles.

To address these issues, needle EMG has been identified as the most reliable technique for precise characterization of tremor features. However, this method is invasive and costly (Xu et al., 2016). Closed-loop tremor suppression systems offer a more advanced alternative to conventional open-loop systems. They are able to deliver stimulation only when symptoms are present, providing greater comfort for patients (Khan et al., 2021).

Therefore, this research studied a simple, compact, and affordable inertial sensor based closed loop wearable device for tremor monitor and suppression system to assist Parkinson's patients can be developed.

2. Material and methods:

2.1 Materials

The components used in this study were NodeMCU, MPU6050, breadboard, LED display and vibration motors.

- NodeMCU: NodeMCU is an open-source development board and firmware based on the popular ESP8266 WiFi module. NodeMCU is one of the most popular IoT development platforms and is widely used to build IoT projects. NodeMCU also provides a rich set of libraries for interacting with different components. NodeMCU is a great platform for building connected devices. It is a great platform for prototyping projects and for developing embedded systems. NodeMCU is easy to set up and use, and it is compatible with a wide range of hardware and software. It is also open source and can be extended with custom code. NodeMCU is an excellent choice for anyone looking to build IoT projects.
- MPU6050: The MPU6050 is an inertial measurement unit (IMU). It combines a 3-axis gyroscope, a 3-axis accelerometer, and a Digital Motion Processor (DMP) all in a single chip. The device can be used to measure both linear acceleration and angular velocity.
- Breadboard: A breadboard is a construction base for prototyping of electronics. It is a reusable solderless device for testing circuit designs before committing to a printed circuit board. It is a great tool for prototyping and experimenting with circuit designs.
- Vibration motors: This study used mini vibration motors rated DC 3V current 85mA, and speed of 12000 RPM. The size is 10mm x 3mm/ 0.39" x 0.12" (D*T) with self-adhesive capability.

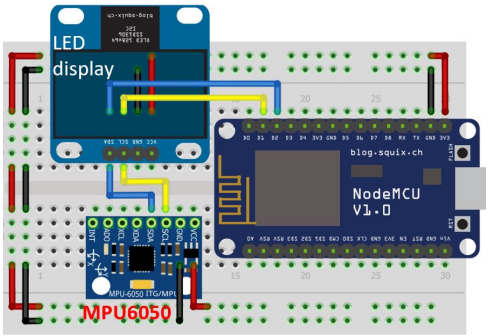


Figure 1: The design of the tremor monitoring system showing major components like NodeMCU, accelerometer and LED display.

2.2 Working principle

This study involved building a portable and wearable device based on inertial sensors (accelerometers and gyroscopes), which

Vol. 2023 (2) 280 – 285
ISSN 2688-3651 [online]

were interfaced with smartphones or tablets through wireless communication protocols (Bluetooth, Wi-Fi, etc.). The design for the tremor monitoring system included NodeMCU, accelerometer and LED display (Figure 1). The tremor suppression system added vibration motor and a few other components like transistors. NodeMCU is an open-source development board and firmware based on the popular ESP8266 Wi-Fi module. The NodeMCU was also used to read the data from the accelerometer and transmit it to a cloud platform after preprocessing. The system was programmed to detect tremors at certain thresholds and record the tremor data. If the tremor intensity crosses a certain limit the system applied a counter force to suppress the tremors. The system included a display to show the readings from the accelerometer and the current status of the tremor suppression system. The system was also programmed to keep track of the number of tremors over a period, thus allowing for monitoring of the patient's condition. After the tremor monitoring system was built, it was hooked up to a laptop to pre-process the raw data. To determine the tremor threshold to detect a tremor vibration was attached to a Parkinson's patient's wrist (Figure 2).

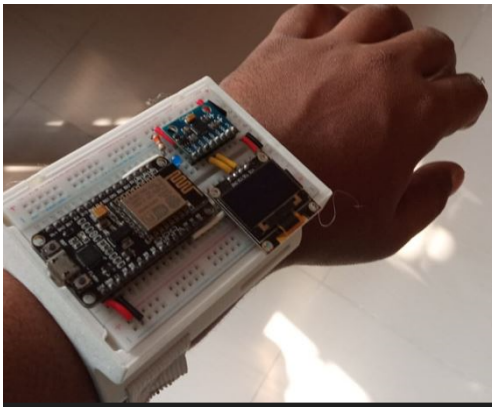


Figure 2: Example of a tremor monitor system hooked up to a patient's wrist.

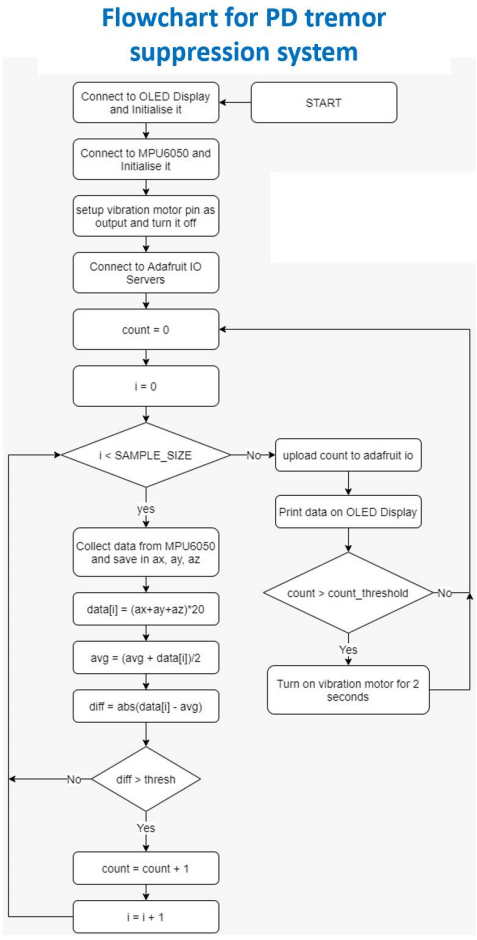


Figure 3: The flowchart showing the steps involved for the tremor suppression system. The left part of the flow chart shows the flow for curve smoothening and tremor count estimation. The right part shows the uploading the tremor count on the cloud (Adafruit) and LED display and also the tremor suppression system flow.

2.3 Methods

The data pre-processing for this study is shows in the left part of the flowchart (Figure 3). This study involved obtaining raw data of the vibrations (changes in acceleration) in all the three-directions. For simplicity the sum of all accelerations is measured as a parameter. Since the actual acceleration was small number it was multiplied by a random number 20 for making it easy to work with.

3. Results:

3.1 Data pre-processing

The movement of the hand is monitored for 4 to 5seconds. The acceleration data was affected by the position of the hand, so moving average was used to smoothen the data (Figure 4). Then the difference between the current data and the moving average was monitored as tremor intensity. This study used an Adafruit cloud platform to save the data and plotted on the dashboard.

3.2 Tremor monitoring:

The tremors were monitored (for a duration of 4 to 5 seconds), and acceleration was plotted for 10 runs using my system. A threshold of 15 m/s^2 was obtained that works for all the

Vol. 2023 (2) 280 – 285
ISSN 2688-3651 [online]

10 experimental runs (Figure 5). Then the test was run 15 times to see if the system was able to pick up the tremor vibrations correctly. My system was able to successfully detect the tremors as soon as the tremor threshold is above 15 m/s^2. We arrived with a threshold acceleration of 15m/s2 as this is the only whole number that separates the tremors and non-tremors (random hand movements) for all the 10 runs performed in this test.

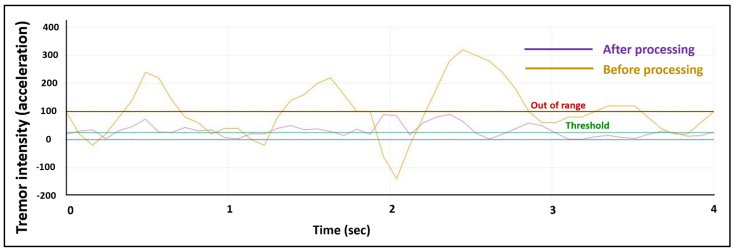


Figure 4: Plot showing the raw data of the tremors before and after smoothening the data.

To calculate the count threshold to determine if a tremor vibration is Parkinson's tremor or not, the device was attached to a Parkinson's patient's wrist again. Ten tremor runs were executed and plotted the data using my system. The number of tremors above 15 m/s^2 in a duration of 2 seconds were counted. The average of the counts for the 10 runs were taken. A count threshold of 27 counts (with a standard deviation of 2.1 counts) was obtained to determine if a particular tremor vibration is a Parkinson's tremor. Then the test was run 15 times to see if my system is able to pick up the Parkinson's tremor vibrations correctly. My system was able to successfully detect the Parkinson's tremors if the tremor count threshold is above 27 counts. The tremor count data is also tracked and plotted on a dashboard (Figure 6).

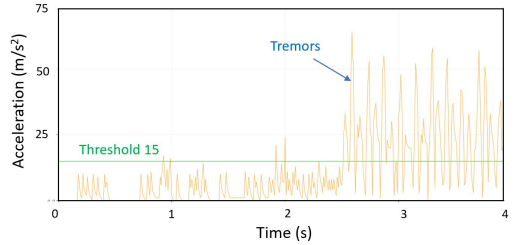


Figure 5: Plot of acceleration versus time showing the tremor threshold as 15 m/s^2

3.3 Tremor suppression:

After the tremor suppression system was built, it was tested on a Parkinson's patient. The test was run 15 times to see if my system was able to activate the vibrational motors as soon as the tremor count threshold is above 27 counts indicating that it i a Parkinson's tremor. My system was able to successfully turn on the vibrational motors whenever the tremor count threshold was above 27 counts. The vibrational motors help reduce the tremors. If the number of tremor counts is greater than 27 and if the acceleration of the tremors is >15 m/s^2 then two vibrational motor turns on using a transistor (connected to the controller). After the 2x motor are turned on if the acceleration of the tremors is still >15 m/s^2 then 2x more vibrational motor turns on (Figure 7). When the tremor acceleration comes below 15 m/s2 the motors are turned off. The vibration motors help reduce the acceleration but doesn't help reduce the tremor count frequency.

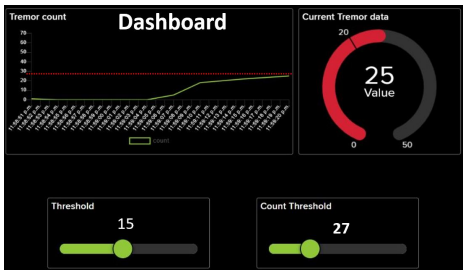


Figure 6: Dashboard tracking the tremor count data and showing the threshold and count threshold values.

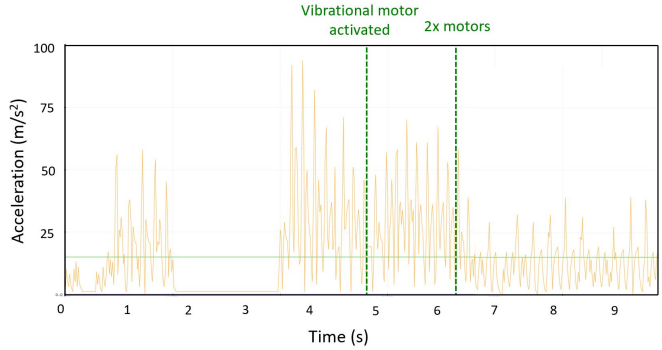


Figure 7: Plot of acceleration versus time showing the tremor acceleration going down when the first vibration motor is turned on. After the 1st motor is turned on, the acceleration of the tremors was still >15 m/s^2 then the second vibrational motor turns on to reduce the acceleration down to <15m/s^2.

4. Discussion:

The total cost of my system was less than $40 which is very affordable. The itemized cost of the components in my device includes a breadboard (~$4), an accelerometer (MPU6050 ~$9), a micro-

Vol. 2023 (2) 280 – 285
ISSN 2688-3651 [online]

controller (~$7), watch base (~$10), a watch strap (~$3) and a protective plastic cover (~$4). The software and cloud used for programing and storing the data are free for the study however these may add cost if the solution should be scaled up.

The tremor frequency for the Parkinson's patient was approximately 1 tremor every 100ms. The accelerometer in the MPU6050 measures the acceleration every 50ms and sent the data to the node-MCU (controller). The software/code in the controller runs the threshold check (Figure 5) every 50ms. This ran for 40 loops which took approximately 2 seconds. The number of tremors higher than the acceleration of 15 m/s^2 in these 2 seconds were tracked and plotted on the dashboard. If this count was greater than 27 counts, then it was flagged as a Parkinson's tremor. This method of determining if the tremors are Parkinson's is not an ideal method but works well because frequency of the Parkinson's tremors are much higher than the normal random hand movements.

Our tremor monitoring system can assist neurologists during dosage planning by providing. Based on the tremor acceleration and tremor count the care takers can adjust the medication dosage accordingly. The vibration motors generated anti-vibrations to compensate the tremor by actively applying a counterforce (equal and opposite forces) whenever the tremor count is exceeded 27 counts. The tremor suppression system was tested 15 times randomly; however, it only cancelled the tremors only four out of the fifteen times. The vibrational motors were installed on the backside of the wearable device that contacts the top of the wrist (Figure 6). If the contact of these motors with the hand was not good, then the impact on the tremors was less. I am currently working on an improved design to have a better contact of the motors with the hand.

Previous studies developed a wearable orthosis for tremor assessment and suppression using sensors that measure rotational motions around the joints. This device, which is placed parallel to the upper limb, suppresses tremor with a tremor suppression rate of 40% (Rocón et al., 2007). However, this device wasn't very popular because it was too large and bulky (Manto et al., 2007).

There are some drawbacks for my system and needs to incorporate the improvements in the future work. The tremor threshold and count threshold were calibrated on only one patient, so it needs to be tested on multiple patients to generate robust values. The vibration motors in my wearable device needs a better design to improve the contact between the hand and the motors. The actual counteracting mechanism of the vibration motors needs to be improved. The method of determining if the tremors are Parkinson's, needed to be improved.

5. Conclusions

A simple, compact, and affordable accelerometer sensor based closed loop wearable device for tremor suppression system to assist Parkinson's patients was built. The system was able to detect the tremors on a Parkinson's patient successfully all 15 times it was tested. The device generates anti-vibrations to cancel the tremors by activating the vibration motors whenever the count threshold was greater than 27 counts however the repeatability needs to be improved. The tremor count data is shown on the LED display and tracked/plotted on a dashboard for easy access.

References

Dexter, D., & Jenner, P. (2013). Parkinson disease: From pathology to molecular disease mechanisms. *Free Radic Biol Med., 62*, 132–144.

Jose, O. et al. (2010). Missing pieces in the Parkinson's disease puzzle. *Nature medicine, 16* (6), 653-661.

Hughes, A. et al. (2002). The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain, 125*, 861–870. doi:10.1093/brain/awf080

Pahuja, G., & Nagabhushan, T. (2016). A novel GA-ELM approach for Parkinson's disease detection using brain structural T1-weighted MRI data. *Cognitive Computing and Information Processing (CCIP), Second International Conference.*

Vol. 2023 (2) 280 – 285
ISSN 2688-3651 [online]

Heller, J. et al. (2014). Evidence for gender differences in cognition, emotion and quality of life in Parkinson's disease. *Aging Dis.. 5*, 63–75.

Chinta, S., & Andersen, J. (2005). Dopaminergic neurons. The international journal of biochemistry & cell biology, 37 (5), 942-946.

Park, H., and Kim, H. (2009). Electrophysiologic Assessments of Involuntary Movements: Tremor and Myoclonus. *Journal of Movement Disorders, 2* (1), 14-17. doi: 10.14802/jmd.09004

Agarwal, S., & Biagioni, M. (2022). Essential Tremor. *Treasure Island (FL): StatPearls Publishing*, PMID: 29763162.

Sigcha, L. et al. (2021). Automatic Resting Tremor Assessment in Parkinson's Disease Using Smartwatches and Multitask Convolutional Neural Networks. *Sensors, 4* (21), 291. doi: 10.3390/s21010291.

Xu, F. et al. (2016). Development of a closed-loop system for tremor suppression in patients with Parkinson's disease. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA*. 1782-1785, doi: 10.1109/EMBC.2016.7591063.

Khan, M. et al. (2021). Wearable Low-power Closed-loop System for Tremor Detection and Stimulation using Electromyography (EMG). *28th IEEE International Conference on Electronics, Circuits, and Systems (ICECS), Dubai, United Arab Emirates*, 1-4, doi: 10.1109/ICECS53924.2021.9665556.

Rocón, E. et al. (2007). Design and Validation of a Rehabilitation Robotic Exoskeleton for Tremor Assessment and Suppression. *IEEE Trans. Neural Syst. Rehabil. Eng., 15*, 367–378. doi: 10.1109/TNSRE.2007.903917

Manto, M. et al. (2007). Evaluation of a wearable orthosis and an associated algorithm for tremor suppression. *Physiol. Meas., 28*, 415–425. doi: 10.1088/0967-3334/28/4/007

Journal of Research High School

Vol. 2023 (2) 286 – 293
ISSN 2688-3651 [online]

The Emergence of Cultural Pan-Asian Identity

Samantha J. Lee[1] *

[1]Chino Hills High School, Chino Hills, CA, USA
*Corresponding Author: sammyjlee290@gmail.com

Advisor: Dr. Hannah Cho, choh@elac.edu

Received June 22, 2023; Revised September 15, 2023; Accepted, October 14, 2023

Abstract

This paper focused on the emergence of cultural or ethnicized pan-Asian identity. While politically based pan-Asian identity during the Asian American movement in the 1960s has been well documented, a cultural pan-Asian identity has newly emerged in the recent years as it hinges on the shared common culture among numerous Asian Americans, especially among second and later generations. The shift of the character of pan-Asian identity suggests that the identity of a racially minority group is bound by racializing ascriptions that a group boundary along racial lines continues to be viable.

Keywords: Pan-Asianism, Pan-Asian identity, Asian American, Racial identity, Ethnicity, Assimilation

1. Introduction

Pan-Asian identity, namely "Asian American" or its consciousness, was first forged during the late 1960s as the Asian American movement emerged. Prompted by the Civil Rights Movement in the 1960s, Asian American activists articulated various Asian groups and created the category "Asian American" for a common goal of achieving group interests. While this identity has been widely recognized as a political category since its emergence, over time it has yielded voluntary associations in which various Asian Americans interact and socialize among themselves at a personal level. Although this phenomenon is largely based on the East Asian American experience and its organizational principle largely remains rooted in similar phenotypes, they successfully articulate and move beyond their ethnic-specific identity and created a meaningful and valid group identity and institutions. This experience is sharply contrasted to the earlier dominance of the assimilationist paradigm of European immigrant children in the U.S. Assimilationists believe in the inevitability of eventual disappearance of distinctive ethnic values as a separate entity (Alba, 1990; Herberg, 1955; Waters, 1992).The immigrant children adopted unitary social system of the host culture and became absorbed into the network of social, political, and cultural values that are blinded to ethnicity. However, for the members of the racial minority like Asian Americans, such a theory becomes inapplicable to understand their experiences. Despite their acculturation and participation in various American institutions, they have created their own new cultural boundary. The purpose of this paper is to explore how Asian Americans negotiated and created a new ethnic boundary in the U.S. How would the acculturation of Asian Americans, who acquired the value and behavior of Anglo-Americans and yet learned to accept their standing as a racial minority, be projected in their identity formation? With this sociological question, I will delve into three voluntary based social arenas – college campuses, Asian inter-ethnic marriages, and religious institutions – in which the cultural/ethnic pan-Asian identity is salient.Because this paper addresses more of the private dimension of pan-Asian identity rather than the public aspect and deals with a contemporary phenomenon and context rather than a historical one, the support for my argument will rely mainly on ethnographic/sociological surveys and interviews from the research of other scholars. Based on the research, it is argued that pan-Asian identity has been more than a political label and has developed into a meaningful

Vol. 2023 (2) 286 – 293
ISSN 2688-3651 [online]

identity to many Asian Americans based on a shared feeling of common culture, which is not based on a definitional or fixed characteristic of culture, rather on a constructive and fluid nature.

2. A Paradigm Shift of Pan-Asian Identity

Subsequent to the emergence of the pan-Asian movement in the 1960s, pan-Asian identity or "Asian American" has been largely employed as a political category, which was well documented by various scholars (Aguilar-San Juan,1994; Chang, 2001; Espiritu, 1992; Lien, 2001, Okamoto, 2006). By the 1970s, the term "Asian American" began to gain a common usage not only by government, but also by various professionals and academics, creating Asian American studies programs on university campuses, and publications. While the term is widely recognized by government officials and in professional fields, some scholars view with skepticism the long-term meaning and significance of pan-Asianism (Shinagawa and Pang, 1996). The primary reason for such a skepticism is due to the fact that a pan-Asian identity which emerged out of the movement is viewed as merely a product of the political process, which was only active when common material concerns or political interests were raised (Shinagawa and Pang, 1996). Such a view is also implied by other scholarly research which predominantly has dealt with a politicized pan-Asian identity (Espiritu,1992; Junn and Masuoka, 2008; Lott, 1998). In fact, some pan-ethnic theorists exclusively identify the nature of pan-ethnicity as a political construct for the purpose of establishing a representative category in response to government policies and external threats (Espiritu, 1992; Junn & Masuoka, 2008; Nakano, 2013; Okamoto, 2014). It is true that with American Indians, Hispanics, and Asian Americans, the concept of "pan-ethnicity" first emerged in the fight for group rights and racial pride; therefore, a political cause was always predominant (Anderson, 1991; Cornell, 1990; Nagel, 1995; Padilla, 1985). If pan-Asian identity is solely defined as a representative category for political or socioeconomic reasons as some suggest, it would be difficult to assess the long-term significance of pan-Asianism or pan-Asian identity at the present since only a situational and inconsistent political pan-Asian identity is periodically formed.

On the one hand, a politically constructed pan-Asian identity seems short-lived and only emerges sporadically; on the other hand, pan-Asian identity, based on shared cultural bonds between various Asian American groups, has developed as a personally recognized identity since the 1990s. Recent studies reveal that an "ethnic" pan-Asian identity rather than a "political" pan-Asian identity is most commonly shared by American-born second and later generations of Asian Americans. In contrast to the pan-Asian identity which is generated by outward, social, or political causes, an ethnic pan-Asian identity is formed internally as a common feeling of a shared culture and of a common experience as part of a distinct group which is shared with acculturated Asian Americans (Chong, 2015).

This phenomenon challenges the traditional assimilation theories found in the U.S. The second and later generations of immigrant experience have traditionally been understood in the U.S. as a linear trajectory of assimilation into the dominant society as ethnic values and practices slowly recede. Assimilation theory assumes that American born generations eventually will put aside ethnic allegiances and ethnic institutional involvement as they desire integration into the host society. Contrary to this view, the second and later generations of Asian American experience reflect a much more complex picture of "assimilation" rather than the simplistic pattern once used to depict European immigrant descendents. Because of the lack of race factors in traditional assimilation theories, it is inadequate to describe the experience of people of color. For Asian Americans, like other people of color, race has been a key factor in understanding the context by which they negotiate affiliations and identity in the U.S. Although the majority of American born Asians are highly acculturated and educated, their status as a racially defined group continues to be viewed by the mainstream society as different and foreign (Tuan, 1998). Instead of following a linear and simplistic assimilation pattern, or maintaining an ethnic specific boundary, the second and later generations of Asian Americans have creatively negotiated and remade their ethnic boundary in a "third space," which reflects their unique experiences and struggles as a racialized minority in the U.S. This ethnic pan-Asian affiliation and identity is emerging within a third space, and a recognizable number of Asian Americans are drawn to other Asian Americans who share a similar experience and find comfort in this new ethnic niche (Kim, 2010)

Vol. 2023 (2) 286 – 293
ISSN 2688-3651 [online]

3. The Third Space

3.1 Pan-Asian Friendship and Organizations on College Campuses

As the number of Asian American students in U.S. colleges has been overrepresented, college campuses have been important places for pan-Asian interactions. According to National Center for Education Statistics, from 2010 to 2021, the number of Asian Americans enrolled in colleges marked 64 to 60 percent, compared to 43 to 38 percent of White students. As the Asian American population in colleges continues to be dominant on campuses, more pan-Asian organizations have been offered in order to provide fellowship and resources. According to Don T. Nakanishi, director of the UCLA Asian American Studies Center, sixty-five Asian American student organizations are pan-Asian (Onishi, 1996). Asian-American clubs on campuses are organized increasingly around race rather than specific ethnicity. Nazli Kibria (2002) interviewed sixty-four Korean and Chinese Americans who revealed that they had encountered and accepted a pan-Asian identity either by joining an on-campus pan-Asian organization or a pan-Asian social circle. As one-third of her interviewees had been actively involved in a pan-Asian organization during their college years, these organizations provided a crucial point through which they adopted an Asian American identity. While these organizations provided both political and social activities, the majority of her interviewees reported that they were more interested in the social aspect. In their socialization along racial lines, they experienced a greater sense of social ease and receptivity among Asian Americans, which become a comfortable social niche as they coped with a new college environment. Kibria describes this "comfortable social niche" with three factors: common family values, generational conflict between immigrants and the second generation, and a racialized or stereotyped image of the Asian.

Kibria's further research, which focused on Asian American college students who were not affiliated with pan-Asian organizations on campuses, also demonstrated the importance of the Asian American social circle in informal ways on campuses. Despite their disinterest in pan-Asian organization affiliation, and a tendency to avoid pan-Asian activities on campus, these students encountered the idea of "Asian American" one way or another so that eventually they were able to reflect the meaning and significance in terms of identity and community. For example, when these students were misidentified as other Asian descendant other than their own, or racially motivated crimes occurred such as in the Vincent Chin case, they inevitably encounter the concept of "Asian American" knowing that they cannot escape their racialized identity as a result of "racial lumping." Regardless of their pan-Asian organization affiliation, many Asian Americans realized the possibility that the concept of Asian American was not just an identity imposed from the outside but a self-claimed identity that can be embraced. In one way or another, Asian American college students encounter the pan-Asian circle and identity in their college years.

3.2 Asian Inter-ethnic Marriage

Traditionally, in the research of inter-group relations, ethnic and racial intermarriage has been an important area of sociological research. While racial and ethnic identities play an important role in maintaining group boundaries, intermarriage challenges the norms of difference and social distance (Bogardus, 1959; Gordon, 1964). In this sense, the pattern of Asian American intermarriage can provide an important clue about possible ethnic or racial consciousness shifts.

Sharon Lee and Marilyn Fernandez (1998) report trends in Asian American racial/ethnic intermarriage based on a comparison of U.S. census data from the 1980s and 1990s. According to Lee and Fernandez, the common assumption about Asian American intermarriage was that it was expected to be low due to "the effects of racial prejudice, social distance, immigration, settlement patterns, and population growth." Asian Americans were perceived as non-whites and a subordinate group in the U.S. racial hierarchy, and anti-miscegenation laws were enacted which forbade marriage between whites and people of color, including Asians. It was expected that Asian intermarriage would remain low. Even after immigration reform in 1965, the influx of new Asian immigrants was expected to cause a low intermarriage rate among Asian Americans since the first generation of immigrants were more closely tied to their

Vol. 2023 (2) 286 – 293
ISSN 2688-3651 [online]

culture and ethnicity. However, contrary to expectations, the research on Asian American intermarriage has reported high levels of intermarriage, with ranges from 25 percent to over 50 percent depending on the ethnic group.

Lee and Fernandez compared census data from the 1980s with census data from the 1990s and noted three findings. First, the overall Asian American intermarriage rate dropped from 25 to 15 percent. Second, while the overall rate decreased, Asian inter-ethnic marriage, i.e., marriage between Asian ethnic groups, has increased from 11 to 21 percent. Third, native born Asian Americans are more likely to outmarry outside of their ethnic group compared with those who are foreign born. American born Asian men are more than four times likely to outmarry than foreign born Asian men, and American born Asian women a little less than two times likely compared to foreign born Asian women. Overall, American born Asians are substantially more likely to outmarry than the foreign born according to the census data from both the 1980s and 1990s.

Based on their research findings, Lee and Fernandez conclude the following important points: If marital distance is a measure of social distance, the large increase in Asian American inter-ethnic marriages suggests the breakdown of social distance among different Asian American groups. They speculate that demographic changes and other social processes, including a growing sense of pan-ethnic Asian identity, may have contributed to the rise in Asian inter-ethnic marriages.

Larry Shinagawa and Gin Yong Pang (1996) also conducted studies on Asian American intermarriage based on data from the 1980s and 1990s. Their research focused on two areas: U.S. and California Asian American marriage patterns from 1980 to 1990. Their findings are similar to those of Lee and Fernandez: on the national level, interethnic marriages rose 400 to 500 percent from 1980 to 1990. In contrast, interracial marriages with the white population substantially decreased. California data also shows a similar pattern that interethnic marriages of Asian American men residing in California increased from 21.1 percent to 64 percent, and for women from 10.8 percent to 45.5 percent. While their data shows findings that are similar to those of Lee and Fernandez (1998), Shinagawa and Pang attempt to interpret such results in historical and sociological contexts. Their sociological analysis leads to five speculations that may contribute to the growth of interethnic marriage among Asian Americans. First, the case of Vincent Chin, who was a victim of anti-Asian violence in 1982 and served as a catalyst who revived a sense of kinship; empathy toward other Asian groups increased interethnic interaction that resulted in the increase of pan-Asian interethnic marriages more than interracial marriages. Second, as the numbers in the Asian American population increased, Asian American social institutions and businesses also multiplied. With the increased social institutions and business tractions, interethnic interaction and social networking among Asian Americans became frequent. Third, socioeconomic attainment among Asian Americans caused them to move from ethnic enclaves into mixed Asian and white suburbs. Fourth, the acculturation of Asian Americans, including using a common English language and sharing American cultural experiences, promote affinity. Lastly, in a multiracial, multicultural society, a growing sense of comfort and familiarity with each other increased by sharing "symbols, foods, behavioral traits, and beliefs about family, community and common histories."

By providing an in-depth analysis of marriage patterns among Asian Americans from the sociological settings, Shinagawa and Pang demonstrate the recent trend of heightened pan-Asian consciousness among Asian Americans.

3.3 The Emergence of Pan-Asian Churches

> The Los Angeles Times on March 8, 1999, reported on emerging pan-Asian churches:
> Some Southland congregations break the mold, courting other ethnic groups rather than staying insular to preserve culture. Scholars see a blueprint for the region's future. Newsong and churches like it are becoming the first truly pan-Asian churches in the country, drawing a mix of second-, third- and even fourth-generation Chinese, Japanese, Vietnamese, and other Asians, whose Americanized upbringing and Christian faith bind ethnically diverse backgrounds.

Newsong has been one of the prominent and successful cases of the pan-Asian church that has branched out to a number of sister churches throughout Southern California. Other representative cases of a pan-Asian church in Southern California include the Garden Christian Fellowship in West Los Angeles, and Evergreen Church in San Gabriel and Rosemead, both of which have been models for a pan-Asian church.

Vol. 2023 (2) 286 – 293
ISSN 2688-3651 [online]

Russell Jeung researches pan-Asian churches in the San Francisco Bay Area, and also reports about burgeoning new congregations (2004). According to his research, pan-Asian churches in the Bay Area increased from one in 1989 to five in 1993. By 1998 an estimated two hundred churches in the area had congregations with a majority of Chinese or Japanese American members. Twenty-two of these were specifically identified as pan-Asian, accounting for 10 percent of the ethnic Asian churches in this region.

While California, where Asians are densely populated, has been the most viable area for pan-Asian congregations, this phenomenon concurrently rises in other parts of the U.S. On "Asian American Parachurch Ministries" website that introduces pan-Asian churches in the U.S., ninety-nine churches are listed under this rubric.

The emergence of pan-Asian churches and parachurch ministries implies a complex interplay between religion and racial identity in the U.S. It indicates that racial dynamics in the U.S. continue to play an important role in the lives of Asian Americans. Such an emergence problematizes Will Herberg's idea of the "triple melting pot" which proposes assimilation based on three religious traditions: Protestant, Catholic, and Jewish. Herberg theorizes that second generation European immigrants abandoned their ethnic ties and customs in order to assimilate into the mainstream society. Herberg asserts that by the third generation immigrants have fully adapted to "the American way of life," while religious divisions continue to exist. In this transitional period, ethnic identity is replaced by religious identity (1955). Contrary to his theory, following their immigrant parents, Asian Americans continue to use religion as an important venue through which they secure both ethnic and religious identities without subsuming one or conflicting with each other.

While the emergence of pan-Asian congregations challenges the theoretical conception of the relationship between religion and ethnic/racial identity, Jeung attempts to articulate three contextual factors that have attributed to the birth of pan-Asian congregations. First, it may be caused by demographic shifts within the Asian American community and their churches. Especially, the Japanese American churches have a shrinking ethnic base since the rate of Japanese immigration is low; the churches needed to widen their target beyond their own ethnic group. Second, generational conflicts within ethnic churches eventually encourage American born Asians to create their own churches. To be free from any cultural baggage, American born Asians tend to choose a pan-Asian church. Third, pan-Asian congregations had already adopted either an Asian American identity or a pan-Asian social connection prior to a church affiliation.

These three factors were summed up by conducting extensive interviews with forty-four Asian American ministers; Jeung believes that the pastors' philosophy in leading their churches, including determining their target group, should play an important role in the birth of pan-Asian churches, because "they are the leading agents of church organizational change, entrepreneurs for their institutions, and cultural narrators of ethnic and racial discourse" (2004). Jeung views the role of ministers, rather than that of congregations, as an explanation for the emergence of the pan-Asian church.

4. The "Common Culture" of Asian Americans

As seen above, campus social organizations, intra-Asian marriage, and religious affiliations such as Protestant pan-Asian churches, have been formed in part through the pan-ethnic consciousness of Asian Americans, which is clear evidence of ethnic pan-Asianism. A common theme that runs through these expressed aspects of society was the shared feelings of a "common culture" among different Asian groups. Scholars such as Kibria, Jeung, and Mia Tuan, who have done research on Americanized Asians, predominantly attribute such phenomena to increased feelings of a shared "common culture." While some use the term "common culture" very loosely, some attempt to articulate what this "common culture" is comprised of, and it is shared by different ethnic Asian Americans.

4.1 Shared Values: Education, Family, Work

By "common culture," many Asian Americans refer to similar values that they share with other Asians. Although these values are not exclusive to Asian Americans, three values are mainly expressed: education, family, and work ethic, by which they reflect their worldview. These values which are held by many Asian Americans are often

Vol. 2023 (2) 286 – 293
ISSN 2688-3651 [online]

explained by a reference to the cultural traditions influenced by Confucianism. These common values are also reinforced and further emphasized by the conditions of Asian immigration. Immigrant parents constantly remind their children of the importance and value of education and of getting good grades as a means of bettering their situation in America. For immigrant parents, educational achievement is emphasized as a key element in the accommodation strategy for their children (Endo,1980). From an early age, immigrant children are reminded of their parents' sacrifice so that they may get a good education. Concurrently, this sense of "common culture" is also shared due to the perception of the white mainstream as the model minority (Park, 2008). The image of the "model minority," albeit an imposed identity, is often accepted by many Asian Americans because of the emphasis on education and work ethics.

4.2 Common Socioeconomic Status

The emphasis on the value of education and a hard work ethic may have led to the growing similarity in the socioeconomic attainment of Asian Americans. Indeed, Asian American socioeconomic achievement has received increased attention in scholarly research. While scholars vary on the causes of the growing similarity in socioeconomic attainment among Asian Americans, education seems to be the primary factor that is highlighted in the research (Barringer, et al.,1980).

In Jeung's interview with pan-Asian church ministers, ministers report that "congregations are remarkably similar in their professional status, their upwardly mobile families, and spare-time activities" (2004). In their similar socioeconomic status, they tend to share common hobbies or leisure activities through which they increase their associations and affinities. Kibria's ethnographic study also reflects middle class upwardly mobile characteristics among her interviewees, and these features certainly play a role in creating a common bond. In connection to this notion, Vijay Prashad argues that the term "Asian American" carries connotations of bourgeois status, while working class Asians such as the Vietnamese, the Hmong "boat-people," Chinese American sweatshop workers, or the South Asian taxi driers and kiosk workers who do not use the term "Asian American" to refer to themselves (1998).

4.3 Racialized Experience in the U.S.

Even though many Asian Americans may have obtained middle class status, this status does not necessarily translate into social acceptance in the larger society. Asian Americans recognize the similarities which link their experiences. Importantly, these similarities are grounded in a common experience of being treated as a distinct racial group in the U.S. Kibria also believes that a shared culture and worldview are fundamentally about a racialized experience characterized by such things as subtle remarks, racial ignorance, and cultural stereotypes. She argues that "Second-generation Chinese and Korean Americans believe that a unique 'Asian American' culture and community emerged out of a central experience and identification with the United States, one that those who had not grown up elsewhere did not fundamentally share" (2002).

All of these sum up to two important points: First, Asian Americans have developed a strong racial consciousness that, for them, race becomes an underlying factor that affects a worldview which is expressed at both a conscious and subconscious level. Second, as a result of strong racial consciousness, the group boundary is shifting. Since race, more than ethnicity, shapes experience and worldview in American society, the group or social boundary is increasingly shaped panethnically. Therefore, pan-Asian or Asian American identity is bound by racializing ascriptions. However, this identity should not be solely understood as a racial category that is ascribed from the outside, but also as a self-defined category based on a common "Asian culture" and friendship.

5. Conclusion: Pan-Asian Identity- Racial or Ethnic Identity?

The pan-Asian identity as a cultural form in recent years has emerged in a complex dynamic between ethnic and racial consciousness. While ethnicity and race are separate entities, they are closely intertwined and mutually reinforce each other, especially among non-white races in the U.S. A consensus exists where pan-Asian identity or "Asian American" identity is largely referred to as a racial identity, while an ethnic-American category, such as Japanese-

Vol. 2023 (2) 286 – 293
ISSN 2688-3651 [online]

American or Korean-American, highlights an ethnic identity. However, for racialized ethnics, the boundary between these two categories is often ambiguous. These two identities crosscut and mutually reinforce or sustain each other, blurring the boundary of the two. For example, Tuan's interview with forty-eight Chinese and Japanese American born descendents reveals that their ethnic identity (e.g., Chinese-American, Japanese-American) was emphasized in a situation in which they were racially grouped together as "Asian" by outsiders, despite the fact that most of them only maintain a minimum affinity with their ethnic values and culture. In this case, it is racial consciousness that reinforces the sense of ethnic identity. However, Kibria's participants report that this specific ethnic identity merely signifies a generic Asian identity rather than a specific ethnonational one, since "in the face of homogenizing racial processes, distinctions of nationality and ethnicity become irrelevant" (2002). In contrast, white Americans, seen as colorless or racially neutral, have the privilege of ethnic options separately from their racial identity (Alba, 1990; Waters, 1992). Ethnic identity for white Americans is not something that affects their lives as a racialized ethnic group. In fact, Asian Americans, especially later generations of immigrants, exercise a great deal of personal choice in cultural values and practices just like their white counterparts. Most of them have chosen to retain very little. Although Asian Americans may retain ethnicity only symbolically in their lives, their racial identity forces them to adopt their ethnicity or ethnic identity.

The symbolic ethnicity of Asian Americans is characterized as Asian American culture in general since their distinct cultural patterns of their specific ethnicity has been watered down through the acculturation. Within this group consciousness, Asian Americans as a collective group have developed their own culture, institutions, and associations, creating a new social boundary as pan-Asian or a hybrid "ethnic" group.

References

Anderson B. 1991. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.

Aguilar-San Juan, K. (1994). *The state of Asian America: Activism and resistance in the 1990s*. Boston, MA: South End Press.

Alba, R. (1990). *Ethnicity in America: The Transformation of White America*. Yale University Press.

Barringer, H., Takechi, D., & Xenos, P. (1990). Education, Occupational Prestige, and Income of Asian Americans. *Sociology of Education*, *63*(1), 27-43. doi:10.2307/2112895

Asian American Parachurch Ministries (n.d.). (Retrieved Oct. 3, 2022) from http://djchuang.com

Bogardus, E. S. (1959). *Social Distance*. Los Angeles: Antioch.

Chang, G. H. (2001). *Asian Americans and politics: Perspectives, experiences, prospects.* Stanford University Press.

Chong, K. H. (2015). "Asianness" under construction: The contours and negotiation of panethnic identity/culture among interethnically married Asian Americans. *Journal of Asian American Studies*, 18(3), 221-246. doi:10.1353/jaas.2015.0031

Cornell S. 1990. Land, labour and group formation: blacks and Indians in the United States. *Ethnic Racial Studies,* 13:368–88. doi: 10.1080/01419870.1990.9993678

Endo, R. (1980). Asian Americans and Higher Education. *Phylon*, *41*(4), 370-371. doi:10.2307/274861

Espiritu, Y. (1992). *Asian American Panethnicity: Bridging Institutions and Identities*. Temple University Press.

Gordon, M. M. (1964). *Assimilation in American Life: The Role of Race, Religion, and National Origins*. New York: Oxford University Press.

Herberg, W. (1955). *Protestant, Catholic, Jew*. Anchor Press.

Vol. 2023 (2) 286 – 293
ISSN 2688-3651 [online]

Hirschman, C., & Wong, M. (1981). Trends in Socioeconomic Achievement among Immigrant and Native-Born Asian-Americans, 1960-1976. *The Sociological Quarterly, 22*(4), 495-514. doi:10.1111/j.1533-8525.1981.tb00677.x

Hirschman, C., & Wong, M. (1984). Socioeconomic Gains of Asian Americans, Blacks, and Hispanics: 1960-1976. *American Journal of Sociology*, *90*(3), 584-607. doi:10.1086/228116

Jeung, R. (2004). *Faithful generations: Race and new Asian American churches.* Rutgers University Press.

Junn, J., & Masuoka, N. (2008). Asian American Identity: Shared Racial Status and Political Context. *Perspectives on Politics*, 6(4), 729-740. doi:10.1017/S1537592708082042

Kibria, N. (2002). *Becoming Asian American: Second-Generation Chinese and Korean American Identities.* Baltimore: Johns Hopkins University Press.

Kim, S. (2010). *A faith of our own: Second-generation spirituality in Korean American churches*. Rutgers University Press.

Lee, S. M., & Fernandez, M. (1998). Trends in Asian American Racial/Ethnic Intermarriage: A Comparison of 1980 and 1990 Census Data. *Sociological Perspectives*, 41(2), 323–342.

Lien, P. (2001). *The making of Asian America through political participation*. Philadelphia: Temple University Press.

Lott, J. T. (1998). *Asian Americans: From Racial Category to Multiple Identities*. India: AltaMira Press.

Nagel, J. (1995). American Indian ethnic renewal: Politics and the resurgence of identity. *American Sociological Review*, 60(6), 947–965. Doi: 10.2307/2096434

National Center for Education Statistics. (2023). *College Enrollment Rates*, 2023.

Okamoto, D. G. (2006). Institutional panethnicity: Boundary formation in Asian-American organizing. *Social Forces*, 85(1), 1-25.

Okamoto, D. G., & Mora, G. C. (2014). Panethnicity. *Annual Review of Sociology*, 40(1), 219-239. https://doi.org/10.1146/annurev-soc-071913-043201

Onishi, N. (1996, May 30). Merging identity - New sense of race arises among Asian-Americans. *New York Times.*

Padilla, F. (1985). *Latino consciousness*. University of Notre Dame Press.

Park, J. Z. (2008). Second-Generation Asian American Pan-Ethnic Identity: Pluralized Meanings of a Racial Label. *Sociological Perspectives*, *51*(3), 541–561. https://doi.org/10.1525/sop.2008.51.3.541

Prashad, V. (1998). Crafting Solidarities. In L. Shankar & R. Srikanth (Eds.), *A Part, Yet Apart: South Asians in Asian America* (pp. 1-22). Philadelphia: Temple University Press.

Shinagawa, L. H. & Pang, G. Y. (1996) "Asian American Panethnicity and Intermarriage". *AmerasiaJournal*, 22:2, 127152, DOI: 10.17953/amer.22.2.x270073668v30918

Tran, T. (1999, March 8). Pan-Asian Churches Emerging. Los Angeles Times.

Tuan, M. (1998). Forever Foreigners or Honorary Whites? The Asian Ethnic Experience Today. New Brunswick: Rutgers University Press.

Waters, M. C. (1992). The Construction of a Symbolic Ethnicity: Suburban White Ethnics in the 1980s. In M. D'Innocenzo & J. Sirefman (Eds.), *Immigration and Ethnicity: American Society – 'Melting Pot' or 'Salad Bowl'?* Westport: Greenwood Press.

Journal of Research High School

Vol. 2023 (2) 294 – 303
ISSN 2688-3651 [online]

Transcriptomics of Cotton Lines with Contrasting Drought Stress Response

Shreyan Sai Kancharla,[1] * Arjun Sai Kancharla[1]

[1]Enloe Magnet High School, Raleigh, NC, USA
*Corresponding Author: shreyansk6@gmail.com

Advisor: Dr. Anita Samuga, anitasamuga@gmail.com

Received July 20, 2023; Revised September 12, 2023; Accepted, October 25, 2023

Abstract

Abiotic stresses affect cotton growth, yield, and development accounting for ~50% of yield reduction around the world. Drought is one of the most important abiotic stresses, however the molecular mechanism and the genes involved in drought tolerance or sensitivity have yet to be completely understood. Cotton is most widely used in the textile industry and is a source of edible oil. For commercial applications, high-yielding cotton that is tolerant to drought stress is highly valuable, and understanding the molecular mechanism would unlock the path to creating higher-yielding cotton that is stress tolerant. Breeding drought-tolerant varieties also leads to sustainable cotton production. In this research, genes that are differentially expressed were identified between two drought-tolerant and two drought-sensitive lines. These results provide a deeper understanding of the genes that play a crucial role in regulation, adaptation to drought stress, and potential targets for molecular breeding.

Keywords: Cotton, Transcriptomics, Drought, RNA sequencing

1. Introduction

Upland cotton (*Gossypium hirsutum*) is an economically important crop accounting for 90% of the overall cotton production (Yuan et al., 2018). It is an important source of fiber for the textile industry and cotton seeds serve as a source of edible oil with the by-products serving as cattle feed such as the meal and hulls. The United States produces about 35% of the cotton globally and is the largest exporter of cotton. Cotton is grown in diverse climatic conditions around the world and is subjected to frequent unfavorable abiotic stresses such as drought, salinity, and heat. According to reports, 57% of cotton worldwide is cultivated in water-deficit areas (World Resources Institute)

Drought is one of the major abiotic factors limiting agricultural production impacting the growth, yield, and fiber quality of cotton (Parida et al., 2007). Generally, drought stress severely restricts cotton growth and development, such as affecting plant height, leaf dry weight, stem dry weight, leaf area index, node number, fiber quality, canopy and root development (Loka et al., 2011). Drought occurs frequently during the boll-forming period in many cotton production regions, and cotton is most sensitive to water stress during this period (Loka and Oosterhuis, 2012). The frequency and intensity of drought will increase with the projected changes in global climate in the future (Giorgi and Lionello, 2008). In order to have sustainable cotton production, it is thus necessary to understand the impacts of drought on cotton and obtain varieties that are drought tolerant hence preserving or increasing yield. The ideal crop is one that requires less water but produces higher yields and better fiber quality. One way to do this is through breeding resistant varieties which is quite challenging due to the narrow genetic base of cultivated cotton species and sensitivity to various stresses. Another way is to understand the molecular mechanisms of drought's effect on cotton growth and modify the gene targets to get the desired phenotype.

Drought tolerance is a complex phenomenon because it is a multigenic system that is related to various morpho-physiological, biochemical, and molecular processes (Ullah et al., 2017). Drought induces plant responses including

Vol. 2023 (2) 294 – 303
ISSN 2688-3651 [online]

but not limited to altered gene expression patterns, accumulation of metabolites such as abscisic acid (ABA), and synthesis of specific proteins, namely hydrophilic proteins, reactive oxygen scavenging proteins, and chaperones (Ghodke et al., 2020). Drought stress causes stomata closure, which leads to decreased CO2 intake, affecting the rate of photosynthesis and consequently, reducing growth and yield (Chaves et al., 2003, 2009). Physiological impacts include an increase in photorespiration, oxidative damage to the chloroplast, and obstruction of adenosine triphosphate (ATP) synthesis that make the plant drought sensitive. Drought-tolerant plants typically increase the reactive oxygen scavenging species or accumulate sugars, amino acids, alkaloids, polyols, and inorganic ions for osmotic adjustment and in turn maintain photosynthesis and other drought-related processes leading to maintaining and improving yield.

Plants including cotton respond or adapt to various stresses, and transcriptional modulation is one of the most important ways that induce or repress gene expression in plants under biotic and abiotic stresses. Transcriptomics is the genome-wide identification of gene expression and RNA sequencing enables the identification of transcripts for the expression of all genes in a single organism under precise physiological conditions or treatments, facilitating global analysis of gene functions and structures. There have been studies published on molecular mechanisms of stress response in diverse cotton species primarily using transcriptomics which facilitated the identification of key genes to be differentially expressed under stress. In *Gossypium hirsutum*, root transcriptomics indicated 1530 transcripts to be co-expressed in natural rain-fed and well-watered (WW) cotton in the field. In wild cotton *Gossypium darwanii*, RNA sequencing of leaf tissue identified 58,961 genes differentially expressed between seedlings with and without drought stress (Xu et al., 2022). RNA sequencing enabled the identification of 6968 transcripts that were differentially expressed with statistical significance in three cotton species *Gossypium hirsutum*, *Gossypium arboreum*, and *Gossypium barbadense* (Hasan et al., 2019). Upland cotton was studied in a hydroponic system with polyethylene glycol (PEG) to simulate drought stress to conduct network analysis and identify key pathways in stress adaptation (Zheng et al., 2022).

In this study, two drought-tolerant (DT) and two drought-sensitive (DS) lines from *Gossypium hirsutum* were selected to be analyzed by RNA sequencing. The lines were selected based on preliminary research indicating differences in yield measurements under water-limiting and well-watered conditions. The plants were subjected to two levels of stress: moderate drought (MD) and severe drought (SD) involving different levels of water limitation. The study also included a recovery period to assess how the plants adapt to water availability after the deficit period. This study aided in the identification of transcripts and shed light on the molecular processes likely involved in drought tolerance of cotton and contributing to yield.

2. Material and Methods

2.1 Genotypes

Four genotypes were chosen based on preliminary research done internal to the lab in water-limited and full irrigation conditions (unpublished data). Performance of the lines based on percent yield estimate difference in water-limited compared to well-watered conditions were considered for line selection. Lines 1179 and 1896 showed a mean of 10% increase yield estimate in water-limited conditions and lines 1189 and 3031 showed a mean of 40% reduction in yield in water-limited conditions. 1179 and 1896 were then chosen as the drought tolerant and 1189 and 3031 as the drought susceptible genotypes for this study.

2.2 Plant growth

Cotton seeds were potted in 4.5-inch pots with soil consisting of peat moss and vermiculite mixture, grown at 24°C and 45% relative humidity, and watered twice a week (Figure 1). Medium-sized pots were selected to ensure that the water limiting conditions were uniform throughout the pot, which would not be possible in smaller pots. The plants were arranged in a randomized design with four replications. When the plants had 5-6 leaves, drought treatment was initiated. The moisture content was measured with a Field Scout Soil Moisture meter. Prior to establishing water limited conditions, the soil had 35-40% Volumetric Water Content (VWC) which marked our well-watered control.

Vol. 2023 (2) 294 – 303
ISSN 2688-3651 [online]

When the moisture reached 11-12% VWC, this constituted the moderate drought (MD) group. The young fully expanded leaves were sampled by flash freezing in liquid nitrogen, placed in dry ice, and kept frozen at -80°C till RNA extraction. The pots were not watered for another 24 hours when the soil moisture content reduced to 7-9% VWC and the soil was visibly dry. Young leaves sampled at this stage belong to the severe drought (SD) group. The plants were then irrigated fully, and leaves were sampled again the next day when the moisture content was 35-40% VWC to constitute the recovery group. Sampling was done in triplicates to account for biological replicates in this experiment. Frozen leaves were ground and stored at -80°C.

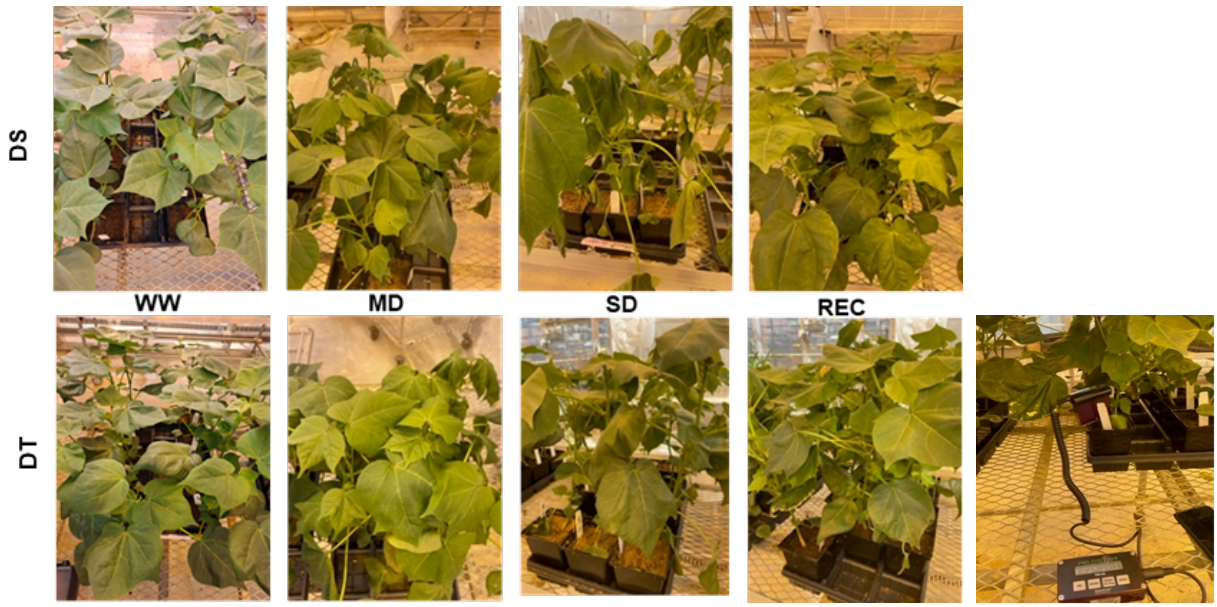


Figure 1. Cotton lines in the well-watered (WW), moderate drought (MD), severe drought (SD), and Recovery (Rec) period. The panel on the right shows the FieldScout Soil Moisture meter.

2.3 RNA Extraction

100 mg of tissue was used for RNA extraction using the Spectrum Plant Total RNA kit as per the manufacturer's instructions. The tissue was first lysed, DNAse treated, washed and RNA was eluted in 50ul of elution buffer. RNA quantity and quality were checked on nanodrop for A260/280 and A260/230 values and on an Agilent tapestation. The ratio of absorbance at 260nm and 280nm indicates the purity of RNA. A ratio close to 2 indicates that the RNA is pure. If the ratios are lower, it indicates contamination with proteins or remnant reagents from the extraction that absorb at 280nm. The ratio of absorbance at 260 nm and 230 nm is a secondary measure of RNA quality and a range of 2.1-2.2 indicates pure RNA and lower ratios indicate contaminants that absorb at 230nm.

2.4 Library Preparation and Sequencing

mRNA was enriched from total RNA using oligo dT beads that attach to the polyA tails of total RNA to select for mRNA. mRNA was reverse transcribed to cDNA using Reverse Transcriptase enzyme. cDNA was end prepped for adaptor ligation and indexing with barcodes followed by polymerase chain reaction (PCR) amplification for 12 cycles. The libraries were purified and then quantified using Qubit and normalized to 2nM and sequenced on a Nextseq 2000 sequencer with 2% PhiX for 300 cycles.

2.5 Data Analysis

Demultiplexed data was checked for quality using FastQC which is a quality control application for sequencing data (Andrews S, 2010). Four main parameters were used as indicators of quality. Per base sequence quality indicates

Vol. 2023 (2) 294 – 303
ISSN 2688-3651 [online]

the quality value across all bases at each position. Per sequence quality score indicates if subset of sequences will have low quality values. Per tile sequence quality indicates quality in each tile across all the bases to detect loss of quality in any part of the flowcell used for sequencing. Per base N content indicates the number of bases called as N if the sequencer is unable to make an accurate call for each base as A,T,G or C. The data was then mapped to the cotton genome (*Gossypium hirsutum* v1.1) using the mapper STAR (Spliced Transcript Alignment to a Reference) that identifies the location of each read in the sequencing data on the reference genome (Dobin et al, 2013). The BAM (binary alignment map) file from mapping and a GFF (general feature format) file with gene models were inputted to htseq_count which counts the number of aligned reads overlapping its exons for each gene. R was used to plot gene names and plot data, and DESeq was used for differential expression analysis. Pairwise comparison was done using the read counts from each of the treatment groups within each line. The output generated gene lists with p-value and fold change for each gene. DESeq uses the Wald test to calculate the p-value and the null hypothesis is that there is no differential expression between the sample groups. If the p-value is very small ($p<0.05$) then the null hypothesis is rejected indicating that the change in expression observed in the sample groups is significant (Anders S, Huber W, 2010). Fold change indicates if a gene is upregulated or downregulated. It is calculated as the ratio of the difference in read counts between the treatment and the control groups over the control and reported in the logarithmic scale to the base 2. A log2 fold change (FC) of greater than 1.5 is typically considered significant. A fold change value of 2 was used here for increased stringency and values greater than 2 indicate upregulation in expression and less than -2 indicate downregulation.

3. Results

3.1 Sequencing Statistics and quality check results

RNA measured on nanodrop resulted in an A260/230 value of 2.0 and an A260/280 value of 2.1 indicating that RNA was very pure. Library distribution was between 300-700 bp (base pair) and 10 million reads per library was generated across a total of 48 samples, of which an average of 9.3 million reads mapped uniquely to the genome indicating a 90% mapping rate. FastQC indicated high quality of the raw data: per base sequence quality >30; per tile sequence quality 0.3-0.4, per sequence quality score >30 and per base N content is <1%. If the per base sequence quality is >28, per tile sequence quality > 0.1, per sequence quality score of >27 and per base N content is less than 5%, then the quality is regarded as high based on the threshold. 88-92% of the reads are uniquely mapped to the genome. 10% of the reads were multi-mapped and 0.2% did not map to the genome indicating good quality sequencing data (Figure 2).

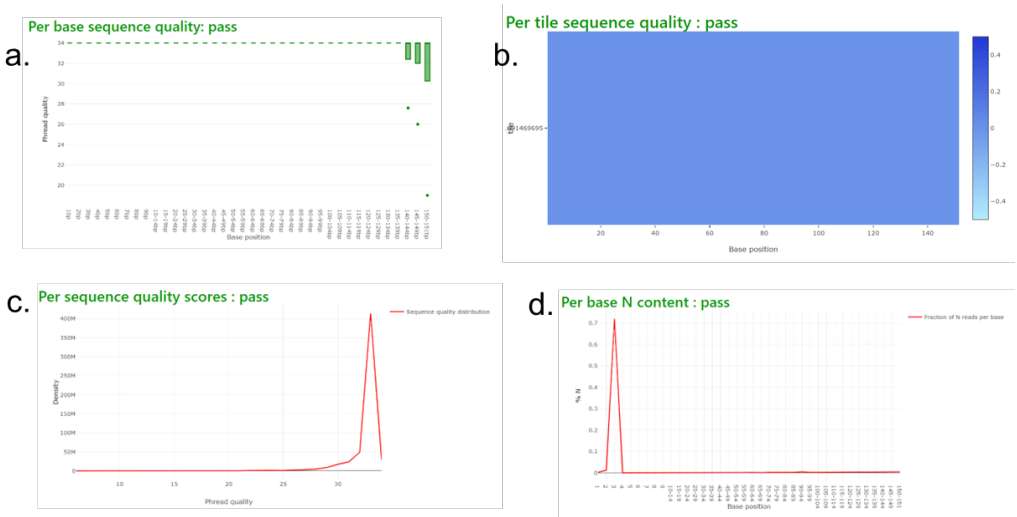


Figure 2. Quality check of the sequencing data. a) Per base sequence quality b) Per tile sequence quality c) Per sequence quality scores d) Per base N content

Vol. 2023 (2) 294 – 303
ISSN 2688-3651 [online]

Principal Component Analysis indicated that the samples clustered by treatment indicate that the treatment effect was stronger than the genotype effect (Figure 3). Well-watered and recovery groups clustered together indicating that the plants reverted to their well-watered state in the recovery period. The leaves after the recovery phase also appeared visually healthy like that of the initial well-watered state. Moderate drought and severe drought samples clustered distinctly from each other and from the control group also validating the correctness of the water-limiting conditions established in the glasshouse.

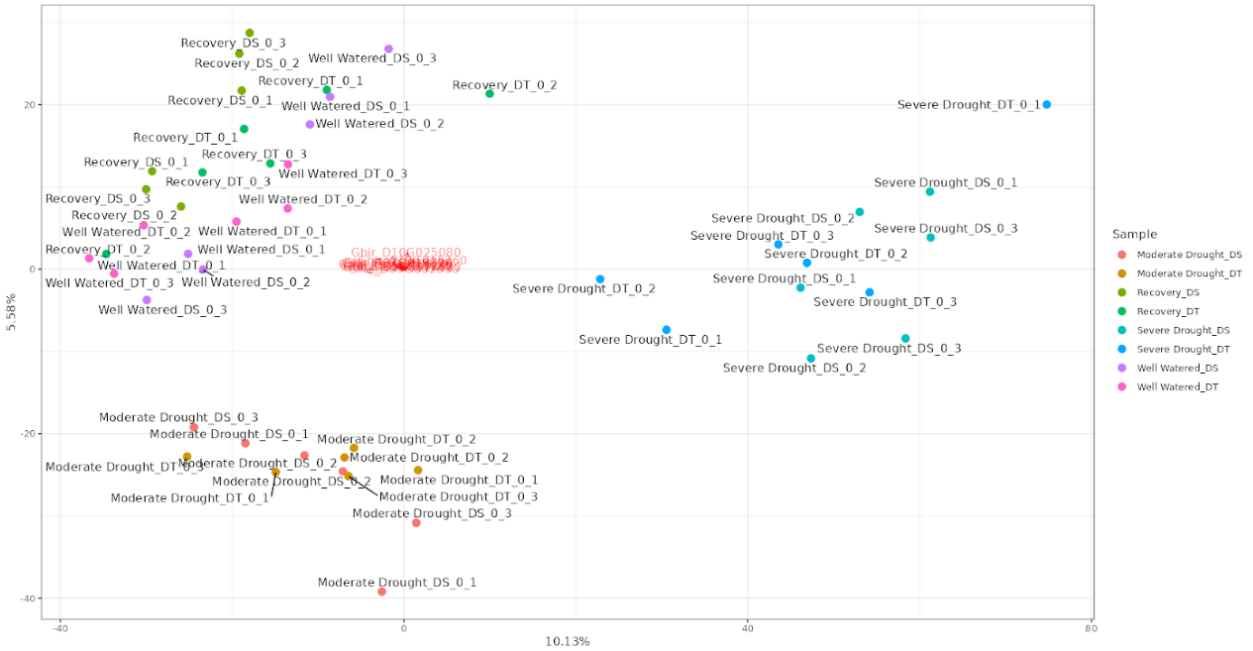


Figure 3. Principal Component Analysis (PCA) plot

3.2 Differential Gene Expression

Differential expression was performed between each drought treatment compared to well-watered control in each line (Table 1). Total number of genes that were differentially expressed included thousands of genes at $p<0.05$. The genes were further filtered to select for those that are changing more than 2-fold indicating hundreds of genes in each line with severe drought having more differentially expressed genes than moderate drought for each line.

Differential Gene expression was performed between the drought-sensitive (DS) and the drought-tolerant (DT) lines and between the drought severity treatments (moderate vs severe) within each group (Figure 4). 466 genes commonly changed in expression between MD and SD in lines 1189 and 197 genes in 3031 (DS lines). 256 genes were commonly changed in expression between MD and SD in lines 1179 and 293 genes in 1896 (DT lines). 1725 genes were commonly changed in expression among the severe drought and 267 in MD (DS lines). 870 genes were differentially expressed in the SD group and 314 genes in the MD group (DT lines). This indicates a strong drought response in all genotypes and the number of differentially expressed genes increased with the severity of the drought.

Table 1. Total number of differentially expressed genes that are statistically significant for each line and treatment type.

Line_treatment	# genes ($p<0.05$)	# genes ($p<0.05$;FC>2)	Line_type
1189_MD	6415	799	DS
3031_MD	8377	1088	DS
1179_MD	5421	759	DT
1896_MD	7996	535	DT
1189_SD	16101	3015	DS
3031_SD	14542	2362	DS
1179_SD	6258	1037	DT
1896_SD	19360	4799	DT

Vol. 2023 (2) 294 – 303
ISSN 2688-3651 [online]

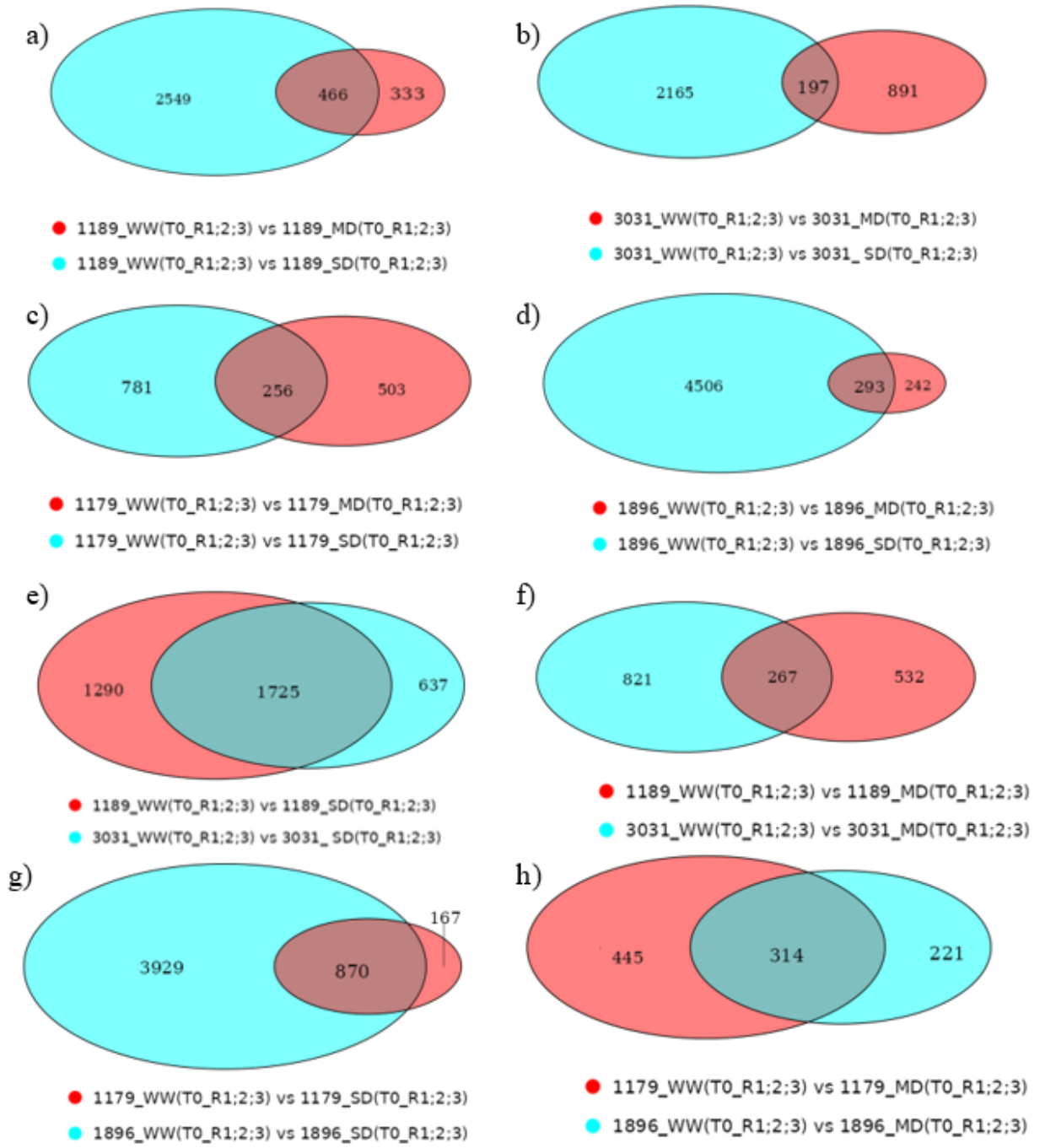


Figure 4. a) DS line 1189 MD vs SD b) DS line 3031 MD vs SD c) DT line 1179 MD vs SD d) DT line 1896 MD vs SD e) DS lines 1189 vs 3031 SD f) DS lines 1189 vs 3031 MD g) DT lines 1179 vs 1896 SD h) DT lines 1179 vs 1896 MD

The genes that were upregulated in SD indicate genes with a potential role in drought tolerance and regulation of stress response in the tolerant plants. NAC domain protein is a classic transcription factor involved in stress response and was upregulated 6-fold in SD. NAC transcription factors are relevant in abscisic acid-dependent and independent pathways in drought stress signaling. Overexpression of NAC genes in rice conferred enhanced drought and salt tolerance (Zheng et al., 2009). OsNAC10 increased tolerance to drought and provided increased grain production in rice (Jeong et al., 2010). Drought-induced unknown protein 1 (DIUP1) was upregulated 10-fold in this study.

Vol. 2023 (2) 294 – 303
ISSN 2688-3651 [online]

Overexpression of DIUP1 from alfalfa in Arabidopsis resulted in increased tolerance to drought, with higher seed germination, root length, fresh weight, and survival rate than in wild-type plants (Luo et al, 2023). Sugar transporter protein was upregulated 12-fold in severe drought conditions. Sugar transporter proteins are key determinants of the influx/efflux of various sugars and their metabolite intermediates that support the plant growth and developmental process. Abiotic stress, especially drought stress-mediated-injury is known to result in reprogramming of sugar distribution across the cellular and subcellular compartments (Kaur et al., 2021). Cytochrome p450 was upregulated 6-fold and it belongs to the family of oxidoreductase enzymes that are known to play a role in plant stress response. Galactinol and raffinose synthase were upregulated 10-fold in SD conditions. Raffinose and its precursor galactinol accumulate in plant leaves during abiotic stress-knockout in maize has been shown to modulate drought sensitivity (Li et al, 2020).

A lot of the drought-responsive genes that are significantly differentially expressed were common in both DT lines and DS lines. The difference was in the level of their upregulation in DT vs DS lines. Genes that were upregulated to a higher extent in DT lines and greater than 4-fold were investigated further (Figure 5).

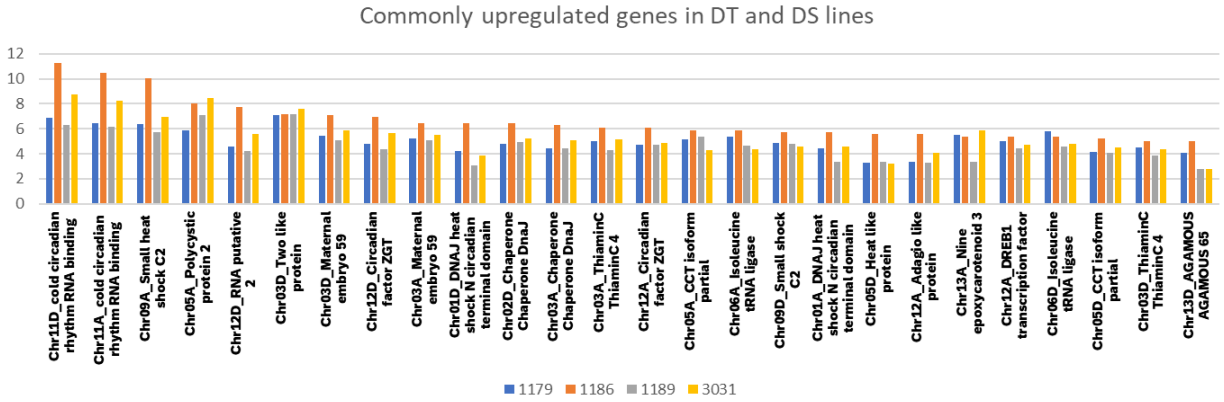


Figure 5. Commonly upregulated genes in DT (1179;1896) and DS (1189;3031) lines

These included cold circadian rhythms RNA binding-like (CCR-like) protein, heat shock protein, dehydration responsive element binding (DREB1) transcription factor, and agamous 65. These genes have supporting evidence in literature for their likely role in abiotic stress response regulation in other plant species. Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation. CCR-like proteins from cotton when overexpressed in tobacco were shown to confer tolerance to abiotic stress (Gurusamy et al., 2015). Heat shock proteins are classic examples of genes that are upregulated in plant stress response. Overexpression of soybean DREB1 has been shown to confer tolerance to water stress in field wheat (Zhou et al., 2020). Agamous-like proteins such as AGL22 have been shown to regulate a transcriptional network during drought stress, linking changes in primary metabolism and the initiation of stress responses in Arabidopsis (Wang et al., 2018). The genes that were commonly downregulated include LHY-like proteins, aquaporins, glutathione transferase, and inositol 3 phosphate synthases. LHY proteins are known to negatively control drought tolerance in soybean and aquaporins regulate water homeostasis in plants. There were no uniquely expressed genes identified in either group that were above the 2-fold threshold indicating that it is the regulation of expression of multiple genes and or pathways that are conferring adaptation of the plant to stress.

3.3 Gene Ontology (GO) Enrichment Analysis

Enrichment analysis was conducted to identify the functional and biological significance of the upregulated genes. The transcripts were grouped into three categories based on the GO annotation: Cellular Component, Biological process, and Molecular Function ($p<0.05$).

Vol. 2023 (2) 294 – 303
ISSN 2688-3651 [online]

In DT lines, the Biological Process category involved enrichment in positive regulation of circadian rhythm (GO:0042753), protein dephosphorylation (GO:0006470), biosynthetic process for thiamine, (GO:0009228), proline (GO:0006562) and galactose metabolic process (GO:0006012). The Cellular Component involved enrichment in the cell wall (GO:0005618), nucleosome (GO:0000786), extracellular region (GO:0005576), and CCAAT-binding factor complex (GO:0016602) were enriched in DT lines. Molecular Function involved enrichment in cation binding (GO:0043169), DNA binding transcription factor (GO:0003700), protein serine/threonine phosphatase activity (GO:0004722), inositol 3 phosphate synthase activity (GO:0004512) were enriched in DT lines.

The DS lines showed similar enrichment in the Chemical Component and Molecular Function categories when compared to the DT lines. The Biological Process includes regulation of DNA replication (GO:0006260), negative regulation of DNA helicase activity (GO:1905775), systemic acquired resistance (GO:0009627), initiation of DNA replication (GO:0006270) that were highly enriched in DS lines. DS lines showed a pronounced enrichment of DNA replication and enzymes involved in the process regulating stress response compared to DT.

4. Discussion

Two contrasting drought-responsive genotypes were analyzed by transcriptome sequencing under 2 drought conditions (moderate and severe drought). More than 3000 genes were differentially expressed in the DT and DS genotypes. Recovery treatment enabled regaining of the leaf health and the gene expression levels reverted to the WT state. Moderate drought is more reflective of a field drought condition as severe drought is an extreme effect not regularly happening in the field. The severe drought treated plants looked visually stressed as well. Key genes were identified that are likely contributors to conferring drought tolerance in cotton. Quantitative Reverse Transcription Polymerase Chain Reaction (qRTPCR) can be used to confirm the identified gene targets in the lab. This would involve the RNA used for sequencing to be first reverse transcribed to cDNA and primers specific to the genes used for PCR. Because the primers and probes are fluorescently labeled, the amount of amplified product would be measured using fluorescence in each PCR cycle. Validation of the genes is not essential by additional wet lab techniques for all targets identified by RNA sequencing, especially with the progress of the sequencing technology and inclusion of biological replicates. However, such validation experiments could benefit the genes with low expression differences such as those with less than 2-fold (Coenye 2021).

The genes identified here provide important targets for molecular breeding and potential use for increasing yield potential. Gene overexpression in cotton for functional analysis of key genes would unravel their role in stress response and contribution to stress tolerance in certain genotypes. This can be achieved by transgenic technologies that include amplification of the gene sequence, cloning into a plasmid vector, and transformation of cotton plant with the Agrobacterium to integrate the target DNA into desired genotypes. These plants can then be tested in well-watered and water limiting conditions to check for yield performance of these transgenic lines. The genes can be introduced in this way to any genotype to improve drought tolerance or combine with other desirable phenotypes existing in the recipient line used to create transgenics.

One of the limitations in this study is that the experiments were conducted in a controlled environment in pots instead of the field. The variability of the field environment cannot be exactly simulated in the controlled environment. In the field, the drought effect is typically in the moderate drought range and seasonal variability cannot be excluded. In the current set up for the study, the primary factor that is changing is the water levels. One way any additional variability was by having 3 biological replicates for each condition and ensuring the selection of statistically significant transcripts. Follow-up studies of these lines in the field would be useful for the validation and refinement of identified targets. In this study, leaf tissue was tested and expanding the tissue sources to include the roots which are responsible for water uptake will help gain a more holistic understanding of the molecular mechanism of drought response and identify potentially more targets. One developmental stage was picked which is in the transition of the vegetative and the reproductive stage as that mimics the timing of maximum stress in nature for cotton plants and this stage is known to be physiologically important for stress adaptation. More vegetative and reproductive stages of tissue sampling could be added for a broader understanding.

Vol. 2023 (2) 294 – 303
ISSN 2688-3651 [online]

The category of Biological Process from Gene Ontology Enrichment Analysis is a strong lead from this study due to the differences between the DT and DS lines. The positive regulation of circadian rhythm is enriched in DT lines, and a cold circadian rhythm RNA binding like (CCR-like) Gm protein was identified as highly differentially expressed in the DT lines compared to DS lines. The ThiaminC gene upregulated more than 4-fold belonging to the biosynthetic process for thiamine is an interesting lead from this study. In a recent paper, foliar application of thiamine on pea plants has enabled the plants to withstand drought stress (Kausar et al., 2023). Recently, ThiaminC was also identified to be induced under stress along with thiamine deficiency in cotton roots and exogenous application of thiamine restored the deficiency induced during stress (Li et al., 2022). The third category which is interesting to follow up is the galactose metabolic processes; galactinol and raffinose synthase was identified as the genes upregulated in drought stress. Raffinose and its precursor galactinol have been identified to accumulate in the leaves of plants during stress. Raffinose synthase transfers the galactosyl group from galactinol to sucrose to catalyze the formation of raffinose. Mutant maize plants lacking raffinose synthase completely lacked raffinose and hyper-accumulated galactinol making them more sensitive to drought stress. Maize raffinose synthase when overexpressed in Arabidopsis plants enhanced drought stress tolerance through either raffinose synthesis or galactinol hydrolysis, depending on sucrose availability in plant cells (Li et al., 2020).

In conclusion, genes that are differentially expressed in contrasting cotton genotypes with different stress tolerance have been identified in this research study. The genes involved in thiamine metabolism have also been reported to be enriched in a drought resistant line of Upland Cotton (Zheng et al., 2022). This indicates both the importance of the thiamine metabolism pathways in drought tolerance of *Gossypium hirsutum* and the probability of genotype independence for drought tolerant varieties as it is enriched in the 2 drought tolerant lines used for this research study as well. While few common gene functions have been identified based on previous research in the cotton field such as thiamine metabolism and heat shock proteins, some new genes and pathways have been identified here such as cold circadian rhythm, galactose metabolism, and agamous-like proteins. These genes and the processes they are involved in cotton demonstrate strong targets for molecular breeding and this study has enhanced our understanding of the molecular basis of cotton drought stress response.

5. References

Anders S, Huber W. (2010). Differential expression analysis for sequence count data. *Nature Precedings*. https://doi.org/10.1038/npre.2010.4282.1.

Andrews S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Chaves, M.M., Maroco, J.P., Pereira, J.S. (2003) Understanding Plant Responses to Drought-from Genes to the Whole Plant. Funct. *Plant Biol*. 30, 239.

Chaves, M.M., Flexas, J. and Pinheiro, C. (2009). Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Ann. Bot*.103, 551–560.

Coenye T. Do results obtained with RNA-sequencing require independent verification? (2021) *Biofilm*.13; 3:100043.

Dobin et al. (2013) STAR: ultrafast universal RNA-seq aligner, *Bioinformatics*, 29, 1, 15–21.

Giorgi Filippo, Lionello Piero. (2008). Climate change projections for the Mediterranean region, *Global and Planetary Change*, 63(2-3), 90-104.

Ghodke P et al. (2020). Comparative transcriptome analyses in contrasting onion (*Allium cepa* L.) genotypes for drought stress. *PLoS ONE* 15(8).

Vol. 2023 (2) 294 – 303
ISSN 2688-3651 [online]

Gurusamy, Dhandapani et al. (2015). A Gene Encoding Cold-Circadian Rhythm-RNA Binding-Like Protein (CCR-Like) from Upland Cotton (*Gossypium hirsutum* L.) Confers Tolerance to Abiotic Stresses in Transgenic Tobacco. *Plant Molecular Biology Reporter.* 33, 22–42.

Hasan, Md Mosfeq-Ul et al. (2019). Comparative Transcriptomic Analysis of Biological Process and Key Pathway in Three Cotton (*Gossypium* spp.) Species Under Drought Stress. *International Journal of Molecular Sciences* 20, no. 9: 2076

Jeong et al. (2010). Root-specific expression of *OsNAC10* improves drought tolerance and grain yield in rice under field drought conditions. *Plant Physiology*, 153 (1), 185–197.

Kaur H et al. (2021). Imperative role of sugar signaling and transport during drought stress responses in plants. *Physiol Plant.* Apr;171(4):833-848.

Kausar A et al. (2023). Alleviation of drought stress through foliar application of thiamine in two varieties of pea (*Pisum sativum* L.). *Plant Signal Behav*. Dec 31;18(1):2186045.

Li, T et al. (2020). Raffinose synthase enhances drought tolerance through raffinose synthesis or galactinol hydrolysis in maize and *Arabidopsis* plants. *J Biol Chem*. June 5;295(23):8064-8077

Li, W. et al. (2022). Thiamine Functions as a Key Activator for Modulating Plant Health and Broad-spectrum Tolerance in Cotton. *The Plant Journal*, 2(111), 374-390.

Luo Dong et al. (2023) DROUGHT-INDUCED UNKNOWN PROTEIN 1 positively modulates drought tolerance in cultivated alfalfa (Medicago sativa L.). *The Crop Journal*, 11 (1), 57-70.

Loka, D.M. et al. (2011). Water-deficit stress in cotton. In Stress Physiology in Cotton (Oosterhuis, D.M., eds), 37–72. 7. The Cotton Foundation Book Series. National Cotton Council of America.

Loka D.A, Oosterhuis D.M. (2012). Water stress and reproductive development in cotton. In: Oosterhuis, D.M. and Cothren, J.T., Eds., Flowering and Fruiting in Cotton, Publ. Cotton Foundation, Memphis, 51-57.

Parida, A.K. et al. (2007). Alterations in photosynthetic pigments, protein and osmotic components in cotton genotypes subjected to short-term drought stress followed by recovery. *Plant Biotechnol Rep*, 1, 37–48

Ullah Abid et al. (2017). Drought coping strategies in cotton: increased crop per drop *Plant Biotechnology Journal*,15, 271–284.

World Resources Institute. www.wri.org.

Wang Zhen et al. (2018). The Flowering Repressor SVP Confers Drought Resistance in Arabidopsis by Regulating Abscisic Acid Catabolism, *Molecular Plant*, 11 (9), 1184-1197.

Xu, C et al. (2022). Transcriptomics for Drought Stress Mediated by Biological Processes in relation to Key Regulated Pathways in *Gossypium darwinii*. *Mol Biol Rep* 49, 11341–11350.

Yuan Y et al. (2018). Genome-Wide Association Study Identifies Candidate Genes Related to Seed Oil Composition and Protein Content in Gossypium hirsutum L. *Front. Plant Sci*. Oct 22,9:1359.

Zheng, J et al. (2022). Whole Transcriptome Sequencing Reveals Drought Resistance-Related Genes in Upland Cotton. *Genes,*13, 1159.

Zheng Xingnan et al. (2009). Overexpression of a NAC transcription factor enhances rice drought and salt tolerance, *Biochemical and Biophysical Research Communications*, 379 (4), 985-989

Zhou Y et al. (2020). Overexpression of soybean DREB1 enhances drought stress tolerance of transgenic wheat in the field. *Journal of Experimental Botany*. Mar;71(6):1842-1857.

Journal of Research High School

Vol. 2023 (2) 304 – 311
ISSN 2688-3651 [online]

An Analysis of The Hyperloop's Feasibility as a Cost Efficient Solution to Public Transportation

Aaron Chen,[1] *

[1]Oakton High School, Vienna, VA, USA
*Corresponding Author: aaron.chen818@gmail.com

Advisor: Julian DuBransky, julian@7edu.org

Received March 17, 2023; Revised September 4, 2023; Accepted, October 4, 2023

Abstract

With growing concerns regarding the impact of car traffic on the environment and the poor quality of existing public transport infrastructure, especially in America, it is necessary to find a cost-efficient solution to the public transportation problem. This paper examined differences between existing public transport systems and the proposed Hyperloop to determine if the Hyperloop is feasible and effective when compared to existing methods. Projections from multiple Hyperloop companies were compared against existing systems using data from various sources including the Washington Metropolitan Area Transit Authority's public records and studies conducted on cargo throughput for freight trains. A focus was placed on the three main aspects of throughput, speed, and cost per mile - comparing each transportation system with an average freeway lane as a baseline. To determine the efficiency of each system, each aspect was evaluated against the aforementioned baseline in order to provide an overall effectiveness for a transport method. Findings revealed that that if the Hyperloop meets expectations for twenty-foot equivalent unit throughput capacity, it may be a far faster method to transport cargo than freight trains, being able to move 2800 twenty-foot equivalent units per day, compared to the 1786 twenty-foot equivalent units per day moved by freight trains in Felixstowe and 1221 twenty-foot equivalent units per day in Southampton. As for passenger throughput, the proposed Hyperloop system may be less efficient than existing high speed rail systems, namely the California high speed rail system which it was intended to replace.

Keywords: Hyperloop, Public Transport, Vactrain, High Speed Rail

1. Introduction

The Hyperloop is a transportation system centered around the concept of placing magnetically suspended pods in low-pressure tubes, first appearing in 1799 before being popularized by Elon Musk in 2013 as a more efficient mode of transport in a white paper titled "Hyperloop Alpha". The system sends pods propelled by magnetic propulsion through low-pressure tubes at high speeds. Existing systems are riddled with passenger violence problems, crowding, delays, and breakdowns - thus creating the need for an efficient method for public transport which works reliably. Despite the Hyperloop being touted as a system that could entirely revolutionize cargo shipping and transportation, some question the practical feasibility of the Hyperloop. Through the analysis of projections and expectations of current Hyperloop projects and data collected on existing methods of transportation, this paper assesses the feasibility of the Hyperloop as a method for cargo shipping and public transport.

Many doubt the feasibility of the Hyperloop, both from a technical and practical standpoint. Not only do some doubt that certain aspects of the Hyperloop Alpha white paper are possible, but also in regards to whether or not the Hyperloop is a worthy investment that can perform better than technologies already available to us.

Despite the Hyperloop Alpha white paper suggesting that the concept is theoretically possible and despite having working pods already built and tested, there are still some concerns on whether having a fully constructed Hyperloop would be a feasible decision. Richard Muller, a physics professor at UC Berkeley, raised concerns about the solar

Vol. 2023 (2) 304 – 311
ISSN 2688-3651 [online]

panels Musk proposes installing atop the tubes, and whether they would be able to power the entirety of the Hyperloop (Wolverton, 2016). Contrariwise, other estimates suggest that it may be possible to power the Hyperloop on solar panels alone, though it may not be an option outside of the Southwestern US due to lower levels of solar irradiation elsewhere (Rana, 2020), and the costs for constructing the Hyperloop may vary wildly due to other geographical constraints. Should powering the system be possible with solar panels alone, there would be a significant amount of emissions savings generated from diverting travelers from automobiles. However, other factors must be considered when it comes to the environmental impact of the Hyperloop in the entirety of its life cycle from the gathering of construction materials to disposal. No extensive analysis on the other environmental impacts of the Hyperloop (other than emissions) has been conducted as of the writing of this analysis, but Hyperloop projects such as the Great Lakes Hyperloop have begun to undertake Environmental Impact Studies.

Furthermore, safety and security concerns were brought up - if the pods were to get stuck, the entire tube would be rendered useless, and no system exists which allows for pods to convert or divert from one tube to another. Not only this, but such a system may require more airlocks, further dragging down throughput (Johnson, 2013) (Wolverton, 2016). However, these are merely concerns - so far, no such quantifiable metric for the safety of the Hyperloop has been calculated, but it can be presumed that as the system operates in a closed and controlled environment, risks from the weather and collisions may be significantly reduced.

Critics have raised concerns about how current prototypes and cost projections fall short of the original proposed concept (Adam Something, 2021). Members of the public when viewing publicity material, namely videos of test runs from Hyperloop One, cite the 100mph speeds and 2-person capacity of the XP-2 prototype and liken it to "reinventing the wheel" despite having the possibility of making improvements to the technologies involved in the future. However, the cost of making these improvements may be far more effort than it's worth.

James Moore, director of the transportation engineering program at the University of Southern California, stated that the development costs are the bigger problem, rather than construction costs (Wolverton, 2016). Despite the Hyperloop Alpha white paper's initial estimates showing that the Hyperloop may be a more affordable option with estimates as low as $11.5 million per mile of tubes on a route from Los Angeles to the Bay Area, leaked documents from the company Hyperloop One revealed that a 107 mile loop along the Bay Area alone could cost anywhere from $84 million up to $121 million per mile (Konrad, 2016).

The promise of being superior to both plane and rail transport is one of the most important reasons the Hyperloop has gained traction in the media. However, even this is a subject of intense debate. Proposed as a cheaper alternative to the California High-Speed Rail system, the Hyperloop promises speeds of 760 miles per hour at projected costs of $6 billion, allowing for costs as little as $11.5 million per mile of track. In comparison, the California High-Speed Rail system would have a budget of $68.4 billion and operate at a speed of 164 miles per hour. However, other sources suggest that it would be theoretically possible to build better high-speed rail systems for less - despite the California High-Speed Rail system costing around $123 million per mile, other projects in Europe can range from $33 to $53 million per mile (Konrad, 2016).

Not only this, but the speed of the Hyperloop and how much faster individual trips for passengers will take seem to be the main perks of the system, as shown on the Virgin Hyperloop website and in the Hyperloop Alpha white paper. However, the main concern from many who oppose the Hyperloop is the passenger throughput and freight train capacity (Adam Something, 2021), as the original Hyperloop Alpha white paper proposed pods with a capacity for 28 passengers, but would be spaced out by around 23 miles or from 30 seconds to 2 minutes on average during operation. However, others doubt that 30 seconds between pods would be enough should a safety issue occur - there wouldn't be enough time to decelerate to prevent crashing (Johnson, 2013). Comparisons between the Hyperloop and the existing train system were also made, arguing that it would be more efficient to have slower cars all connected rather than much faster pods all with independent motors being launched at intervals (Adam Something, 2021).

Setting aside speed and cost, throughput is also an important factor to consider. The Hyperport, proposed by Hyperloop TT would have pods capable of carrying 2 TEUs, or Twenty-Foot Equivalent Units, (Drăgan, 2021), as stated on their website, in comparison to a mean capacity of around 55 - 60 TEUs per train in the ports of Felixstowe, Southampton, Tilbury, and Thamesport (Woodburn, 2011). Woodburn also found that rail TEU throughput in the ports of Felixstowe and Southampton were around 1,786 and 1,221 TEUs per day, respectively - in contrast to the

Vol. 2023 (2) 304 – 311
ISSN 2688-3651 [online]

2,800 TEUs per day promised by the Hyperport. It should be noted that the study only measured actual TEU throughput rather than capacity, as throughput is usually lower than theoretical maximum operating capacity. Hyperloop TT stated that their system "can move 2,800 Twenty-foot Equivalent Units, or TEUs, a day in an enclosed operating environment".

Estimates for the efficiency of passenger throughput vary wildly as well. The original white paper established that it would be possible to meet 840 passengers per hour, yet Virgin Hyperloop's proposed system is estimated to have the capacity for 50,000 passengers per hour, per direction. However, considering that a rough estimate for the time taken to decelerate to avoid collisions would probably be around 80 seconds minimum, the maximum number of passengers who would be able to depart in one hour hovers closer to 1,260 (Johnson, 2013). Even if 30-second time between departures were used during peak hours as suggested in Elon Musk's white paper, only 3,360 passengers could depart in one hour.

These estimates are not as exciting when compared to already existing transportation technologies. For instance, the California High-Speed Rail project is estimated to have a capacity of 12,000 passengers an hour, a subway running at 3-minute headways such as the WMATA Red Line can carry 36,000 passengers per hour, and a freeway lane on its own can carry 2,000 cars per hour (Johnson, 2013). Furthermore, to reach or surpass rail throughput, more tubes would be needed. As building costs are a crucial factor, more tracks may not be a viable option (Eichelberger et al., n.d.).

Prior literature has primarily conducted studies on the economic and technical feasibility of the Hyperloop, with an emphasis placed on the speed that the system is capable of. However, aside from assessing if the system is physically possible to engineer in the first place and if a market exists, little has been done to determine if the rate of passengers moving through the Hyperloop is actually more efficient in terms of throughput than other methods of transportation, rather than looking at how fast the system can move an individual passenger. As such, this paper has aimed to compare the throughput of the Hyperloop for both freight/cargo and passengers with existing methods to examine the overall cost efficiency of the Hyperloop system against existing methods of transport.

2. Materials and Methods

This research focuses on the efficiency of various different public transport methods in terms of passenger throughput, speed, and cost. The analysis is based entirely on secondary research, collected through projections and estimates provided by companies, public records, and other studies conducted on rail network efficiency, placing focus on the three main aspects of throughput, speed, and cost per mile - comparing each transportation system with an average freeway lane as a baseline, expressed as a value of 1.00. To determine the efficiency of each system, each aspect was evaluated against the baseline, with the averages of each value for throughput, speed, and cost per mile in order to provide an overall effectiveness for a transport method. The values for each individual aspect are expressed as the ratio of how much "better" it is - for example, if the passenger throughput of the baseline transportation system is 100 passengers per hour, while another system can move 120 passengers per hour, the values for the "passenger throughput" for each system is 1.00 and 1.20 respectively, (second system divided by baseline) as a higher throughput is better. Contrariwise, if the baseline system were to cost $1 million per mile whilst another cost $1.5 million per mile, the values for each system are 1.00 and 0.67 respectively, (baseline divided by second system) as a higher cost is worse. As such, a higher value indicates higher performance and lower cost. The overall "score" for a transportation system is the average of all three values for passenger throughput, speed, and cost.

It is important to note that due to a lack of previous studies, measurements, data points, and the fact that different studies have a different definition of "commercial operating speeds", the true average speed of certain systems over an entire trip may be slower than what is listed in the tables below. For the purposes of this paper, we will be using the maximum hypothetical operating speed for the Virgin Hyperloop, Hyperloop TT and Hyperloop One estimates, while the Great Lakes study has provided more accurate estimates including acceleration and deceleration between stations/during curves. For the high speed rail speed estimates (save for the California HSR), we will be using the commercial operating speed rather than the maximum operating speed - assumed to be the speed at which most vehicles are observed to travel at in free-flow, favorable conditions. The California HSR estimate was obtained by calculating average speed using the duration of the trip along with the total track distance. For the freeway estimate,

Vol. 2023 (2) 304 – 311
ISSN 2688-3651 [online]

a 100 kph speed was adopted, as observed in the speed limits of most urban freeways across most states in the United States. The actual observed speed may differ from this as Americans typically drive considerably faster than the speed limit under favorable weather conditions.

3. Results

3.1 Hyperloop Proposals

It is to be noted that passenger throughput and speed estimates for Hyperloop One and HyperloopTT systems are assumed to be the same as the system outlined in the Hyperloop Alpha white paper, as they intend on using the same design, while Virgin Hyperloop's design utilized a different multi-track system. Cost statistics for the Bay Area hyperloop and HyperloopTT were taken by averaging the minimum and maximum projected cost per mile. Additionally, Virgin Hyperloop and Hyperloop One are the same company (having changed their name), but have provided different estimates for different systems - and as such have been listed separately so as to avoid ambiguity. The Great Lakes study only provided a wide range of costs for the Hyperloop system overall, and as such the cost per kilometer has been averaged. More importantly, their speed estimates were calculated with a 0.1G acceleration time and deceleration needed to navigate curves along a planned route, while the other studies use a more optimistic maximum operating speed.

Table 1. A table outlining rough estimates and projections for the cost per mile, passenger throughput, and speed of several proposed Hyperloop systems.

Transportation Method	Cost (Cost/Km)	Passenger Throughput (Passengers/Hr)	Speed (Km/Hr)
Virgin Hyperloop	$7,130,000[a]	50000[b]	1080[b]
Hyperloop Alpha	$7,130,000[a]	840[a]	1220[a]
Hyperloop One (Bay Area)	$63,550,000[d]	840[a]	1220[a]
Hyperloop One (Abu Dhabi)	$32,240,000[d]	840[a]	1220[a]
Hyperloop One (Helsinki/Stockholm)	$39,680,000[d]	840[a]	1220[a]
HyperloopTT	$7,750,000[e]	840[a]	1220[a]
Great Lakes (Hypothetical, No Curves)	$38,835,000[c]	840[a]	954[c]
Great Lakes (Chicago/Cleveland)	$38,835,000[c]	840[a]	706[c]
Great Lakes (Hybrid Chicago/Cleveland)	$38,835,000[c]	840[a]	891[c]
Great Lakes (Cleveland/Pittsburgh)	$38,835,000[c]	840[a]	545[c]
Great Lakes (Hybrid Cleveland/Pittsburgh)	$38,835,000[c]	840[a]	719[c]

[a](Hyperloop Alpha, n.d.). [b](Virgin Hyperloop, n.d.). [c](HyperloopTT, 2020). [d](Konrad, 2021). [e](HyperloopTT, n.d.).

Several key points can be identified: the first of which being the vast jump between the original cost per mile estimates in the Hyperloop Alpha white paper and the Great Lakes and Hyperloop One estimates. Furthermore, it can be seen that there is another large gap between the Hyperloop Alpha proposal's initial estimate for passenger throughput and Virgin Hyperloop's proposal - Virgin Hyperloop's concept appears to be able to move nearly 60 times the amount of passengers in an hour.

Table 2. A table outlining the cost per mile, passenger throughput, and speed for several high speed rail systems in Europe, as well as the average statistics for all three individual aspects.

Transportation method	Cost (Cost/Km)	Passenger Throughput (Passengers/Hr)	Speed (Km/Hr)
Rome-Naples HSR	–	416[f]	300[f]
High Speed 1	$35,278,000[a]	–	300[a]
Naples-Salerno HSR	–	416[c]	250[c]
Mannheim-Stuttgart HSR	$12,400,000	–	280
LGV EST	$15,190,000[e]	–	320[d]
LGV Sud-Est	$6,200,000[b]	–	300[b]
Average	$17,267,000	416	292

[a]("High Speed 1", 2022). [b]("LGV Sud-Est", 2022). [c]("Naples–Salerno high-speed railway", 2022). [d](LGV EST, n.d.). [e]("LGV Est", 2023). [f](Brown, 2017).

3.2 High Speed Rail in Europe

As this paper relies entirely on secondary research, it was not possible to retrieve data for certain aspects of certain rail lines. As such, the incomplete data for each rail line has been

Vol. 2023 (2) 304 – 311
ISSN 2688-3651 [online]

3.3 Overall Statistics

Table 3. A table combining the statistics of the previously mentioned HSR systems and Hyperloop proposals while also incorporating statistics for the WMATA Green Line and the average American freeway lane.

Transportation method	Cost (Cost/Km)	Passenger Throughput (Passengers/Hr)	Speed (Km/Hr)
Freeway Lane	$4,774,000[j]	2000[i]	100[k]
Virgin Hyperloop	$7,130,000[a]	50000[b]	1080[b]
Hyperloop Alpha	$7,130,000[a]	840[a]	1220[a]
WMATA Green Line	$62,000,000[f]	6300[g]	121[h]
California HSR	$81,840,000[d]	12000[i]	270[i]
Hyperloop One (Bay Area)	$63,550,000[d]	840[a]	1220[a]
Hyperloop One (Abu Dhabi)	$32,240,000[d]	840[a]	1220[a]
Hyperloop One (Helsinki/Stockholm)	$39,680,000[d]	840[a]	1220[a]
HyperloopTT	$7,750,000[e]	840[a]	1220[a]
HSR Composite	$17,267,000	416	292
Great Lakes (Hypothetical, No Curves)	$38,835,000[c]	840[a]	954[c]
Great Lakes (Chicago/Cleveland)	$38,835,000[c]	840[a]	706[c]
Great Lakes (Hybrid Chicago/Cleveland)	$38,835,000[c]	840[a]	891[c]
Great Lakes (Cleveland/Pittsburgh)	$38,835,000[c]	840[a]	545[c]
Great Lakes (Hybrid Cleveland/Pittsburgh)	$38,835,000[c]	840[a]	719[c]

[a](Hyperloop Alpha, n.d.). [b](Virgin Hyperloop, n.d.). [c](HyperloopTT, 2020). [d](Konrad, 2021). [e](HyperloopTT, n.d.). [f](McGowan, 2005). [g](WMATA, 2019). [h](WMATA, 2015). [i](Johnson, 2013). [j]("Speed limits in the United States", 2023). [k](Strong Towns, 2020).

top. However, it is important to note is that the Hyperloop's success hinges on its ability to maintain its speed for long periods of time - the more detailed Great Lakes speed estimates which accounted for deceleration inside of curves rarely ever had capsules reaching their maximum speed. Not only this, but a perfectly straight tube accounting for acceleration and deceleration at the beginning and ends of the trip, respectively, was estimated to only be 3.15 times as efficient as a freeway lane, only slightly more efficient overall than the California HSR project. The value of 2.92 in the California HSR also fails to account for curves, acceleration, and deceleration, though, and as such may fall even lower when measured after its completion. Furthermore, the European HSR composite showed to be only condensed by taking the average of all existing statistics, now referred to as the "HSR Composite" from hereon after. It is to be noted that the cost statistic for the Mannheim-Stuttgart HSR line was taken by averaging the minimum and maximum estimated costs, converted from 1973 Deutsche Marks - not accounting for inflation.

Despite slightly lower costs, passenger throughput and speed for high speed rail cannot compete with the Hyperloop - with the proposed system estimated to be able to reach 1220 kph compared to existing operating speeds of 250 to 320kph, and with the Hyperloop's projected 840 to 50000 passengers per hour compared to existing high speed rail systems' throughput of 416 passengers per hour.

3.4 Comparison with Baseline

Virgin Hyperloop's proposal appears to be an outlier, promising to be 12.16 times more efficient than a freeway lane, far surpassing any other proposal or existing transport method. Again, the statistics gathered from their website are most likely optimistic projections due to the lack of any existing Hyperloop system. The original Hyperloop Alpha proposal can be estimated to be 4.43 times as efficient as a freeway lane, and the California HSR project which the Hyperloop was intended to derail was found to be around 2.92 times more efficient, putting the original Hyperloop Alpha concept on

Table 4. A table comparing the statistics of every transportation method previously mentioned against the average freeway lane, expressed as a ratio of how much "better" a transportation method is. A larger value means higher performance.

Transportation Method	Passenger Throughput	Speed	Cost per Mile	Overall
Freeway Lane	1.00	1.00	1.00	1.00
Virgin Hyperloop	25.00	10.80	0.67	12.16
Hyperloop Alpha	0.41	12.20	0.67	4.43
WMATA Green Line	3.15	1.21	0.08	1.48
California HSR	6.00	2.70	0.06	2.92
Hyperloop One (Bay Area)	0.42	12.20	0.08	4.23
Hyperloop One (Abu Dhabi)	0.42	12.20	0.15	4.26
Hyperloop One (Helsinki/Stockholm)	0.42	12.20	0.12	4.25
Hyperloop One (Abu Dhabi)	0.42	12.20	0.15	4.26
Hyperloop One (Helsinki/Stockholm)	0.42	12.20	0.12	4.25

Vol. 2023 (2) 304 – 311
ISSN 2688-3651 [online]

Table 4_Continued

Transportation Method	Passenger Throughput	Speed	Cost per Mile	Overall
HyperloopTT	0.42	12.20	0.62	2.88
HSR Composite	0.20	2.92	0.28	1.13
Great Lakes (Hypothetical, No Curves)	0.42	9.54	0.12	3.36
Great Lakes (Chicago/Cleveland)	0.42	7.06	0.12	2.53
Great Lakes (Hybrid Chicago/Cleveland)	0.42	8.91	0.12	3.15
Great Lakes (Cleveland/ Pittsburgh)	0.42	5.45	0.12	2.00
Great Lakes (Hybrid Cleveland/Pittsburgh)	0.42	7.19	0.12	2.57

from a wider variety of examples. Even with limited data, the Hyperport appears to have a larger TEU throughput capacity than that of the ports of Felixstowe and Southampton. It is also important to note that the 2800 TEUs per day outlined by HyperloopTT's estimates is an operating capacity rather than actual recorded throughput.

1.13 times as efficient when compared to a freeway lane - though the lack of complete statistics on specific lines makes it hard to provide an accurate average.

3.5 TEU Throughput

HyperloopTT's HyperPort aims to utilize the Hyperloop as a high-speed cargo transport and freight solution, using individual pods rather than trains to move TEUs. As the studies conducted within this paper rely entirely on secondary research, it was difficult to gather TEU throughput statistics for freight trains

Table 5. A table comparing TEU throughput per day between rough estimates for the HyperPort's TEU throughput capacity and actual measured TEU throughput for rail freight in the ports of Felixstowe and Southampton.

Port	TEU Throughput (TEUs/Day)
HyperPort	2800[a]
Felixstowe	1786[b]
Southampton	1221[b]

[a](Drăgan, 2021). [b](Woodburn, 2011)

4. Discussion

This paper has analyzed the efficiency of various transportation methods in contrast to the proposed Hyperloop. With limited published information, many holes in existing data, and very few existing prototypes, this paper has relied heavily on optimistic estimates - most prior research was conducted on whether the Hyperloop is feasible from a technical standpoint, rather than if investing into Hyperloop technologies is worth it in the first place. The preceding analysis identified that if the Hyperloop meets expectations for TEU throughput capacity, it may be a far faster method to transport cargo than freight trains, being able to move 2800 TEUs per day, compared to the 1786 TEUs per day moved by freight trains in Felixstowe and 1221 TEUs per day in Southampton. However, the 2800 TEUs per day is an operational capacity rather than an accurate estimate of the actual amount of cargo that will move through the system, compared to the actual measured results at Southampton and Felixstowe.

Furthermore, Virgin Hyperloop's proposal stood out as an outlier, promising to be almost 12 times as efficient than a traditional freeway - though it should be noted that the statistics gathered were only projections, and that actual data may vary heavily if the Hyperloop is to be established. Additionally, current prototypes such as the XP-2 are not even close to reaching the promised capabilities of the Hyperloop, which makes it even more difficult to accurately determine if the Hyperloop can outcompete other methods of transport. Setting the Virgin Hyperloop proposal aside, the analysis determined that the proposed Hyperloop system may be less efficient in terms of passenger throughput than existing high speed rail systems, namely the California HSR system which it was intended to replace. When factoring in speed, cost, and efficiency as a whole, the Hyperloop may only be more efficient than high speed rail when it is capable of sustaining its maximum speed for long periods of time without needing to decelerate at curves - and when deceleration in non-linear routes is factored in, the Hyperloop system may even be less efficient than high speed rail overall. Again, the ability of the Hyperloop to meet its original expectations for speed is still not completely certain.

Further investigation placing a heavier emphasis on primary research, analyzing a wider range of aspects such as safety and environmental impacts, calculating the time required for a system to pay itself off, and using data from functional prototypes would be desirable in order to make a more conclusive decision on the viability of the Hyperloop

Vol. 2023 (2) 304 – 311
ISSN 2688-3651 [online]

as a cost-efficient transportation method as speed, throughput, and cost are not the only factors that should be considered. Quantifiable measures for safety should be obtained such as deaths per 100,000 passengers, and an extensive analysis on the environmental impacts of the Hyperloop's full life cycle should be conducted rather than simply considering its emissions while operational.

References

Adam Something. (2021, July 25). The HyperPort Is Dumb And Will Most Likely Explode [Video file]. YouTube. https://www.youtube.com/watch?v=dNzjk-kiUmQ

Adam Something. (2021, May 7). The HYPERLOOP Will Never Work, And Here's Why [Video file]. YouTube. https://www.youtube.com/watch?v=CQJgFh_e01g

Atkins. (n.d.). First Interim Evaluation of the Impacts of High Speed 1.

Decker, K., et al. (n.d.). Conceptual Feasibility Study of the Hyperloop Vehicle for Next-Generation Transport. Cleveland, Ohio; NASA Glenn Research Center.

Drăgan, O. (2021, July 7). Forget standard cargo shipping, this sustainable capsule moves at airplane speed. autoevolution. Retrieved July 23, 2022, from https://www.autoevolution.com/news/forget-standard-cargo-shipping-this-sustainable-capsule-moves-at-airplane-speed-164816.html

Eichelberger, M., et al. (n.d.). High-Throughput and Low-Latency Hyperloop. Z¨urich, Switzerland; ETH Z¨urich.

HyperloopTT. (2022, May 18). Feasibility study. HyperloopTT. Retrieved September 27, 2022, from https://www.hyperlooptt.com/projects/feasibility-study/

Johnson, M. (2013, August 14). Musk's hyperloop math doesn't add up. Greater Greater Washington. Retrieved July 23, 2022, from https://ggwash.org/view/32078/musks-hyperloop-math-doesnt-add-up

Konrad, A. (2016, November 18). Leaked hyperloop one docs reveal the startup thirsty for cash as costs will stretch into billions. Forbes. Retrieved July 23, 2022, from https://www.forbes.com/sites/alexkonrad/2016/10/25/hyperloop-one-seeks-new-cash-amid-high-costs/?sh=154c4ba8125c

McGowan, P. (2019, August 18). Fort Meade proposes metro extension. Baltimore Sun. Retrieved September 27, 2022, from https://www.baltimoresun.com/bal-md.ar.bases09jun09-story.html

Milne, A. (2017). Napoli Afragola – Italy's remarkable new station. RailEngineer. Retrieved October 1, 2022, from https://www.railengineer.co.uk/2017/06/28/napoli-afragola-italys-remarkable-new-station/

Rana, Y. (2020). On the feasibility of the Hyperloop Concept. Retrieved July 23, 2022, from https://dspace.mit.edu/handle/1721.1/127876.

Shrikant, A. (2018, September 26). Why US public transportation is so bad - and why Americans don't care. Vox. Retrieved July 23, 2022, from https://www.vox.com/the-goods/2018/9/26/17903146/mass-transit-public-transit-rail-subway-bus-car

Strong Towns. (2020, January 27). How much does a mile of road actually cost? Strong Towns. Retrieved September 27, 2022, from https://www.strongtowns.org/journal/2020/1/27/how-much-does-a-mile-of-road-actually-cost

The Antiplanner. (2020, November 8). 10 reasons not to build high-speed rail. The Antiplanner. Retrieved October 2, 2022, from https://ti.org/antiplanner/?p=17764

Thompson, D. (2020, February 12). California bullet train adds another $1.3 billion to projected price tag. Los Angeles Times. Retrieved October 2, 2022, from https://www.latimes.com/california/story/2020-02-12/california-bullet-train-adds-another-1-billion-to-projected-costs-for-project

Vol. 2023 (2) 304 – 311
ISSN 2688-3651 [online]

Un nouvel espace temps avec des TGV circulant à 320 km/h. LGV EST. (n.d.). Retrieved October 1, 2022, from http://www.lgv-est.com/rff-construit-la-lgv.php?pg=289. Archived at https://web.archive.org/web/20040506050638/http://www.lgv-est.com/rff-construit-la-lgv.php?pg=289

Virgin Hyperloop. (n.d.). Retrieved October 23, 2022, from https://virginhyperloop.com/
Wikimedia Foundation. (2022, August 22). Speed limits in the United States. Wikipedia. Retrieved September 27, 2022, from https://en.wikipedia.org/wiki/Speed_limits_in_the_United_States

Wikimedia Foundation. (2022, February 10). Naples–salerno high-speed railway. Wikipedia. Retrieved October 1, 2022, from https://en.wikipedia.org/wiki/Naples%E2%80%93Salerno_high-speed_railway#cite_note-tr32008-1

Wikimedia Foundation. (2022, September 28). High speed 1. Wikipedia. Retrieved October 1, 2022, from https://en.wikipedia.org/wiki/High_Speed_1

Wikimedia Foundation. (2022, September 5). LGV Sud-Est. Wikipedia. Retrieved October 2, 2022, from https://en.wikipedia.org/wiki/LGV_Sud-Est

Wikimedia Foundation. (2023, May 29). LGV Est. Wikipedia. Retrieved August 20, 2023, from https://en.wikipedia.org/wiki/LGV_Est

Wikimedia Foundation. (2023, August 19). Speed limits in the United States. Wikipedia. Retrieved August 20, 2023, fromhttps://en.wikipedia.org/wiki/Speed_limits_in_the_United_States

WMATA. (2015). 2015 METRO FACTS.

WMATA. (2019). Metrorail Average Weekday Passenger Boardings.

Wolverton, T. (2016, August 12). Wolverton: Elon Musk's Hyperloop hype ignores practical problems. The Mercury News. Retrieved July 23, 2022, from https://www.mercurynews.com/2013/08/13/wolverton-elon-musks-hyperloop-hype-ignores-practical-problems/

Woodburn, A. (2011, April). An Investigation of Container Train Service Provision and Load Factors in Great Britain. Retrieved July 23, 2022, from https://www.researchgate.net/publication/237397871_An_Investigation_of_Container_Train_Service_Provision_and_Load_Factors_in_Great_Britain.

Journal of Research High School

Vol. 2023 (2) 312 – 317
ISSN 2688-3651 [online]

Mental Health and Physical Activity in Adolescents: Analysis of 2021 Youth Risk Behavior Study Data

Nathaniel J. Choi[1] *

[1]Marriotts Ridge High School, Ellicott City, MD, USA
*Corresponding Author: stchoi79@gmail.com

Advisor: Erin Kim, erinmkim@msn.com

Received July 1, 2023; Revised September 14, 2023; Accepted, October 12, 2023

Abstract

The latest 2021 Youth Risk Behavior Survey (YRBS) data were analyzed to explore association between mental health and physical activity. Through analyzing YRBS data, females were more likely to be impacted by mental health during Covid than males. Baseline statistics were carried out and used chi- square to test for differences. Race and sex were controlled for in the study and odds ratios were calculated for subgroups. Twenty-nine percent of youth reported mental health was most of the time or always not good. Highest reports of poor mental health were reported among American Indians/Alaskan Natives. Ninth graders were less likely to report poor mental health compared to 11th and 12th graders. The prevalence of poor mental health was 29.3 %. The results demonstrated an association between poor mental health and lack of physical. Thirty two percent of youth were overweight and obese. Controlling for race and gender, the statistically significant groups were who were found to have an association between reports of not being physically active and poor mental health were Black males and white females. In addition, white females, American Indian/Alaskan Native and Asian Females were at increased risk of reporting poor mental health during Covid pandemic among those who were not physically active for at least 60 minutes on at least 1 day. These findings point for the need for improved school–based services for mental health and the need for schools to provide mental health services or referral systems to community resources. These findings of association between mental health and physical activity were consistent with previous studies.

Keywords: Mental Health, Physical Activity

1. Introduction

Obesity is a major contributor to morbidity and mortality in the US population. Obesity impacts public health with obese having higher risks for diabetes, stroke, cancer, and premature death. Previous research indicates that obesity can affect youth's psychological as well as cardiovascular health and their overall physical health. The association between obesity and other conditions makes it a public health concern for youth. Due to the increase in the prevalence of obesity among children. (Sanyaolu, et al., 2019) Mental health impacts our emotional state, how we act and think. Mental disorders are chronic health conditions that last a long time and often there is no cure. Without early diagnosis and treatment, youth with mental disorders can have problems at home, in school, and in friendships. Mental disorders can also interfere with a child's healthy development, causing problems that can continue into adulthood. (CDC,2023) A previous study found that teens with both physical and mental health disorders were more likely to have very low quality of life as adults, in comparison with teens who suffered from only one form of disorder. (Chen, et al., 2006) Previous research also suggests that comorbid physical and mental health problems persist into adulthood (Cohen, et al., 1998) Given the previous research findings, additional research studies are needed

Vol. 2023 (2) 312 – 317
ISSN 2688-3651 [online]

to discover what associations and risk factors increase the probability that youth will present with obesity and mental health issues and whether they are associated.

This study aims to seek if there is an association between poor mental health and physically activity among youth. Hypothesis is that poor mental health is associated with not being physically active. Also, is there an increased risk for less physical activity among those who reported poor mental health? Youth that report poor mental health will be more likely to report physical inactivity than youth who do not report poor mental health during Covid-19. In addition, it is hypothesized that there will be statistically significant differences by gender and race. Given the public health implications this research study can provide insights in identifying potential higher risks youth for planning targeted public health interventions for youth. Poor mental health was determined by respondents who answered their mental health was most of the time or always not good (CDC, 2021). A significant percentage (37%) of youth reported poor mental health during Covid.

2. Methods

Statistical analysis was performed on YRBS data imported into Epi Info 7 using procedures that accommodate the weighted sampling design of YRBS. YRBS is used to monitor priority health risk behaviors among youth in the United States. The national Youth Risk Behavior Survey (YRBS) uses a three-stage cluster sample design to produce a representative sample of 9th through 12th grade students (CDC, 2021). National data in YRBS High school student survey for 2021 were used to analyze the data. (CDC, 2021) Baseline statistics were carried out and used chi- square to test for differences. Statistical analysis was carried out using Epi Info 7 software. Race and sex were controlled for in the study and odds ratios were calculated for subgroups.

Poor mental health was defined by specific questions on the YRBS. “During the past 30 days, how often was your mental health not good? (Poor mental health includes stress, anxiety, and depression.” (CDC, 2021) The other YRBS question included the time period of during the pandemic. “During the COVID-19 pandemic, how often was your mental health not good? (Poor mental health includes stress, anxiety, and depression.” (CDC, 2021)

The following are statistical definitions and purpose of p values, 95% confidence interval and odds ratios used in this study. P values are defined as the probability assuming no effect or difference of obtaining a result that what was observed. P value measures how likely it is that any observed difference between groups is due to chance. A p-value of 0.05 or lower is generally considered statistically significant. 95% confidence interval is a range of values that is likely to include a populations value with a certain degree of confidence. It is expressed as a percentage whereby a population mean lies between an upper and lower interval. Odds ratio is the ratio of the odds of the event happing in an exposed group versus a non-exposed group. The odds ratio is used to report the strength of association between an exposure and an event. The larger the odds ratio the more likely the event is to be found with exposure.

Calculation of odds ratio is described below as shown in Figure 1. The odds ratio is odds of the event in the exposure group (a/b) divided by the odds of the event in the control or non-exposure group (c/d). Therefore, the odds ratio is (a/b) / (c/d) which can be simplified to ad/bc.

		Outcome	
		Yes	No
Exposure	Yes	a	b
	No	c	d

$$Odds\ Ratio = \frac{\text{Odds of the Outcome in Exposed Group}}{\text{Odds of the Outcome in Non} - \text{Exposed Group}}$$

$$Odds\ Ratio = \frac{a/b}{c/d} = \frac{a \times d}{b \times c}$$

Figure 1. Calculating Odds Ratio (OR)

The total youth sample size for YRBS was 17,232 youth with 48.3% being female and 51.7% male. White youth accounted for 50.7%, black 12.1%, Hispanic 25.4% and other 11.8%.

3. Results

The analysis from the YRBS, indicated that 36.9% of youth reported mental health was poor during the Covid-19 pandemic. Compared with those without mental health issues, those with mental health issues were generally

Vol. 2023 (2) 312 – 317
ISSN 2688-3651 [online]

Table 1. Mental Health and Sex

Reported Mental Health was most of the time or always not good during the Covid-19 Pandemic			
Sex	Percentage	95% Confidence Interval	P value
Females	51.00	48.42 – 53.57	0.00
Males	23.71	21.89 – 25.63	

younger and female, less likely to be Native Hawaiian or other Pacific Islander, be less physically active and overweight.

Female students (51%) were more likely to have poor mental health than male students (23%), p = 0.00

Comparison by race: White students were more likely to have poor mental health than Black, Asian students (p = 0.00) and Hispanic students. (p =0.02) There was no statistically significant difference in poor mental health between white and American Indian/Alaska Native students (p = 0.72) There is no statistical difference since the 95% confidence intervals overlap. Only when 95% confidence intervals do not overlap is there a statistically significant difference.

Using 95% confidence intervals, the only grades that had statistically significance differences in reporting poor mental health was between 9th graders and 11th, and 9th and 12th graders. 9th graders were less likely to report poor mental health compared to 11th and 12th graders.

Table 2. Mental Health and Race

Reported Mental Health was most of the time or always not good during the Covid-19 Pandemic		
Race	Percentage	95% Confidence Interval
American Indian/Alaska Native (AI/AN)	41.49	30.90 – 52.93
Asian	29.78	27.50 – 32.16
Black	30.16	27.24 – 33.26
Hispanic	35.76	34.25 – 37.30
White	39.53	36.81 – 42.31
Race	P Values	
White & Black	0.00	
White & Asian	0.00	
White & Hispanic	0.02	

Table 3. Mental Health and Grade Level

Reported Mental Health was most of the time or always not good during the Covid-19 Pandemic		
Grade Level	Percentage	95% Confidence Interval
9	31.71	29.28 – 34.24
10	36.97	34.17 – 38.85
11	40.56	38.16 – 43.00
12	39.06	36.01 – 42.20
Grade		P Values
9 & 10		0.00
9 & 11		0.00
10 & 11		0.04
11 & 12		0.47

9th graders (32%) were less likely than 11th graders (40%) to report poor mental health (p = 0.00) 10th graders more likely to have poor mental health than 9th graders. (p = 0.00) 11th graders (41%) more likely to have poor mental health than 10th graders (37%), (p = 0.04) 12th graders (39%) were not statistically different than 11th graders (41%) in likelihood to have poor mental, (p = 0.47)

Table 4. Overweight Youth by Sex

Reported Mental Health was most of the time or always not good during the Covid-19 Pandemic		
Sex	Percentage	95% Confidence Intervals
Female	17.36	15.95 – 18.88
Male	14.81	13.82 – 15.86

Table 5. Overweight Youth by Race

Reported Mental Health was most of the time or always not good during the Covid-19 Pandemic		
Race	Percentage	95% Confidence Interval
American Indian/Alaska Native (AI/AN)	13.97	7.75 – 23.88
Asian	8.29	4.29 – 13.61
Black	18.65	16.96 – 20.46
Hispanic	20.88	19.68 – 22.14
White	14.57	13.51 – 15.71
Race	P Values	
White & Asian	0.00	
White & Black	0.00	
White & Hispanic	0.00	
White & AI/AN	0.89	

White youth (14.57%) were more likely to be overweight than Asians (8.29 %), (p = 0.00). Black youth (18.65%) were more likely to be overweight than White youth (14.57%), (p = 0.00). Hispanic youth 20.88% were more likely to be overweight than White youth (14.57%), (p = 0.00). There was no statistically significant difference between White youth (14.57%) than AI/AN youth (13.97%), (p = 0.89)

Vol. 2023 (2) 312 – 317
ISSN 2688-3651 [online]

White youth (13.69%) were more likely to be obese than Asians (7.69%), (p = 0.00). Black youth (21.24%) were more likely to be obese than White youth (13.69%), (p = 0.00). Hispanic youth (20.2%) were more likely to be obese than White youth (13.69%). (p = 0.00). AI/AN youth (29.4%) were more likely to be obese than White youth (14.57%), (p = 0.03).

Table 6. Obesity by Race

Reported Mental Health was most of the time or always not good during the Covid-19 Pandemic		
Race	Percentage	95% Confidence Interval
American Indian/Alaska Native (AI/AN)	29.43	17.60 – 44.87
Asian	7.69	4.97 – 11.71
Black	21.24	18.48 – 24.28
Hispanic	20.23	17.87 – 22.81
White	13.69	11.98 – 15.61

Race	P Values
White & Asian	0.00
White & Black	0.00
White & Hispanic	0.00
White & AI/AN	0.03

Table 7. Mental Health and Physical Activity

Reported mental health was most of the time or always not good	Were not physically active for at least 60 minutes	
	Yes	No
Yes	803	2870
No	1365	7579

Odds Ratio (OR) = 1.55
Odds Ratio of 1.55 indicates increased risk for less physical activity among those who reported poor mental health.

Table 8. Mental Health and Physical Activity Among Black Males

Reported mental health was most of the time or always not good	Were not physically active for at least 60 minutes	
	Yes	No
Yes	44	117
No	160	716

OR=1.68
Odds ratio of 1.68 indicates increased risk among black males for poor mental health among those who reported they were not physically activity for at least 60 minutes on at least 1 day.

Table 9. Mental Health and Physical Activity Among White Females

Reported mental health was most of the time or always not good	Were not physically active for at least 60 minutes	
	Yes	No
Yes	241	989
No	201	1419

OR=1.72
The odds ratio of 1.72 indicates increased risk among white females for poor mental health who were not physically activity for at least 60 minutes on at least 1 day.

Table 10. Mental Health during Covid and Physical Activity for American Indian/Alaska Native Females

Reported mental health was most of the time or always not good during COVID-19 pandemic	Were not physically active for at least 60 minutes	
	Yes	No
Yes	703	2798
No	899	5254

OR= 1.47
This indicates an increased risk among AI/AN females for poor mental health during the Covid pandemic among those who were not physically active for at least 60 minutes on at least 1 day.

Table 11. Mental Health during Covid and Physical Activity for Asian Females

Reported mental health was most of the time or always not good during COVID-19 pandemic	Were not physically active for at least 60 minutes	
	Yes	No
Yes	36	989
No	26	1419

OR = 2.38
This indicates an increased risk among Asian females for poor mental health during the Covid pandemic among those who were not physically active for at least 60 minutes on at least 1 day

Table 12. Mental Health during Covid and Physical Activity for White Females

Reported mental health was most of the time or always not good during COVID-19 pandemic	Were not physically active for at least 60 minutes	
	Yes	No
Yes	221	956
No	116	894

OR = 1.78
The odds ratio of 1.78 Indicates increased risk for poor mental health among white females during Covid pandemic for those who are not physically active for at least 60 minutes on at least 1 day.

Vol. 2023 (2) 312 – 317
ISSN 2688-3651 [online]

Poor mental health and not being physically active were associated and statistically significant (p = 0.000) for youth in YRBS but when controlled for my race and sex only White females, Asian females p = 0.000 and AI/AN females showed a statistically significant association between poor mental health during Covid-19 and low physical activity.

4. Discussion

This study analyzed most recent 2021 YRBS survey data. In this large survey of high school students, the prevalence of poor mental health was 29.3%. As seen in previous research, females reported poor mental health more than males. 17.36% of females and 14.82% of males were overweight. Current mental health 40.77% of females and 18.08% of males reported that their mental health was most of the time or always not good. Fifty one percent of females and 23.71% of males reported their mental health was most of the time or always not good during the Covid-19 pandemic. This is consistent with findings from Reinherz, et al 1993. "Moreover, consistent with national-level data, females in this population exhibited a significantly greater risk for both mental and physical health issues than males."[3] In addition, "… studies found relationships among physical activity, sedentary behavior, and depression, but more recent information is needed to inform research and practice." (Das, et al., 2016)

Controlling for race and gender, the only statistically significant groups were who were found to have an association between not being physically active and poor mental health were Black males, white females, and Native Hawaiians. AI/AN youth 29.4% were more likely to be obese than White 14.57% youth (p=0.03) During the pandemic, there was an increased risk for poor mental health among those who were not physically active for at least 60 minutes on at least 1 day among American Indian/Alaska Native females, Asian females and white females. The strength of the association (odds ration =2.38) was highest among Asian females during Covid-19 pandemic. Poor mental health and not being physically active were associated and statistically significant for youth in YRBS but when controlled for by race and sex. only White females, Asian females p=0.000 and AI/AN females showed a statistically significant association between poor mental health during Covid-19 and low physical activity.

According to Reinherz, et al., (1993), "there is a need to report differential effects for gender, age groups, socioeconomic status, and geographic settings since the impact of mental health interventions might vary according to various contextual factors." This study controlled for race and gender and found statistically significant results.
Moreover, these findings on the association between mental health and physical activity are consistent with research by Wang and Peiper published in Preventing Chronic Disease. Wang and Peiper (2022) found that "Inadequate physical activity and excessive sedentary behavior are associated with depressive symptoms among US high school students." (Wang and Peiper, 2022) According to Reinherz, et al., (1993) "approximately one in every four to five youth in the U.S. meets criteria for a mental disorder with severe impairment across their lifetime. The likelihood that common mental disorders in adults first emerge in childhood and adolescence highlights the need for a transition from the common focus on treatment of U.S. youth to that of prevention and early intervention."
The results demonstrate an association between poor mental health and the lack of physical.

Given that over 29% of youth had poor mental health and 37% reported poor mental health during covid, these findings point for the need for improved school-based services in the area of mental health. Therefore, there is a need for schools to assess youth for mental health and provide mental health services or referral systems to community resources. Results of community surveys in regions of the United States have shown one in every four children experiences a mental disorder with few affected youth receiving adequate mental health care. There has been a lack of data on the prevalence and distribution of a wide range of mental disorders from a nationally representative sample of children or adolescents. This information is necessary to establish resource allocation priorities for prevention and treatment. (Cohen, et al.,1993) (Reinherz, et al., 1993) (Lahey, et al.,1996) (Braumer and Stephens, 2006)

Further research is needed to identify specific causes of poor mental health and identification of effective early interventions for at risk youth to prevent poor physical activity and related health outcomes. The findings in this study that mental health is associated with physical health in youths suggest the need for a closer look at this population and their risks for health problems. Future research should focus on tracking incidence and prevalence of mental health in youth, identify effective treatment and intervention strategies that promote mental and physical health of youth.

Vol. 2023 (2) 312 – 317
ISSN 2688-3651 [online]

Acknowledgements

Thanks to my parents for encouraging me to research the important topic of mental health.

References

Brauner, C., and Stephens, C. (2006). Estimating the prevalence of early childhood serious emotional/behavioral disorders: Challenges and recommendations. Public Health Rep, 121.

Centers for Disease Control and Prevention. (2021) Youth Risk Behavior Survey. Available at: www.cdc.gov/YRBSS. Accessed on 6/1/2023.

Centers for Disease Control and Prevention. (2021) Youth Risk Behavior Survey Data. Available at: www.cdc.gov/yrbs. Accessed on 6/1/2023.

Centers for Disease Control and Prevention. (2023, June 1) What is children's mental health Medium. https://www.cdc.gov/mentalhealth/children/index.htm? Accessed on 6/1/2023.

Chen, H., et al. (2006) Adolescent axis I and personality disorders predict quality of life during young adulthood. Journal of Adolescent Health. 2006;39:14–19.

Cohen, P., et al. (1993). An epidemiological study of disorders in late childhood and adolescence—I: Age- and gender-specific prevalence. Journal of Child Psychology Psychiatry, 34.

Das J., et al. (2016) Interventions for Adolescent Mental Health: An Overview of Systematic Reviews. Journal of Adolescent Health. 2016;59:S49–S60. doi: 10.1016.

Lahey, B., Flagg, E., and Bird, H. (1996). The NIMH Methods for the Epidemiology of Child and Adolescent Mental Disorders (MECA) Study: Background and methodology. Journal of American Academy of Child Adolescent Psychiatry, 35.

Merikangas K., et al. (2010). Lifetime Prevalence of Mental Disorders in US Adolescents: Results from the National Comorbidity Study-Adolescent Supplement. Journal of American Academy of Child Adolescent. Psychiatry. 2010;49:980–989. doi: 10.1016/j.jaac.2010.05.017.

Reinherz, H., et al. (1993.) Prevalence of psychiatric disorders in a community population of older adolescents. Journal of American Academy of Child Adolescent Psychiatry, 32.

Sanyaolu A., et al. (2019) Childhood and Adolescent Obesity in the United States: A Public Health Concern. Global Pediatric Health. 2019;6.

Wang, C., and Peiper, N. (2022). Association Between Physical Activity and Sedentary Behavior With Depressive Symptoms Among US High School Students, 2019 Preventing Chronic Disease November 17 2022;19:E76

Journal of Research High School

Vol. 2023 (2) 318 – 327
ISSN 2688-3651 [online]

Predicting Breast Cancer Diagnoses using Supervised Classification Models

Kyle Wang[1] *

[1]Radnor High School, Radnor, PA, USA
*Corresponding Author: kylewang122205@gmail.com

Advisor: Daniela Perry, dsperry@ucsd.edu

Received June 14, 2023; Revised September 23, 2023; Accepted, October 17, 2023

Abstract

Machine learning has many applications in the healthcare industry with the potential to save lives, one of which is detecting and diagnosing diseases based on images or predicting the likelihood of breast cancer given gene expression data. As a result, researchers have considered using machine learning techniques for faster diagnoses, which is critical for diseases like cancer when early detection can lead to a better prognosis. This study utilized the impact of supervised classification models, RNA-seq data from control patients and breast cancer patients. Gene expression read counts were subsequently normalized during the exploratory data analysis phase and split into training and testing data to create models that would help doctors draw conclusions about the presence of breast cancer. The study then introduced a separate validation set, to which the model could be used to predict a diagnosis. The paper explored various techniques to improve accuracy, such as reducing the number of significant genes used, altering the hyperparameters of each model, and normalizing data with a zero-inflated negative binomial distribution. The research yielded results with a maximum accuracy of 90.1% was obtained with both logistic regression models, and their performances were further analyzed using sklearn (Python machine learning tool) metrics. The models also discovered that patients with the gene markers ENSG00000201908, ENSG00000216184, and ENSG00000221326 exhibited the greatest variation in gene counts between breast cancer patients and control patients, which could be worth exploring in future studies.

Keywords: Machine learning, Neural network, Logistic regression, Random forest

1. Introduction

Twelve percent of women in the United States will be diagnosed with breast cancer at some point in their lives (Waks & Winer, 2019). Though somewhat common, breast cancer remains difficult to predict and prevent. Early diagnosis is one of the best methods for a full recovery (Sun et al., 2017). However, for low-income countries, equipment and resources for detecting breast cancer are not readily available, which decreases the rate of survival. Using various models in machine learning, trends based on genetics and demographics can be used to create models that predict the likelihood that someone has breast cancer. Using models to predict breast cancer is not new; using mammogram images to predict breast cancer has yielded accuracies as high as 94.20% (Lin et al., 2022). If one could predict the occurrence of breast cancer based on quantitative gene counts even before the mammogram imaging that occurs at an older age, people could take more intentional actions in reducing future breast cancer risk.

The data and demographics containing both breast cancer patients and control patients comes from a public dataset that was previously used for different purposes. The genes were sequenced and analyzed through a process known as RNA-seq, which allows us to measure RNA transcripts that are transcribed. The transcribed genes are broken into small cDNA segments, upon which adaptors are attached, further allowing us to duplicate these segments through a polymerase chain reaction (PCR) and analyze accordingly (Wang, Gerstein, & Snyder, 2009). When comparing

Vol. 2023 (2) 318 – 327
ISSN 2688-3651 [online]

expression levels between breast cancer patients and normal patients, a significant difference in gene counts could often be found and used to create a model for prediction.

The best prediction type of model that can be used to analyze this data is the supervised classification model, meaning the model has predetermined and labeled input values (specific genes and respective gene counts), and the output values are discrete variables (presence or absence of breast cancer). Three main classification models were used to find a relationship between genes and diagnosis: neural networks, logistic regression, and a random forest. Neural networks randomly select variables and add random weight values to calculate a predicted value, which is compared to the actual value to determine whether the prediction was accurate. If not, the model begins backpropagation, regenerating new weights for the variables and repeating the process for many epochs until the model converges. Logistic regression can be used with data with binary outcomes, converting continuous probabilities into a discrete 0 or 1. Random forests generate many decision trees and randomly selecting variables, maintaining trees that predict most accurately in the model. With many decision trees, the chance of having just one strong predictor is very high (Ren, Cheng, & Han, 2017). After determining the probability of a patient having breast cancer, a diagnosis can be made based on which outcome is more likely. These models will be employed to predict the probability of the presence of breast cancer based on gene expression data.

2. Materials and Methods

The data for this project was obtained through SILVER-seq (Small Input Liquid Volume Extracellular RNA Sequencing), which utilizes extracellular RNA found in human serum to compare expression levels of about 25% of the human genes (Zhou et al., 2019). The original training data consisted of 128 patients: 96 of the patients were breast cancer patients, while 32 of them were normal control patients. Using this dataset, various models created their parameters based on the gene counts for each patient. Molded from the training data, the new models were later compared to a validation set to determine how well each model predicts breast cancer diagnosis of a new set. For the validation data set, there were 161 total patients: 86 breast cancer patients and 75 normal control patients. Throughout this study, both training accuracies and validation accuracies were compared: the classification processes strive to find models for the gene count data that would correctly predict the breast cancer diagnoses from the training set to obtain a high training accuracy. However, generating a high validation accuracy is more important, as the previous model is then used on a new validation dataset to check the applicability of the model in a slightly different context. The challenge is generating a model with high training and validation accuracy, meaning the model cannot overfit the training data so heavily that it is useless and inaccurate in a slightly different context.

2.1 Neural Network

In order to effectively analyze high-dimensional data, the deSEQ-2 package, a program in the Python programming language that allows for easier processing of raw data like gene counts for use in a supervised model, was used to normalize and visually display the gene expression data. In the first iteration of the neural network, the dataset for breast cancer patients and normal patients were imported and transposed in order to ensure that each patient served as the sample with the characteristics of gene counts. Lots of data cleaning was required to remove the unnecessary data and add a diagnosis column so that the model can compare the model results to the actual diagnosis to determine the proportion of diagnoses that was correctly predicted. Gene count values were also converted from integer values to proportions, so that larger SILVER-seq samples won't affect the dataset. Using the standard scaler normalization from the sklearn package, data was inputted into a sequential neural network model. This model contained 4 dense models with 16, 8, 4, and 2 neurons, with the final output layer having 1 neuron and a sigmoid activation. One hundred epochs of training were run, with gradual improvement in performance after each epoch (high val_acc and low val_loss). After fitting this model into a new test dataset, the model could now try to predict the validation set.

In iteration two of the neural network, it was observed that some genes seemed to be extraneous, and the addition of these genes could add extra variables into the model that would only ruin its performance. As a result, genes were

Vol. 2023 (2) 318 – 327
ISSN 2688-3651 [online]

filtered and only included genes with the most drastic count difference between normal patients and breast cancer patients (having a p-value less than 0.05 and suggesting a strong correlation between the gene and the occurrence of breast cancer), which narrowed the number of genes from 60675 to 2109. The neural network model's accuracy improved by 14.9% with this change.

2.2 Logistic Regression

The process also involved experimenting with other model types, such as the logistic regression model. The data cleaning process was similar to that of the neural network, with a similar train-test split of 0.50, meaning half of the data values are split between the training and test data. The implementations of the actual models were as simple as importing the logistic regression function from the sklearn linear model class and fitting the initial model using the gene counts and diagnosis of the 64 training values. Comparing the model-predicted diagnoses of the test data and the actual diagnoses had an 100% accuracy, but a comparison to a different validation set was necessary to ensure that the model was not overfit and feasible in a slightly different context. The validation set's predicted results were compared to its actual results, and the accuracy was slightly less than 100%. Therefore, the models were not a perfect fit, but the data was also not significantly overfit.

2.3 Random Forest

The random forest model had a similar initial setup, but the sklearn ensemble model class imported the random forest regressor model. The model had two hyperparameters that were controlled: n_estimators and random_state. In a random forest model, the n_estimators determine the number of trees that will be included in the random forest, with a greater number of trees typically correlated with a greater accuracy but slower model loading times. However, adding more trees also makes the model more prone to overfitting, leading to a lower accuracy when comparing the training data model to a different validation set. Choosing 20 trees seemed to maintain the highest accuracy without overfitting. The random_state keeps the end accuracy relatively consistent after every trial.

2.4 Zero-inflated Negative Binomial Normalization

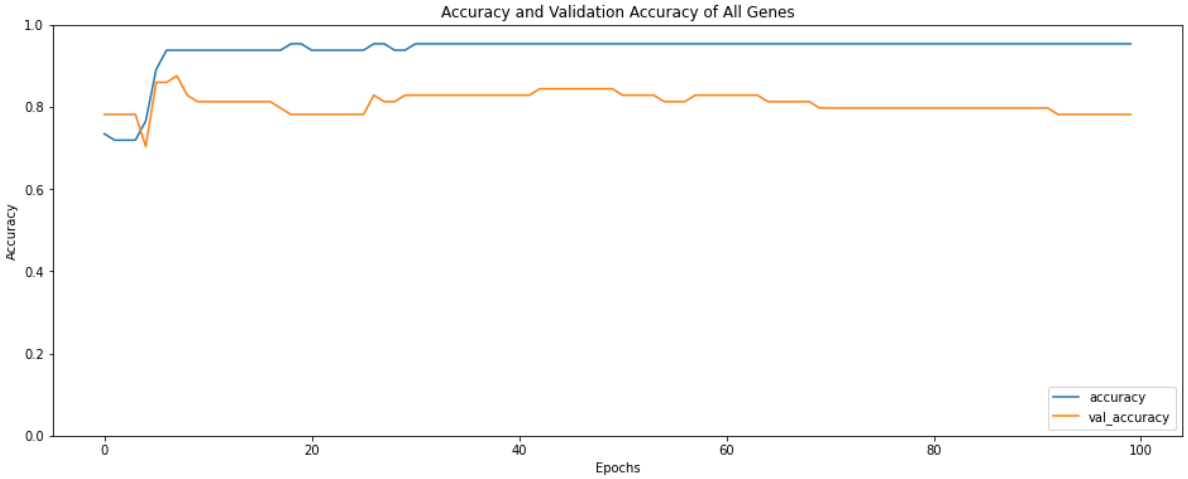
In the third iteration, a new normalization technique is employed that could potentially increase accuracy. A new normalization method – zero-inflated negative binomial normalization – was used, in which the excess zeros are modeled independently of the rest of the data. This normalization is used often in scientific research in overdispersed data that contains lots of zeros, as the zeros increase the amount of variability, which distorts the standard deviation and prevents an accurate normalization from being created (Alam, Al Mahi, & Begum, 2018). This new normalization method created slightly different normalized values with varying results in its predictions using the neural network model, logistic regression model, and random forest model.

3. Performance and Results

The model prediction of the training set data for the neural network model tracked accuracy and validation accuracy most closely. The first neural network model iteration that contained 60675 different genes had the highest training accuracy of 95.31%, but the validation accuracy never reached higher than 78.12%, decreasing as the model ran, which is a sign of overfitting. Admittedly, final accuracy and validation accuracy for all neural networks had variation due to the nature of neural networks, especially the original model, as 60,675 variables are randomly assigned weights with a small sample size to which the models are trained and validated. Using the same model for the significant genes only showed both a higher training accuracy of 100%, but the validation accuracy also continued to increase, with a maximum test accuracy of 96.88%. This contrast can be viewed more clearly in figure 1.

Vol. 2023 (2) 318 – 327
ISSN 2688-3651 [online]

1a



1b

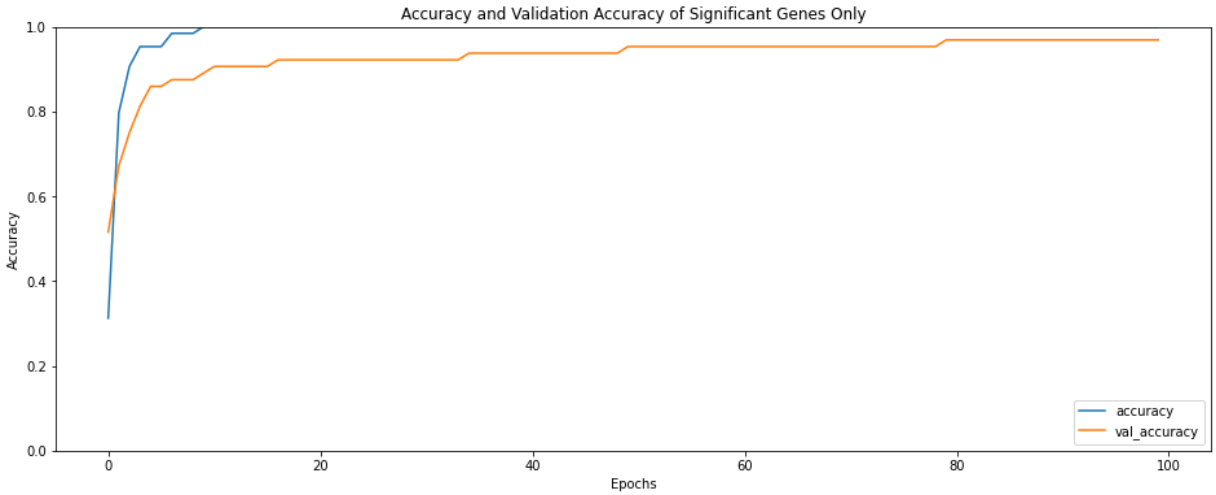


Figure 1: Comparison of accuracy and validation accuracy increase in neural networks with 1a: 60,675 genes (top) to 1b: 2,109 genes (bottom)

The first iteration of the neural network contained too many different types of gene counts to yield an accurate neural network model, resulting in a test accuracy of 55.9% - just barely better than guessing. However, the effect of reducing the number of variables to include only the significant gene was massive, as the accuracy rate increases to 70.8% for the neural network model with significant genes only.

At this stage, the random forest and logistic regression models were also introduced, and accuracy rates were markedly higher than the neural network rate, correctly predicting the breast cancer diagnosis 82.0% and 90.1% of the time, respectively.

Although higher accuracies for the zero-inflated negative binomial distribution were expected, accuracies actually seemed to remain relatively the same, if not worse. Using this normalization for both the training and validation sets, decreased the neural network model to an accuracy of 62.1%, slightly increased the random forest model's test accuracy to 82.6%, and kept the logistic regression's test accuracy at the same value of 90.1%. The various validation accuracies for each iteration are shown in table 1.

Vol. 2023 (2) 318 – 327
ISSN 2688-3651 [online]

3.1 Confusion Matrix

Table 1: Final validation accuracies in seven models used.

Model	Accuracy Scores
Iteration 1: Neural Network (60,675 genes)	55.9%
Iteration 2a: Neural Network (2,109 genes)	70.8%
Iteration 2b: Random Forest (2,109 genes)	82.0%
Iteration 2c: Logistic Regression (2,109 genes)	90.1%
Iteration 3a: Neural Network (Zero-inflated negative binomial normalization)	62.1%
Iteration 3b: Random Forest (Zero-inflated negative binomial normalization)	82.6%
Iteration 3c: Logistic Regression (Zero-inflated negative binomial normalization)	90.1%

The accuracies of each model can be visually represented using a confusion matrix. Each prediction made by the model is split into four categories. The top-left box represents true-negative, meaning that the patient was correctly predicted to be a normal control patient, and the bottom-right box represents true-positive, meaning that the patient was correctly predicted to be a breast cancer patient. The top-right box represents a false-positive result, in which the model predicts that the patient has breast cancer, but the actual diagnosis of the patient is normal. Finally, the bottom-left box represents false-negative, in which the model predicts that the patient is normal when the patient actually does have breast cancer (Scikit-learn Developers., n.d.). Figure 2 shows the accuracy of each prediction made by each model in a visual manner, with colors indicating the number of predictions that fell under each category. Accuracy can be determined from the matrix by summing the values in the top-left and bottom-right boxes and dividing that value by the sum of all four values. An ideal model should have the majority of samples in the top-left and bottom-right boxes.

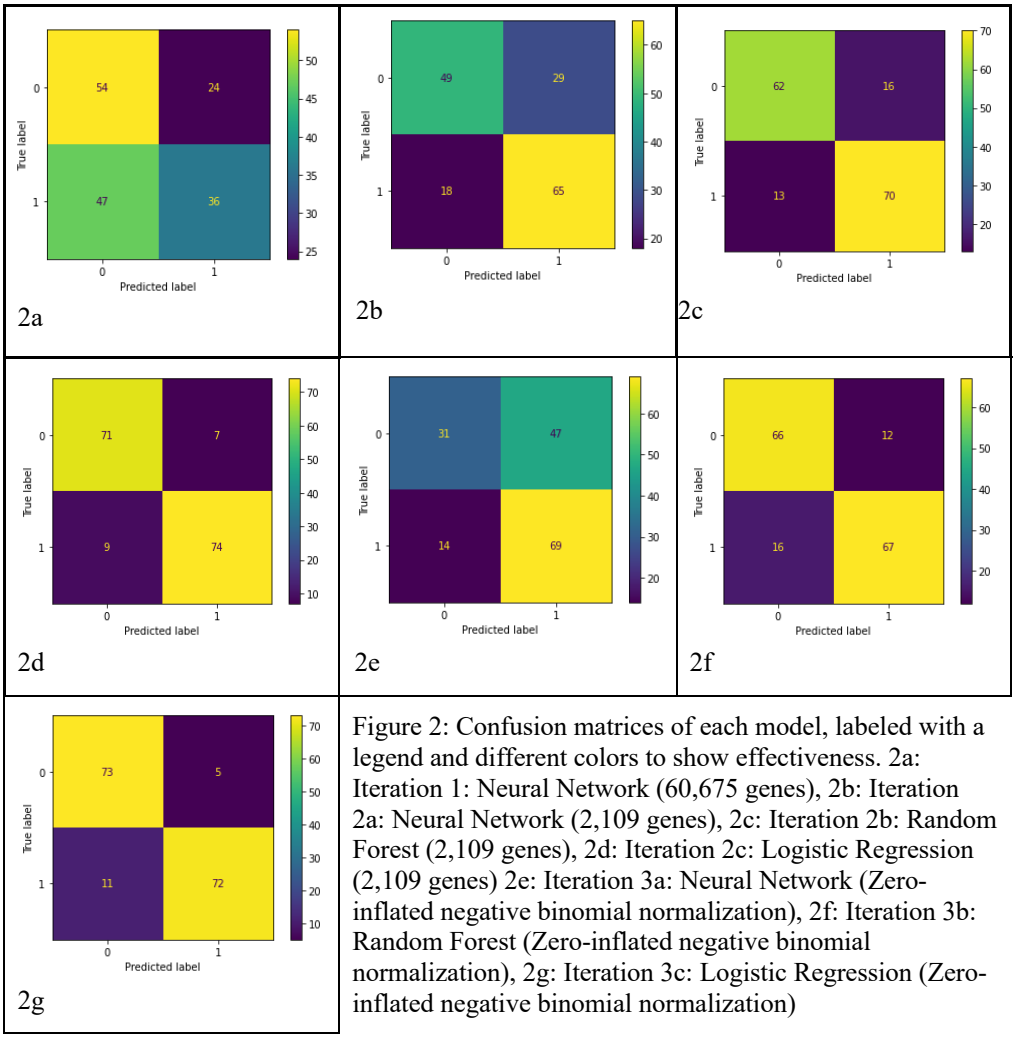


Figure 2: Confusion matrices of each model, labeled with a legend and different colors to show effectiveness. 2a: Iteration 1: Neural Network (60,675 genes), 2b: Iteration 2a: Neural Network (2,109 genes), 2c: Iteration 2b: Random Forest (2,109 genes), 2d: Iteration 2c: Logistic Regression (2,109 genes) 2e: Iteration 3a: Neural Network (Zero-inflated negative binomial normalization), 2f: Iteration 3b: Random Forest (Zero-inflated negative binomial normalization), 2g: Iteration 3c: Logistic Regression (Zero-inflated negative binomial normalization)

The confusion matrix in figure 2 is further labeled with colors that indicate the number of samples in each category, with yellow being the greatest and dark purple being the least. In a strong model like the one in figure 2g, a mix of yellow and dark purple indicates that the model correctly predicts the diagnosis most of the time and rarely makes mistakes. Looking at both the accuracy and the confusion matrix, one can see that the logistic regression models are the strongest, while the neural networks are

Vol. 2023 (2) 318 – 327
ISSN 2688-3651 [online]

the weakest because there are lots of blue colors in the matrices, indicating that the matrix is about equally likely to predict correctly as incorrectly, resulting in a lower accuracy.

The performance of each model can be further investigated by utilizing various classification metrics.

3.2 Receiver Operating Characteristic Curve

Exploration began with the ROC curve and the ROC_AUC score. The ROC curve, which stands for a receiver operating characteristic curve, plots the sensitivity of the model compared 1 - specificity on the x-axis (Mandrekar, 2010). The sensitivity model represents how often the model correctly predicts an outcome correctly marked as positive or having a "cancer" diagnosis, compared to the total number of positive predictions, which would include normal patients who were predicted to have breast cancer. In contrast, the specificity model represents an outcome that is correctly marked as negative with a "normal" diagnosis, compared to the total number of negative predictions, which includes cases in which a patient with breast cancer is predicted to not have breast cancer.

It's often difficult to have a model with perfect sensitivity and specificity, so the model should typically prioritize one over the other. In this case, it's better to take precautions and do tests on a patient without breast cancer versus neglecting a breast cancer patient by claiming that they do not have cancer. Having a false negative is not ideal, so a model with higher sensitivity, or recall, is preferred (Lalkhen & McCluskey, 2008).

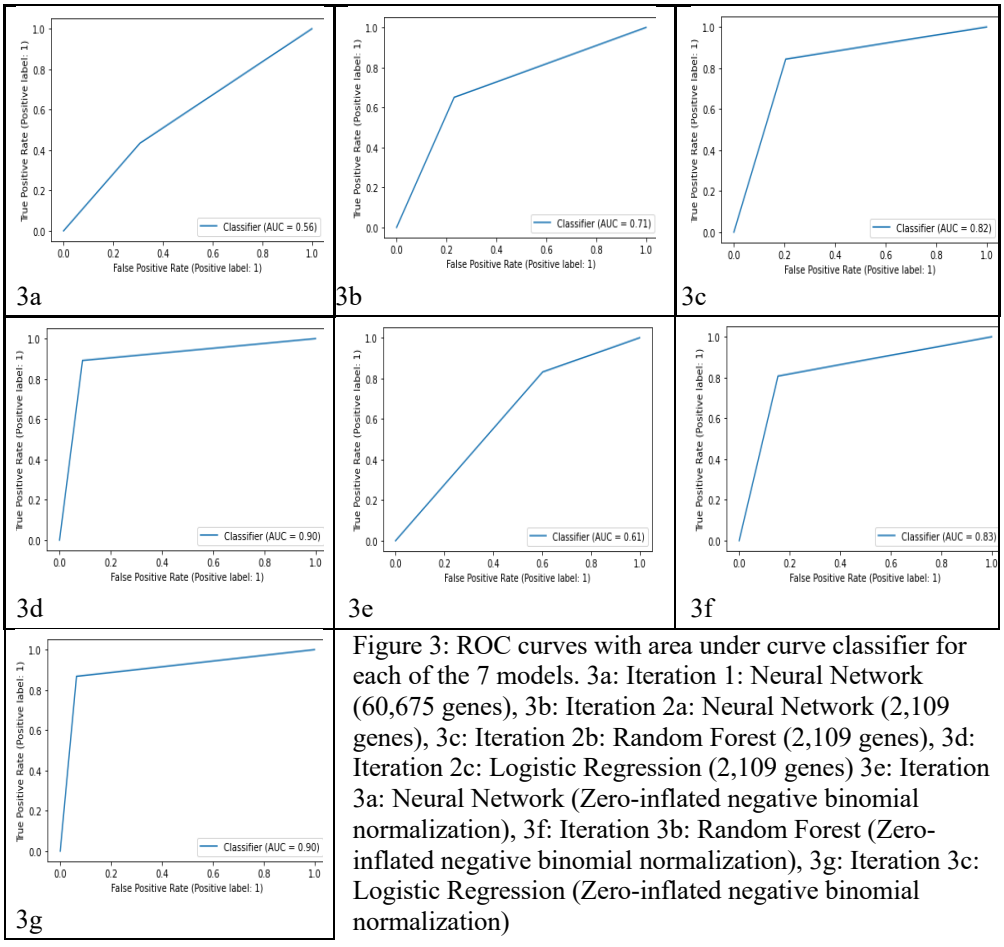


Figure 3: ROC curves with area under curve classifier for each of the 7 models. 3a: Iteration 1: Neural Network (60,675 genes), 3b: Iteration 2a: Neural Network (2,109 genes), 3c: Iteration 2b: Random Forest (2,109 genes), 3d: Iteration 2c: Logistic Regression (2,109 genes) 3e: Iteration 3a: Neural Network (Zero-inflated negative binomial normalization), 3f: Iteration 3b: Random Forest (Zero-inflated negative binomial normalization), 3g: Iteration 3c: Logistic Regression (Zero-inflated negative binomial normalization)

The ROC curves for the models in figure 3 were generated by comparing results from the actual diagnoses in the validation set and the predicted diagnoses in the validation set according to the corresponding model. The curves in these models are composed of two line segments with different slopes, beginning at (0.0, 0.0) and ending at (1.0, 1.0)

Vol. 2023 (2) 318 – 327
ISSN 2688-3651 [online]

in all cases. Models, like the one seen in figure 3d, that start with a large slope are typically better-performing curves, whereas curves with a generally constant slope like 1 are poorer-performing curves. The ROC_ AUC score (located in the bottom left of each graph) also ranges from values between 0.5 and 1, with numbers closer to 1 also indicating better performance (Hoo, Candlish, & Teare, 2017).

3.3 Recall and Precision

The recall and precision are two other metrics that are worth noting. The recall is the same as sensitivity, measuring the number of breast cancer patients who were correctly predicted as having breast cancer, as some breast cancer patients were predicted to have a normal diagnosis, causing a type II error, as seen in equation 1.

$$TP \,/\, (TP + FN) \qquad \text{eq. (1)}$$

where TP = true positives and FN = false negatives

On the other hand, precision measures the number of breast cancer patients who were correctly predicted as having cancer as a percentage of the total number of patients who were given a positive diagnosis, causing a type I error, as seen in equation 2.

$$TP \,/\, (TP + FP) \qquad \text{eq. (2)}$$

where TP = true positives and FP = false positives

As mentioned previously, a type II error would be more detrimental in this model, so the number of false negatives should be mitigated, so recall and sensitivity should ideally have a proportion closest to 1.

Table 2 highlights the precision and recall values found after generating each model. As expected, model 1 performed poorly with the lowest precision and recall of 60.0% and 43.3%. This model is extremely prone to predicting normal diagnoses in patients with breast cancer, which would be dangerous, since people would not receive the proper treatment. Iteration 2c had the highest precision and recall rate of 91.4% and 89.2%, respectively, meaning out of these seven models, logistic regression currently looks like the best model to predict breast cancer. It should be noted that iterations 2a and 3a, neural network models, have recall percentages that are significantly higher than the precision percentages: in 3a, the difference is 23.6%. Neural networks that prioritize recall over precision would be more likely to have false positives that lead to a higher cost of additional testing but more breast cancer screening that ensures that patients are healthy.

Table 2: Precision and recall scores for each of the seven models.

Model	Precision	Recall
Iteration 1: Neural Network (60,675 genes)	60.0%	43.3%
Iteration 2a: Neural Network (2,109 genes)	69.1%	78.3%
Iteration 2b: Random Forest (2,109 genes)	81.3%	84.3%
Iteration 2c: Logistic Regression (2,109 genes)	91.4%	89.2%
Iteration 3a: Neural Network (Zero-inflated negative binomial normalization)	59.5%	83.1%
Iteration 3b: Random Forest (Zero-inflated negative binomial normalization)	84.8%	80.7%
Iteration 3c: Logistic Regression (Zero-inflated negative binomial normalization)	93.5%	86.7%

The precision recall curve for each iteration in figure 4 visually represents the relationship between precision and recall and is a metric that is suitable for skewed breast cancer gene count data, in which the proportions are typically zero or very close to zero (Davis & Goadrich, 2006). The precision is labeled on the y-axis, and the recall is labeled on the x-axis, with the graph starting from the top-left part of the graph and ending at the bottom-right part of the

Vol. 2023 (2) 318 – 327
ISSN 2688-3651 [online]

graph. In an ideal situation, both the precision and recall should be high, so the precision-recall graph should have a horizontal line as high on the graph as possible.

The classifiers in these graphs are found using average precision, which is the area under the curve (Kielwagen, Grosse, & Grau, 2014). A higher number correlates with a better model in terms of the precision and recall metric.

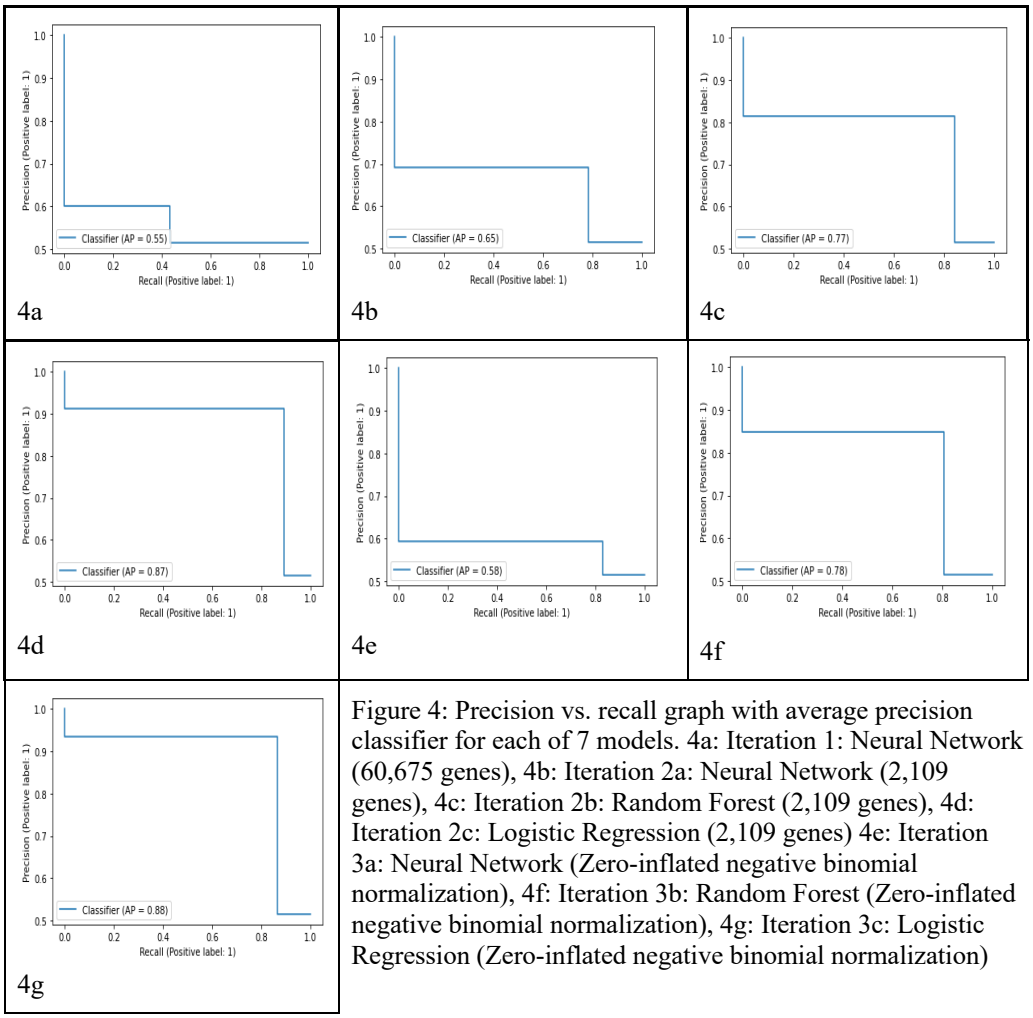


Figure 4: Precision vs. recall graph with average precision classifier for each of 7 models. 4a: Iteration 1: Neural Network (60,675 genes), 4b: Iteration 2a: Neural Network (2,109 genes), 4c: Iteration 2b: Random Forest (2,109 genes), 4d: Iteration 2c: Logistic Regression (2,109 genes) 4e: Iteration 3a: Neural Network (Zero-inflated negative binomial normalization), 4f: Iteration 3b: Random Forest (Zero-inflated negative binomial normalization), 4g: Iteration 3c: Logistic Regression (Zero-inflated negative binomial normalization)

As anticipated from previous metrics, the original neural network with 60,675 genes performed the worst, as the precision drops to 0.6 immediately, and the average precision is the lowest of the seven models at 0.55. Both logistic regression models performed well, as average precision was the highest at 0.87 and 0.88 in figures 4d and 4g, respectively. Average precision is correlated with accuracy, as the two metrics never deviated by more than 0.05 for each corresponding model, so this is an effective metric for this dataset.

4. Discussion

4.1 Limitations

Although the models seem to have lower numbers, the zero-inflated negative binomial distribution did not effectively normalize the data because of the nature of the zeros. The use of this model typically assumes that both structural zeros (zeros that occur because of some restriction that forces a value to be null) and random zeros (zeros that occur randomly in the dataset without possible restrictions) are present in the model. All genes are present in each

Vol. 2023 (2) 318 – 327
ISSN 2688-3651 [online]

patient, so all zeros that occur in the sample are random and not structural, making the use of this model questionable (Hawinkel et al., 2020). To further increase accuracy rate, it is necessary to try to find a better normalization technique that makes sense in the context of the data. A larger sample size should also be considered, as having only 128 samples limits the amount of data that can be used to more effectively train the model. A larger sample size would also mitigate the amount of standard deviation that made accuracies in each trial slightly different.

4.2 Future Directions

While many genes were considered in combination to generate the model, it is important to take note the gene counts with the greatest difference between breast cancer patients and normal patients: ENSG00000201098, ENSG00000216184, ENSG00000221326, all of which had the lowest p-value. ENSG00000201098 is also referred to as the RNY1 gene that plays a role in chromosomal DNA replication, which makes sense as cancer is often caused by mutations in the DNA (GeneCaRNA., 2023, May 22). Future investigation into these genes may discover additional links to breast cancer that could serve as the starting point for developing a drug that could combat it.

5. Conclusion

Artificial intelligence has made significant advancements in the healthcare industry in terms of diagnosing diseases like cancer and finding ideal drugs and solutions to remedy these diseases (Davenport & Kalakota, 2019). In addition to the neural networks, logistic regression, and random forest models that can be used, machine learning can continue to be optimized to convert data and numbers on a spreadsheet into actionable steps that can help machines make the most reasonable and unbiased diagnoses and treatments. Though hyperparameters and normalizations could be further optimized, the logistic regression models had the highest accuracies of 90.1% when determining breast cancer diagnoses from gene counts. Normalizing the logistic regression models with the standard scaler or using the zero-inflated negative binomial distribution had no significant effect on the performance, as verified by metrics such as precision, recall, and receiver operating characteristic curves.

References

Alam, M., Al Mahi, N., & Begum, M. (2018). Zero-Inflated Models for RNA-Seq Count Data. *Journal Articles: Biostatistics*. 4, 1. https://digitalcommons.unmc.edu/coph_biostats_articles/4

Davenport, T. & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthc J.* 6(2), 94-98. 10.7861/futurehosp.6-2-94.

Davis, J. & Goadrich, M. (2006). The relationship between precision-recall and ROC curves. *ICML '06: Proceedings of the 23rd international conference on Machine learning*. 233-240.

GeneCaRNA. (2023, May 22) *RNY1 Gene - RNA, Ro60-Associated Y1*. GeneCards. https://www.genecards.org/cgi-bin/carddisp.pl?gene=RNY1

Hawinkel, S., et al. (2020). Sequence count data are poorly fit by the negative binomial distribution. *PLoS One.* 15(4), e0224909. https://doi.org/10.1371/journal.pone.0224909

Hoo, Z. H., Candlish, J., & Teare, D. (2017). What is an ROC curve?. *Emergency medicine journal : EMJ.* 34(6), 357-359. 10.1136/emermed-2017-206735.

Keilwagen, J., Grosse, I., & Grau, J. (2014). Area under precision-recall curves for weighted and unweighted data. *PloS One.* 9(3), e92209. 10.1371/journal.pone.0092209.

Lalkhen, A. G. & McCluskey, A. (2008). Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care & Pain.* 8(6), 221-223. https://doi.org/10.1093/bjaceaccp/mkn041

Vol. 2023 (2) 318 – 327
ISSN 2688-3651 [online]

Lin, R.-H., et al. (2022) Application of Deep Learning to Construct Breast Cancer Diagnosis Model. *Applied Sciences*. 12(4):1957, https://doi.org/10.3390/app12041957

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*. 5(9), 1315-1316. 10.1097/JTO.0b013e3181ec173d.

Ren, Q., Cheng, H., & Han, H. (2017). Research on machine learning framework based on random forest algorithm. *AIP Conference Proceedings*. 1820, 080020. https://doi.org/10.1063/1.4977376

Scikit-learn Developers. (n.d.) *sklearn.metrics.ConfusionMatrixDisplay*. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.ConfusionMatrixDisplay.html

Sun, Y.-S., et al. (2017). Risk factors and preventions of breast cancer. *Int J Biol Sci.* 13(11), 1387-1397. 10.7150/ijbs.21635.

Waks, A. G. & Winer, E. P. (2019). Breast cancer treatment: a review. *JAMA.* 321(3), 288–300. 10.1001/jama.2018.19323.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 57–63. https://doi.org/10.1038/nrg2484

Zhou, Z., et al. (2019). Extracellular RNA in a single droplet of human serum reflects physiologic and disease states. *National Academy of Sciences.* 116(38), 19200-19208. 10.1073/pnas.1908252116.

Journal of Research High School

Vol. 2023 (2) 328 – 336
ISSN 2688-3651 [online]

Impact of Technology on Education

Yeondoo Lee[1] *

[1]Henry M. Gunn High School, Palo Alto, CA, USA
*Corresponding Author: yd13014367@gmail.com

Advisor: Patricia Holmes, pholmes@pausd.org

Received June 13, 2023; Revised September 26, 2023; Accepted, October 27, 2023

Abstract

The incorporation of technology has resulted in a significant shift away from traditional methods of instruction and study, which has had a profound impact on educational systems all over the world. The purpose of this research is to evaluate and investigate the myriad ways that technology has changed education on several levels, including the curriculum design, student participation, and educational results. The essay presents a comprehensive examination of the existing body of information as well as empirical research, putting an emphasis on the positive aspects of incorporating technology into educational settings. It discusses how greater accessibility to knowledge and resources, made accessible by digital platforms, results in personalised and independent learning experiences for students. The research also examines how students may be helped by technology to collaborate on projects, think critically, and find solutions to challenges, so creating a learning environment that is more dynamic and entertaining. It also covers the challenges that arise and the elements that must be taken into consideration when combining technology with education, such as the requirements for infrastructure, the digital divide, and concerns regarding online privacy and safety. It emphasizes the necessity of pedagogical practices that make good use of technology and offers advice for teachers and policymakers on how to optimize the benefits of utilizing technology in educational settings. Curriculum, educational results, accessibility of knowledge and resources, independent learning, and cooperation are some of the keywords that can be associated with this topic.

Keywords: Curriculum, Blended Learning, Education System, Technology, Interactive Learning

1. Introduction

The integration of technology into educational settings has resulted in major changes to a variety of aspects of the traditional classroom setting. This article makes an effort to study and investigate the various implications that contemporary technology has had and is having on education, particularly instructional techniques in the classroom, the development of curriculum, student involvement, and academic results (Jangjarat et al., 2023). Technology has become indispensable to our day-to-day lives, and the inclusion of technology into educational settings has fundamentally changed the methods traditionally used for teaching and learning. Technology has increased the availability of educational materials and made them easier to access. This is accomplished by making it possible to utilize digital tools, software programs, and internet resources. This helps students overcome geographical limitations and improves their chances of learning.

The dynamics of the classroom have been altered, and accessibility has been improved, as a result of technological advancements. Traditional lectures have either been complemented by virtual learning environments, online debates, and multimedia presentations, or they have been completely replaced by these methods. The use of digital platforms improves the quality of collaborative learning experiences and encourages communication and collaboration among students. In addition, the unique requirements of each student are satisfied through the implementation of systems of adaptive learning and individualized learning paths, which encourage students to learn at their own speed and provide individualized educational experiences.

Vol. 2023 (2) 328 – 336
ISSN 2688-3651 [online]

The impact of technology on education extends far beyond the confines of the traditional classroom setting. Massive open online courses (MOOCs), online courses, and virtual classrooms have revolutionized the concept of distance learning by making it possible for everyone, regardless of location, schedule, or budget, to have access to educational opportunities. These innovations have given rise to a transnational and global educational community that provides opportunities for education to continue throughout one's life. However, the implementation of technology in educational settings frequently faces difficulties. Infrastructure requirements, such as consistent internet connectivity and sufficient hardware, need to be addressed in order to make sure that all of the students have access to technology on an equal basis. It is imperative that concerns such as the digital divide, security, and online safety be adequately addressed in order to guarantee a risk-free setting for educational activities.

This paper makes an effort to give a comprehensive examination of how technology is changing education by taking into account both the advantages that it provides and the potential disadvantages that it may have. By analyzing recently published research and conducting original empirical investigations, the purpose of this article is to contribute to the growing body of knowledge regarding the ways in which technological advancements have altered education, given students more agency, improved educational outcomes, and reduced educational disparities.

This article will provide crucial insights for educators as well as policymakers by analyzing the possibilities and problems related with technology integration. These evaluations will assist educators and policymakers in making decisions that will help educational institutions make the most of the benefits that technology has to offer. The findings of this research will not only contribute to the existing body of knowledge, but they will also educate educational stakeholders on the most effective ways to utilize technology in an educational setting that is undergoing rapid transformation.

2. The Development of Technology Throughout the Educational System

2.1 The evolution of educational technology in its historical setting

There is a connection that can be drawn between the early tools and technologies that were designed to assist teaching and learning and the development of educational technology. In the early 20th century, the arrival of projectors, film strips, and instructional radio shows marked the beginning of early attempts to incorporate technology into educational settings (Hattie, 2023). These early developments had the intention of enhancing the delivery of teaching and involving students through visual and auditory channels. However, the emergence of the computer in the 1970s marked the beginning of a major revolution in educational technology. The rise in popularity of computer-based lessons and educational software can be attributed to the fact that these mediums provide individualized instruction in addition to opportunities for interactive learning. During this time period, there was a significant change in the way that technology was integrated into educational environments, which opened the door for further advancement.

2.2 Recent advances in technology and their influence on education are as follows

The introduction of new technologies has led to substantial shifts in the approaches that are taken to education, both in terms of teaching and learning. The proliferation of internet use throughout the 1990s brought about changes in the ways in which people could access information and resources (Garzón & Acevedo, 2019). Knowledge of a broad scope was made accessible to both students and teachers, which encouraged learning, collaboration, and the sharing of information.

The proliferation of individual electronic devices and personal computers in the 2000s accelerated the process of incorporating technology into the classroom. Students were able to learn whenever they wanted, whenever they wanted, and at their own pace thanks to the tools that were provided. In addition, internet platforms and learning management systems (LMS) came into existence, which expanded the opportunities for online education, virtual classrooms, and other forms of interactive educational environments.

Vol. 2023 (2) 328 – 336
ISSN 2688-3651 [online]

2.3 Integration of contemporary technologies into present educational settings

Integration of technological tools is increasingly often practiced in educational institutions. Through the implementation of interactive whiteboards, video technologies, and computer laboratories into the educational setting, both the delivery of education and the engagement of students can be significantly enhanced (Jangjarat et al., 2023). As a result of the explosive growth of online education in recent years, educational institutions today provide students at all levels with a variety of learning options, including blended and fully online classes.

Businesses that specialize in educational technology are continually developing cutting-edge solutions in order to enhance students' learning environments. The instructional content has been improved through the use of flexible learning systems, educational software, and multimedia resources in order to support various learning styles. There has also been a rise in the development of technologies such as virtual and augmented reality, which provide chances for immersive learning and foster higher comprehension as well as involvement (Radović et al., 2021). To give equal access to technology, to close the digital gap, and to motivate instructors and students to become digitally literate is still a difficult task. Efforts are being undertaken in order to eliminate these gaps and provide equivalent opportunities for all of the students.

Important new insights on how technology has transformed and influenced educational practices can be gained from an understanding of the historical context, the technical developments, and the status of the integration of technology in education at this moment (Jangjarat et al., 2023). If educators and policymakers are aware of these trends, they will be able to make better use of technology to its full potential, which will ultimately lead to improved teaching and learning outcomes.

3 Positive impacts

The numerous positive benefits that technology has had on education have had the effect of transforming the ways in which teaching and learning are done. The following is a list of the primary benefits that technology brings to education

3.1 Encouraged cooperation and coordination

Because of advancements in technology, which make it easier to solve problems and think critically. The use of online tools, video conferencing, and collaborative software gives students the opportunity to work together on projects, discuss ideas, and provide feedback. Because of the collaborative learning environment, which encourages students to work together, communicate effectively, and find solutions to problems, students are more prepared for the challenges they will face in the real world.

Technology enables the creation of learning environments that are more dynamic and exciting by way of mediums such as multimedia content, interactive simulations, and gamification. Videos, animations, and other forms of multimedia presentations keep students interested while also breaking down complex concepts (Radović et al., 2021). Hands-on learning possibilities can be found through the use of interactive simulations and virtual experiments, which encourages active research and leads to higher understanding. Features of gamification such as rewards and leaderboards are two examples of elements that make learning more interesting and encourage active engagement from students.

3.2 Learning at a distance and Use of technology

Access to education has been substantially improved because to technological advancements, particularly for individuals who face financial or geographical barriers. Through the utilization of online learning technologies and virtual classrooms, students are able to pursue their education from a distance, overcoming both the constraints of time and location. Those individuals who previously may not have had access to educational opportunities now do so as a direct result of this accessibility. Technology makes it easier for educators and institutions to carry out their

Vol. 2023 (2) 328 – 336
ISSN 2688-3651 [online]

administrative obligations (Radović et al., 2021). Automating processes like record-keeping, grading, and attendance monitoring with digital technologies helps save time and minimizes the amount of paperwork required. Because of its efficiency, the classroom teacher is able to concentrate more on teaching and providing tailored assistance to each student.

Because of technological advancements, information and educational resources are now more readily available than at any time in history. Students who are able to use the internet to access it can gain access to a multitude of material, including academic websites, research journals, and digital libraries. Because information is so readily available, students have the opportunity to investigate a wide range of subjects, improve their level of subject comprehension, and conduct research.

3.3 Learning Opportunities at Your Own Pace

Chances for personalised learning are made available by innovation, and these chances can be tailored to meet the needs of each learner. Systems for adaptive learning and intelligent tutoring services use data analytics and algorithms to determine a student's strengths and limitations in order to provide individualized instruction. The students receive individualized content, feedback, and learning paths based on this research, which enables them to study at their own pace and concentrate on the topics in which they have the greatest room for improvement.

4 Problems Associated with Technology Integration

4.1 Issues of unequal access and the requirement for improved infrastructure

One of the most significant challenges that must be overcome in order to successfully integrate technology into educational settings is the construction of a stable infrastructure. For educational institutions to be able to support learning that is enhanced by technology, they need to have stable internet access, sufficient bandwidth, and up-to-date hardware and software. Despite this, there are still access barriers, which are especially prevalent in developing regions or remote locations with inadequate infrastructure (Tamim et al., 2011). It is essential to ensure that everyone has equal access to the technical resources that are available in order to put a stop to the widening educational gaps. It is important that there be an increase in broadband connectivity, that there be funding for the construction of infrastructure, and that students in underserved locations have access to the resources that they require.

4.2 Disparity in Peoples Levels

The disparity in people's levels of access to various forms of technology is referred to as the "digital divide." Variables of socioeconomic status, geographic location, and disparities in the availability of technical resources all contribute to the widening of this chasm (Raja & Nagasubramani, 2018). Accessing online resources and actively participating in classes that make use of technology may be challenging for students who do not have access to electronic devices or a reliable internet connection in their homes. To bridge the digital gap, complete solutions are required, some examples of which include the creation of community technology centers, the implementation of mobile learning programs, and the distribution of electronic devices and internet access to students from disadvantaged backgrounds. To ensure that all students are provided with an equal opportunity to profit from the implementation of technology in the classroom, it is vital to place an emphasis on equality.

The use of technology in the classroom raises concerns over students' right to privacy and their protection when online. The protection of the privacy of students' data is absolutely necessary given that the collection and storage of personal information could be risky if it was not well safeguarded. Educational institutions are obligated to lay out clear guidelines, policies, and procedures in order to protect the privacy of their students and ensure compliance with applicable data protection legislation (Tamim et al., 2011). This comprises safeguarding platforms, obtaining the required consent prior to data collection, and putting protections in place to prevent data breaches. Both teachers and students need to receive training in online safety, responsible online citizenship, and cybersecurity so that they can

Vol. 2023 (2) 328 – 336
ISSN 2688-3651 [online]

navigate the online learning environment in a manner that does not compromise their personal safety. Ongoing education and awareness campaigns should be provided in order to effectively address issues relating to the protection of personal information and public safety.

4.3 Pedagogical concerns, as well as continued professional development for teachers

In order to successfully integrate technology into the classroom, careful pedagogical preparation and ongoing professional development for educators are required prerequisites. Simply adopting new technology might not be enough to provide the desired educational outcomes if there are not also new pedagogical techniques. It is essential for educators to be able to integrate appropriate uses of technology into their pedagogy so as to foster enhanced student learning. To do this, it is vital to align the use of technology with the learning goals, to create activities that are entertaining and relevant, and to offer opportunities for student participation and critical thinking. It is important that educators have access to the knowledge and skills necessary to make successful use of the various technology tools, platforms, and apps that are accessible (Tamim et al., 2011). It is essential to provide teachers with continual opportunities for professional development, education, and support in order for them to be able to use technology to its fullest potential. Collaboration between educators, technology specialists, and instructional designers can be an effective means of facilitating the sharing of effective best practices and the development of effective instructional strategies.

Taking into account these difficulties and concerns is one way for educational organizations and governments to ensure the successful incorporation of technology into educational settings (Tamim et al., 2011). Establishing infrastructure, encouraging equitable access, resolving issues regarding security and privacy, and providing pedagogical help are some of the ways in which educational systems can capitalize on the benefits of technology while also minimizing any potential drawbacks. It is vital to approach the incorporation of technology in a holistic manner, taking into account the many different requirements of students and aiding them in the process of acquiring knowledge in a way that is both welcome and effective.

5 Case Studies and Other Forms of Empirical Evidence

Numerous studies have been carried out to investigate the effects that technology has had on education, and the efficacy of this technology as well as its benefits have been brought to light through empirical data and case studies (Sinacori, 2020). The following is a comprehensive review of case studies and actual data that demonstrates how technological advancements have increased educational opportunities:

The use of technology in the classroom has been linked in a number of studies to improved academic performance, which supports the hypothesis that this connection exists. For instance, a meta-analysis conducted by (Harris et al., 2015) found that technology-enhanced teaching significantly improved student success across a variety of subject areas and grade levels. In addition, a case study conducted by shown that the implementation of technological tools, like as interactive whiteboards, improved learning outcomes and increased the level of student involvement. The participants in this quantitative study were Fourth Grade students from two different classrooms, but in the same Title 1 School, located in Central Illinois. According to the Illinois Interactive Report Card (2013), the school has a low-income rate of 84.3%, with 40.5% of the students being African-American, 15.2% Multiracial, 32.3% Caucasian, 10.2% Hispanic, 1.0% American-Indian, and 0.7% Asian.

Students can be inspired and kept interested through the use of technology. A study conducted by Wang and Eccles (2011) found that technology-enhanced educational activities fostered better levels of student participation when compared to more conventional forms of instruction. Case studies have also highlighted the use of gamification and multi-media interaction as useful approaches for attracting students' attention and encouraging active engagement in the learning process. Gamification refers to the practice of turning a learning experience into a game.

According to (Mallon et al. 2021). The researchers utilized a mixed method approach to understanding how the integration of technology affected students' learning. A survey was developed and administered through Qualtrics to collect data. The survey contained 14 questions that utilized a variety of questions formats such as open-ended,

Vol. 2023 (2) 328 – 336
ISSN 2688-3651 [online]

multiple choice, and Likert scale. Plano, et al. (2010) describe the survey research design process as being fairly flexible. The researchers utilized this approach and synthesized trends revealed by the data. The survey was sent to K-12 educators at a local school district in central Illinois. Participation was voluntary. The data was analyzed using descriptive statistics such as means, standard deviation, and percentages. Qualitative data was analyzed and organized into emerging themes. The quantitative and qualitative data was triangulated to help answer the research questions.

Learning that Is individualized and customized to the learner's needs and preferences technology has made it feasible for students to have learning experiences that are individualized and customized to meet their particular needs and preferences in terms of how they learn best. According to the findings of a study computer-based education enables individualized instruction, which ultimately results in higher levels of student achievement than traditional instruction. Case studies have demonstrated how intelligent tutoring systems and adaptive learning platforms can modify content and provide students with personalised feedback to assist students in improving their academic performance.

Technology fosters an atmosphere that encourages cooperative learning among students by making it simpler for them to work together and communicate with one another. According to the findings of research conducted by the integration of technology facilitated collaborative learning, which in turn increased students' capacity for critical thinking and problem-solving (Raja & Nagasubramani, 2018). Case studies have shown that encouraging student participation, peer review, and group problem-solving through the use of online discussion boards, collaborative activities, and virtual classrooms may be accomplished with great effectiveness.

Students are able to engage in more realistic learning experiences thanks to the advancements in technology that make it possible for learning to be applied in real-world settings. The research conducted by Herrington highlighted the benefits of realistic learning environments that are mediated by technology and in which students engage in activities that are relevant to the real world and address problems that are relevant to the real world. Case studies have made use of simulations, virtual reality, and technologies for augmented reality in order to provide immersive and authentic learning experiences for students. These types of experiences have enhanced students' conceptual grasp and application.

Access to Education Has Increased The number of children who are able to receive an education has increased, particularly those who live in remote or disadvantaged areas. Case studies have placed an emphasis on the utilization of mobile devices, video conferencing, and online learning platforms in order to provide high-quality education to students who may not have access to traditional educational resources. Students are now able to complete their educations at a distance thanks to the advancements in technology, which has eliminated the need for geographical constraints.

The positive effects that contemporary technology has had on the educational system are consistently backed up by factual data and case studies. The findings demonstrate increased access to education, improved academic accomplishment, increased student engagement, personalized learning environments, collaborative learning settings, and the application of information in the real world (Radović et al., 2021). By making judicious use of technology, educators are able to create transformative learning experiences for their pupils that cater to the various educational requirements of their pupils and set them up for academic success in the digital age.

6 Advice and suggestions

6.1 Explicit Regulations and Policies for Technology Integration and sufficient funds

Make explicit regulations and policies that encourage efficient and fair technology integration in education. This is one of the many proposals that lawmakers and educators should consider implementing in order to make the most of technology in the educational setting. A few examples of the challenges that these guidelines ought to address are infrastructure requirements, access disparities, curricular integration, and user data privacy.

And promote the availability of enough funding for technology resources. It is essential to make certain that educational institutions have access to the necessary computer hardware, software, and other digital tools (Raja, & Nagasubramani, 2018). Spending on technological infrastructure should be the top expenditure priority for governments if they want to close the digital divide, particularly in underdeveloped areas.

Vol. 2023 (2) 328 – 336
ISSN 2688-3651 [online]

6.2 Collaboration of Stakeholders and Resolution Strategies

Policymakers, educators, and technologists ought to collaborate in order to draft laws and regulations that are able to fulfill the requirements posed by children and their instructors (Kuppusamy, 2019). By cooperating with one another, we will be able to ensure that the rules allow for the effective integration of technology and reflect best practices.

Improve Internet Connectivity

Policymakers should make it their goal to expand broadband access in educational institutions and communities that are currently underserved (Radović et al., 2021). This goal can be accomplished with the assistance of governmental financing, infrastructure development programs, and collaborative efforts with internet service providers.

Ensure Access to Technological Resources

See to it that students have access to a variety of technological resources and devices. Every student should be able to use a computer, iPad, or Chromebook, and schools ought to make that availability a priority. Initiatives should be taken to provide electronic devices to children who do not have access to such items in their homes in order to promote a more equitable environment for learning.

Utilize Mobile Technology

It's possible that using a mobile device to get around access problems is a good idea. Policymakers and educational institutions should look into programs that make use of mobile devices like tablets and smartphones to provide instructional content and encourage online learning.

6.3 Ensuring Data Privacy and Cybersecurity

To preserve students' privacy, maintain their safety, and ensure that their use of technology does not compromise their morals, the following measures should be taken:

Clear Guidelines for Data Privacy

Schools and other types of educational institutions need to have procedures and policies in place to keep student information secure. These policies should include guidelines for the collection, storage, communication, and retention of data in accordance with any applicable data protection regulations.

Cybersecurity Education

Students and teachers both need to receive training on how to behave in a moral and responsible manner when using the internet, as well as how to be safe while using it (Radović et al., 2021). Utilizing the internet in a secure manner, recognizing and evading online dangers, and guarding personal information ought to be the primary focuses of this education.

Educational institutions such as schools and colleges ought to routinely examine the IT infrastructure and systems they use in order to identify any areas of vulnerability and install the necessary security controls. Installing firewalls, antivirus software, and encryption techniques are some of the measures that need to be taken in order to keep data and networks secure. Professional development is necessary for educators if they are to make effective use of technology. It is imperative that educational institutions provide teachers with ongoing opportunities to further their professional development in the field of technology (Tamim et al., 2011). Training sessions and seminars may place an emphasis on particular software programs, platforms, and instructional strategies that facilitate effective technology integration.

Make it simpler for teachers to collaborate with one another and share successful strategies for integrating technology into the classroom setting. This will encourage more effective use of technology in classrooms (Wang & Eccles, 2011). This can be accomplished with the support of educational networks, online forums, or specific platforms

Vol. 2023 (2) 328 – 336
ISSN 2688-3651 [online]

where instructors can share materials, ideas, and stories of their students' achievements. Incorporate technology into teacher education programs: Both the coursework and the opportunities for hands-on learning that are included in programs for teachers should be geared to better educate future teachers on how to utilize technology in the classroom. This may be done through mentoring programs, specialist technology classes, and field trips in classrooms with plenty of technology.

These tips can help educators and decision-makers make the best use of technology within the classroom (Radović et al., 2021). A successful and revolutionary incorporation of contemporary technology in education may be assisted by appropriate legislation, increased infrastructure, privacy controls, and chances for professional development for instructors.

7 Conclusion

Technology has benefitted education in a number of ways. It has enhanced access to knowledge and assets, created personalised learning opportunities, stimulated teamwork and critical thinking abilities, and produced exciting and interesting learning situations. The current level of technology integration in education is the product of historical events and breakthroughs driving the development of technology in education.

There are concerns and aspects to be taken into account. The need for educational issues and teacher professional development are among them, as are infrastructure needs and access disparities, the digital gap and equity challenges, security and internet safety concerns, and the demand for a safe online environment.

New technological breakthroughs have the possibility to further improve instructional techniques because technology is always improving. Ongoing investigation demonstrates unique techniques to using technology and maximizing its effects on education.

To address new issues that could emerge as technology proceeds to transform the educational landscape. In order to understand and solve these difficulties, such as assuring equal access, resolving privacy concerns, and appreciating the implications of developing technology on teaching and learning, research is needed.

To encourage effective technology integration strategies in education, research delivers evidence to support evidence-based practices. It provides educators with guidance in making decisions on the selection, application, and evaluation of various technology tools and resources.

Technology has the potential to totally transform education and create learning environments that are more inclusive and effective for all students. It has the potential to eliminate barriers to accessing high-quality education, adapt to the specific educational needs of each student, foster collaborative problem-solving and imaginative thought, and prepare young people for life in the digital age. However, in order to make good on this promise, it will be necessary to devise an all-encompassing plan that takes into account issues such as infrastructure, equity, privacy, education, and professional growth.

If they want to fully achieve the revolutionary potential of technology, educators, lawmakers, and stakeholders need to collaborate and prioritize projects to address problems, provide vital support, and foster research and innovation. Only then will they be able to fully realize the potential of technology. If we do this, we can unlock the full potential of technology to revolutionize education and give kids the tools they need to be successful in a world that is always evolving.

References

Garzón, J., & Acevedo, J. (2019). Meta-analysis of the impact of Augmented Reality on students' learning gains. *Educational Research Review*, 27, 244-260.

Harris, J., & Al-Bataineh, A. (2015, April). One to one technology and its effect on student academic achievement and motivation. In Global learn (pp. 579-584). *Association for the Advancement of Computing in Education (AACE).*

Hattie, J. (2023). Visible learning: The sequel: A synthesis of over 2,100 meta-analyses relating to achievement. *Taylor & Francis*.

Vol. 2023 (2) 328 – 336
ISSN 2688-3651 [online]

Jangjarat, K., et al. (2023*). The impact of Smart Education on learning outcomes in the digital era* ... ResearchGate.

Kuppusamy, P. (2019). Smart education using internet of things technology. *Advances in Data Mining and Database Management,* 385–412.

Radović, S., Hummel, H. G., & Vermeulen, M. (2021). The mARC instructional design model for more experiential learning in higher education: theoretical foundations and practical guidelines. *Teaching in Higher Education*, 1-18.

Raja, R., & Nagasubramani, P. C. (2018). Impact of modern technology in Education. *Journal of Applied and Advanced Research*.

Sinacori, B. C. (2020). How nurse educators perceive the transition from the traditional classroom to the online environment: *A qualitative inquiry. Nursing education perspectives*, 41(1), 16-19.

Tamim, R. M., et al. (2011). What forty years of research says about the impact of technology on learning. *Review of Educational Research, 81(1)*, 4–28.

Wang, M.-T., & Eccles, J. S. (2011). Adolescent behavioral, emotional, and cognitive engagement trajectories in school and their differential relations to educational success. *Journal of Research on Adolescence,* 22(1), 31–39.

Journal of Research High School

Vol. 2023 (2) 337 – 346
ISSN 2688-3651 [online]

JRHS Outstanding Research Paper Award

Assessing the Impact of Climate Change on the Long-Range Transport of Smoke from the Canada Wildfire Event of June 2023

Rachel Kim[1] *

[1]Bergen County Academies, Hackensack, NJ, USA
*Corresponding Author: rachelkim1120@gmail.com

Advisor: Myoungwoo Kim, information@bluebellenvirobotics.com

Received September 1, 2023; Revised October 11, 2023; Accepted, November 3, 2023

Abstract

In the wake of the 2023 Canada Wildfires, this research analyzes the impact global warming had on intensifying the environmental effects of wildfires, focusing in on the Canada Wildfire Smoke Event. Using climate modeling techniques, this study explored the long-range transport and aerosolization of particulate matter, especially fine particulate matter ($PM_{2.5}$). Emphasis was placed on regions with the highest $PM_{2.5}$ values such as Queens College, with the investigation further extending to other cities across New York City, spanning a 32-mile radius. It was discovered that $PM_{2.5}$ values in New York surged to 132.23 $\mu g/m^3$, marking a significant increase compared to New York City's normal $PM_{2.5}$ value of 7.88 $\mu g/m^3$. Across all seven sites in NYC, the net contribution of the wildfire varied between 109 $\mu g/m^3$ and 196 $\mu g/m^3$. Back-trajectory analysis over a 48-hour period and dispersion analysis with and without the deposition scheme revealed that the 2023 Canada Wildfire Smoke Event significantly compromised air quality in areas like New York, New Jersey, and Pennsylvania. Our backward dispersion modeling indicated that the influence of global warming during the Canada wildfire event has elevated NYC's air quality to levels that are up to 89 times higher than what might have been anticipated without such global warming influences. This observation aligns with the conclusion that rising temperatures, due to global warming, dry out the atmosphere by depleting vital moisture and relative humidity at semi-arid regions. Such conditions hinder crucial processes like coagulation and deposition, and because there is reduced atmospheric moisture, there are fewer moisture particles for smoke to attach to or interact with, meaning that it can travel longer distances without significant deposition. This research underscores the exacerbated effects of wildfires on air quality under the influence of global warming and the results accentuates the urgent need for proactive environmental strategies that address these temperature-induced atmospheric changes, along with the integration of these insights into climate models for effective global climate and air quality management.

Keywords: Canada wildfire smoke event, Global warming, Long-range transport, $PM_{2.5}$, Aerosolization

1. Introduction

Exposure to $PM_{2.5}$, or particulate matter less than 2.5 micrometers in diameter, has been linked to a variety of adverse health effects, including respiratory issues, cardiovascular diseases, and increased mortality rates (Brook et al., 2010; Pope et al., 2002). These tiny particles can penetrate deep into the lungs and even enter the bloodstream, exacerbating preexisting conditions like asthma and heart disease, and potentially leading to more serious complications such as heart attacks or chronic obstructive pulmonary disease (COPD) (Dominici et al., 2006; Dockery et al., 1993).

During the period from June 5th to 7th, a concerning escalation in the air quality index in New York City (NYC) was observed, triggered by the formidable Canada Wildfire events located in Ottawa Province (Cohen, 2023). The

Vol. 2023 (2) 337 – 346
ISSN 2688-3651 [online]

scope and severity of this issue came to light when the New York Times reported that wildfires in Canada had already engulfed an astonishing 25 million acres since the start of the year, surpassing the country's previous annual record set in 1989, when over 18 million acres were devastated by fire (Popovich, 2023). The enormity of the impact was further substantiated as the Copernicus Atmosphere Monitoring Service (CAMS) estimated that these wildfires released nearly 160 million tons of carbon into the atmosphere (CAMS, 2023). An equally alarming aspect is the long-distance dispersion of nearly 600 million tons of carbon dioxide and particulate matter generated by these wildfires, affecting areas as far-reaching as New York, New Jersey, and Pennsylvania (Salahieh et al., 2023).

The underlying cause of fires in the boreal forest of northern Canada can be attributed primarily to lightning (Romps et al., 2014). In conjunction with a one-degree Celsius increase in temperature, there is approximately 12% more lightning, making the warming climate a significant factor in the frequency of fire ignition (Price, 2009). The science behind this phenomenon is grounded in the understanding that global warming increases temperature through the accumulation of greenhouse gases (IPCC, 2021), which trap heat from the sun in the Earth's atmosphere. In arid and semi-arid regions, global warming has the potential to influence the long-range transport of wildfire smoke. Warmer and drier conditions can lead to reduced coagulation and deposition of smoke particles, allowing them to remain airborne for extended periods (Liu et al., 2010). While global warming enhances evaporation rates, the low relative humidity characteristic of these regions can limit the condensation of this evaporated water into droplets or clouds, reducing the atmosphere's natural "cleansing" mechanisms (Reid et al., 2005). Moreover, the increased temperatures associated with global warming can also amplify the frequency and intensity of wildfires, producing more smoke (Abatzoglou et al., 2016). Thus, in these dry regions, global warming can promote conditions conducive to the long-distance travel of wildfire smoke.

In light of these urgent and complex issues, this study aimed to achieve two objectives. The primary objective of this study was to validate the impact of the Canada wildfire smoke long-range transport on NYC's $PM_{2.5}$ levels using back trajectory analysis (Liu et al., 2020) and to quantify its contribution to $PM_{2.5}$ concentrations over and above the local background levels. The secondary aim was to demonstrate and quantify the influence of global warming on this long-range transport of wildfire smoke (Ford et al., 2018) by employing an air dispersion modeling. Through this dual-focused approach, our research delves into the profound interconnections between climatic changes, wildfires, and their extended atmospheric consequences.

2. Materials and Methods

2.1 HYSPLIT4 Modeling

The HYSPLIT 4 (Hybrid Single-Particle Lagrangian Integrated Trajectory) model allows for the tracking and analyzation of the effects of the Canada wildfire. HYSPLIT 4 is a powerful atmospheric dispersion model that enables the simulation and visualization of the movement of air masses and pollutants over time (Draxler & Hess, 1997). Inputting relevant meteorological data and fire emissions information allowed the model to generate trajectory paths of smoke particles originating from the Canada Wildfire event. These trajectory paths provided insight to the long-range transport of particulate matter and smoke, allowing for assessment of potential impacts on air quality and atmospheric composition in distant regions.

2.2 Lagrangian Dispersion Scheme

The Lagrangian Dispersion Scheme, rooted in the domain of atmospheric science, is a specialized method designed to model the transport and dispersion of pollutants within the atmosphere (Draxler & Hess, 1997). Its foundation lies in the Lagrangian perspective of fluid dynamics. Unlike the Eulerian viewpoint, which focuses on observing changes in fluid properties at fixed spatial points (akin to watching a river from a stationary bridge), the Lagrangian perspective emphasizes tracking individual particles or air parcels as they traverse through space and time.

The primary technique of the Lagrangian Dispersion Scheme involves meticulously simulating the trajectories of a multitude of "particles." These particles, in the context of the model, can symbolize air parcels, droplets, or actual

Vol. 2023 (2) 337 – 346
ISSN 2688-3651 [online]

particulate pollutants. As these particles journey through the atmosphere, the model integrates various forces and dynamics that influence their movement. Such influences range from wind advection and turbulent diffusion to more intricate processes like gravitational settling for larger particulates or potential chemical transformations.

One noteworthy feature of this scheme is its incorporation of stochastic (randomly determined) processes, especially when representing turbulence. This inclusion mirrors the inherently random nature of turbulent motions in the atmosphere. Consequently, by analyzing the collective behavior of these simulated particles over a period, the Lagrangian Dispersion Scheme offers a nuanced understanding of pollutant dispersion patterns. Such insights prove invaluable, especially in scenarios demanding knowledge of pollutant pathways from sources like industrial power plants or expansive wildfires.

2.3 Back Trajectory

A back trajectory, an essential instrument in atmospheric science, delineates the antecedent path of an air parcel or particle (Stein et al., 2015). Through the meticulous amalgamation of meteorological data and sophisticated computational models, this tool traces in reverse the journey an air parcel has undergone, spanning from its concluding location to its origin (Rolph, Stein, & Stunder, 2017). The core objective of such an analysis is the identification of probable source regions or pathways for air masses and the pollutants they might carry.

By discerning an air parcel's origins and the atmospheric conditions it encountered, researchers can pinpoint potential pollution sources and interpret the atmospheric composition observed at specific sites. Historically archived meteorological data, encompassing wind patterns, temperatures, and other salient atmospheric metrics across multiple altitudes, underpins the precision of back trajectory computations. Depending on the research aims, the temporal scale of these trajectories can vary widely; intercontinental air mass transport might necessitate a more extended trajectory analysis, while local or regional pollution sources could be discerned with shorter trajectories. The practical implications of back trajectory analyses are multifaceted. Whether utilized for source apportionment endeavors to demystify pollutant origins or to fathom phenomena like transcontinental smoke transport from forest fires, dust storms, or the encroachment of polluted air into pristine zones, its value remains indisputable. In essence, the back trajectory offers a retrospective lens, granting profound insights into the historical movement of air masses and enhancing our comprehension of atmospheric transport dynamics.

2.4 Dispersion Simulation With and Without Deposition

Deposition pertains to the transfer of atmospheric particles or gases from the atmosphere to the Earth's surface. This process is primarily categorized as wet and dry deposition. While wet deposition involves particles or gases being purged from the atmosphere from precipitation, dry deposition refers to their settling onto surfaces independently of precipitation (Seinfeld & Pandis, 2016).

Models that operate "with deposition" inherently incorporate both the transport and deposition of particles or gases. Such models provide invaluable insights into potential accumulation points of pollutants on the ground, crucial for environmental and health risk assessments. For instance, they can highlight regions where acid rain might jeopardize soil or water resources or locations susceptible to the harmful accumulation of particulates. In these models, pollutants often exhibit truncated atmospheric lifetimes due to their removal by deposition processes.

Conversely, models that function "without deposition" emphasize purely the transport and dispersion of particles or gasses. They eschew considerations of pollutants being removed from the atmosphere either by settling or precipitation-led processes. This perspective assumes an extended presence of pollutants in the atmosphere and is typically simpler in nature. Such models can be particularly insightful when the core objective is discerning pollutant movement without the intricacies associated with deposition.

The choice to employ models "with" or "without" deposition hinges on the objectives of the study. Comprehensive insights into the fate of pollutants, including terrestrial or aquatic destinations, warrant the inclusion of deposition. However, if the study's focus leans predominantly towards understanding pollutant dispersion and movement, omitting deposition might be more appropriate. This paper employs both options in order to analyze maximal results.

Vol. 2023 (2) 337 – 346
ISSN 2688-3651 [online]

2.5 Estimations of Ambient Air Concentration Provided by HYSPLIT4

Within the horizon of atmospheric and environmental sciences, estimations of ambient air concentration represent a critical metric. They provide an estimate of specific pollutants, gases, or particulate matter in the air of a designated locale. Typically expressed in units such as micrograms per cubic meter ($\mu g/m^3$) or parts per million (ppm), these estimations afford a snapshot of the air quality at a given moment. By combining trajectory modeling with emission data and meteorological inputs, HYSPLIT4 provided insight into estimating the ambient air concentrations of these pollutants over the course of the Canada wildfire event. This estimation offered valuable insights into the spatial and temporal distribution of air pollutants, aiding in the assessment of their potential impacts on air quality and public health in the affected regions.

2.6 Data Collection

The data's comprehensive scope not only ensures a robust understanding of $PM_{2.5}$ and Ozone fluctuations but also helps discern the potential regional variances in air quality. Factors such as traffic density, industrial activities, local weather conditions, and green spaces could contribute differently to the air quality in these locations. Morrisania, for instance, might experience different pollutant sources compared to Freshkills West, Staten Island, due to its unique geographical and infrastructural characteristics. By examining such a vast array of locations, this breadth of data paves the way for more targeted interventions and strategic air quality improvement initiatives tailored for specific regions within New York.

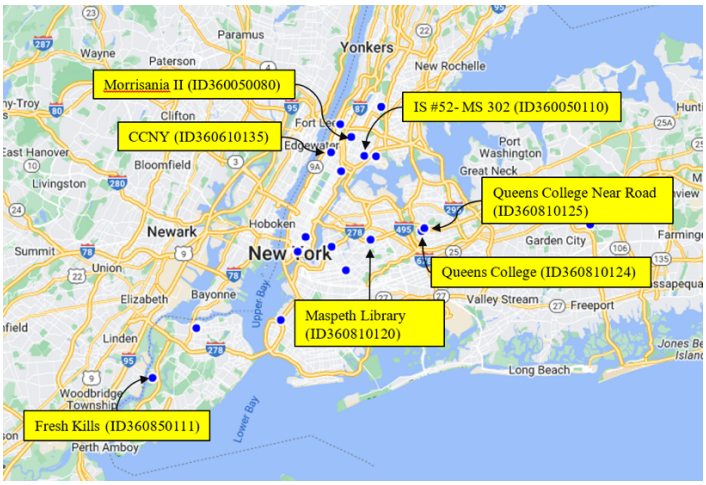


Figure 1. Air Monitoring Stations Map (Source: the New York State DATA.NY.GOV. All captioned sites are the ones used in this paper)

2.7 Trend Analysis

Trend analysis emerges as an indispensable instrument for discerning nuanced patterns within temporally sequenced datasets. The methodology is underpinned by rigorous examination and assessment of sequential data points collected over discrete intervals. Its primary objective is to illuminate the trajectory, whether ascending, descending, or remaining static, of a specified metric or phenomenon. By adeptly employing trend analysis, historical tendencies of the subject variable offer projections about its prospective evolution. Specifically, the trend analysis conducted for this study showcased the diurnal variations of pollutant concentrations spanning from June 4th to June 7th. This temporal depiction furnishes a robust visualization elucidating the oscillations in pollutant concentrations during the delineated period.

2.8 Overlaying Wildfire Map and Back Trajectory

Overlaying the wildfire map and back-trajectory patterns using the program Microsoft Excel allowed for the tracing of wildfire patterns and a further analyzation of the patterns and the routes they took to reach locations like New Jersey and NYC. Overall, this process provided significant information to analyze the steadfast growth and the concerning impact of these fires as it revealed the pathway the fires followed to reach locations in North America.

Vol. 2023 (2) 337 – 346
ISSN 2688-3651 [online]

2.9 Theoretical Background for HYSPLIT4 Simulations (Draxler & Hess, 1998)

Lagrangian Dispersion Scheme represented in general terms by stochastic differential equation (equation below represents a particle's position)

$$d\mathrm{X}(t) = \mathrm{V}(\mathrm{X}(t), t)dt + \mathrm{B}(t)d\mathrm{W}(t) \quad (1)$$

Where $\mathrm{X}(t), \mathrm{V}(\mathrm{X}(t), t), \mathrm{B}(t)$, and $d\mathrm{W}(t)$ represent the position of the particle at time *t,* the mean wind velocity at the position of the particle, the turbulence parameter, and a Wiener process which represents the random nature of turbulent motion, respectively.

Particle's concentration (if considering a puff model)

$$\frac{dC(t)}{dt} = -\frac{Q}{V(t)} + S(t) \quad (2)$$

Where $C(t), Q, V(t)$, and $S(t)$ represent the concentration of the pollutant within the puff at time *t,* the rate at which the puff is losing mass, the volume of the puff (which might increase due to dispersion), and any sources or sinks of the pollutant within the puff, respectively.

Back trajectory equation

$$\mathrm{X}(t - \Delta t) = \mathrm{X}(t) - \mathrm{V}(\mathrm{X}, t)\,\Delta t \quad (3)$$

Where $\mathrm{X}(t), \mathrm{V}(\mathrm{X}, t)$, and Δt represent the particle's position at time *t,* the wind velocity at the particle's position at time *t,* and the time step, respectively.

Gaussian plume model for dispersion without deposition

$$C(x, y, z) = \frac{Q}{2\pi U \sigma_y \sigma_z} e^{-\frac{y^2}{2\sigma^2 y}} e^{-\frac{(z-H)^2}{2\sigma^2 z}} \quad (4)$$

Where $C(x, y, z), Q, U, \sigma_y$ and σ_z, H, and $e^{-\frac{y^2}{2\sigma^2 y}}$ and $e^{-\frac{(z-H)^2}{2\sigma^2 z}}$ represent the concentration of the pollutant at a specific location given the coordinates (x, y, z), the emission rate of the pollutant, the wind speed, the dispersion coefficients or standard deviations of the concentration distribution in the *y* (crosswind) and *z* (vertical) directions, the effective release height of the pollutant, and the exponential functions that describe the Gaussian distribution of the pollutant concentration in the crosswind and vertical directions, respectively.

Dispersion with deposition

$$\frac{dC}{dt} = \text{dispersion term} - V_d C \quad (5)$$

Where $\frac{dC}{dt}$, dispersion term, V_d, and C represent the rate of change of concentration time, the spreading and dilution of the pollutant, the deposition velocity, and the concentration of the pollutant, respectively.

3. Results

3.1 $PM_{2.5}$ and Ozone Concentrations in NY

According to the EPA, a $PM_{2.5}$ concentration of 12 μg/m^3 or lower is generally considered healthy, posing minimal risks from exposure. However, when the level exceeds or equals 35 μg/m^3 within a 24-hour period, the air quality is deemed unhealthy, especially for individuals with pre-existing respiratory conditions such as asthma (EPA, 2012). During July 6th to July 8th, the NYC/NJ area experienced an alarming peak of $PM_{2.5}$, with levels reaching an unprecedented 203.5 μg/m^3 (on 6/7/2023) which is about 17 times the healthy $PM_{2.5}$ level.

Vol. 2023 (2) 337 – 346
ISSN 2688-3651 [online]

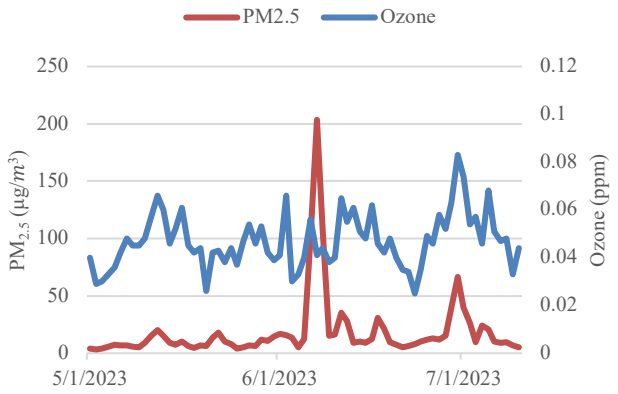


Figure 2. $PM_{2.5}$ and ozone concentrations before and after the Canada Wildfire event (June 6-8) in the Queens College 2 site

In terms of ozone measurements, while wildfires emit certain pollutants that are involved in ozone formation, their contribution to ozone levels in distant states like NYC is minimal. Ozone levels in a specific area are predominantly influenced by local emissions such as VOCs (volatile organic compounds), NOx (nitrogen oxides), and meteorological conditions, not long-range transmissions, making it unlikely for a Canada Wildfire to bring forth a significant drop throughout ozone levels in NYC. The lack of a peak in ozone levels further indicates that the ozone has less association with Canada wildfire in terms of long-range transport.

As shown in Figure 2, the peak value of $PM_{2.5}$ observed at NYC during the Canada Wildfire events was as high as 203.5 µg/m^3, with the average of the $PM_{2.5}$ values during the peak days (6/6/2023 – 6/8/2023) being 137.13 µg/m^3. The typical annual average of $PM_{2.5}$ in NYC was about 9.35 µg/m^3 from May 1st, 2023, to June 5th, 2023 (5/1/2023 – 6/5/2023). Therefore, not only was there a 194.15 µg/m^3 difference in the peak day compared to the average $PM_{2.5}$ level, but the contribution from the Canada Wildfires in terms of its long-range transport on the peak day also sparked a 2076% increase from the average $PM_{2.5}$ value, and the overall average of the peak days sparked a 1366% increase from the average $PM_{2.5}$ value. Additionally, the average $PM_{2.5}$ value is only 7% of the average $PM_{2.5}$ value of the peak days, indicating that the peak days made up 93% of the $PM_{2.5}$ in NY during the height of the wildfire in NY.

The lasting effects of the Canada Wildfires are further visible through a notable increase in average $PM_{2.5}$ values during the aftermath of the critical days (6/9/2023 – 7/10/2023). The average $PM_{2.5}$ managed to lower to 17.32 µg/m^3, and although this suggests a 119.81 µg/m^3 decrease from the average $PM_{2.5}$ levels during the peak dates, the average $PM_{2.5}$ value from the dates before and after the fire had increased 7.97 µg/m^3 (an 85% increase), which shines a light on established mark of the Canada Wildfire.

During late June and early July, another Canada wildfire event seemingly influenced $PM_{2.5}$ levels but had a less pronounced impact on ozone concentrations. While the $PM_{2.5}$ levels showed a significant increase, the ozone levels remained relatively consistent with typical days.

Fig. 3 delineates the $PM_{2.5}$ concentrations across a diverse array of locations in NYC, covering the period from 5/2/2023 to 6/28/2023. Prior to the apex of the Canada Wildfire (5/2/2023-6/5/2023), the mean $PM_{2.5}$ concentration stands at 7.88 µg/m^3. This value escalates to an average of 140.2 µg/m^3 during the wildfire's highest peak days (6/6/2023-6/8/2023), marking an increase of 132.32 µg/m^3 in and a percent increase of 1679% in $PM_{2.5}$ concentration between normal days and the peak dates of the fire.

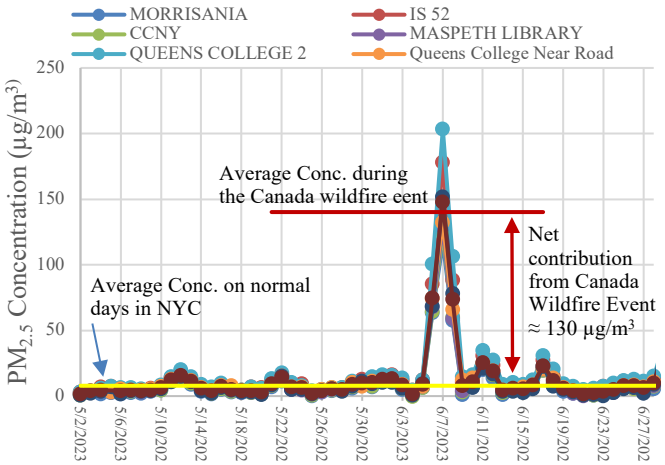


Figure 3. Analysis of $PM_{2.5}$ levels throughout larger range of locations in NYC (locations labeled above)

Furthermore, by taking the $PM_{2.5}$ values from 6/7/2023, the peak date of the wildfires, it can be noted that the highest value is 203.5 µg/m^3 (from Queens College 2), and the lowest value is 116.9 µg/m^3 from Morrisania). Compared to the average $PM_{2.5}$ level in NYC, the value 203.5 µg/m^3 presents a 195.62 µg/m^3 increase and the value 116.9 µg/m^3 presents a 109.03 µg/m^3 increase from the normal average of 7.88 µg/m^3. Therefore, the net contribution

Vol. 2023 (2) 337 – 346
ISSN 2688-3651 [online]

from the Canada Wildfire ranged from 109 µg/m^3 to 196 µg/m^3, with the average of 132 µg/m^3 as listed above. These observations were drawn from sites situated within a 32-mile radius, stretching from Staten Island to Queens, underscoring the extensive adverse influence of the Canada Wildfire across New York State.

3.2 HYSPLIT 4 Back Trajectory Models Using Langrangian Dispersion Scheme

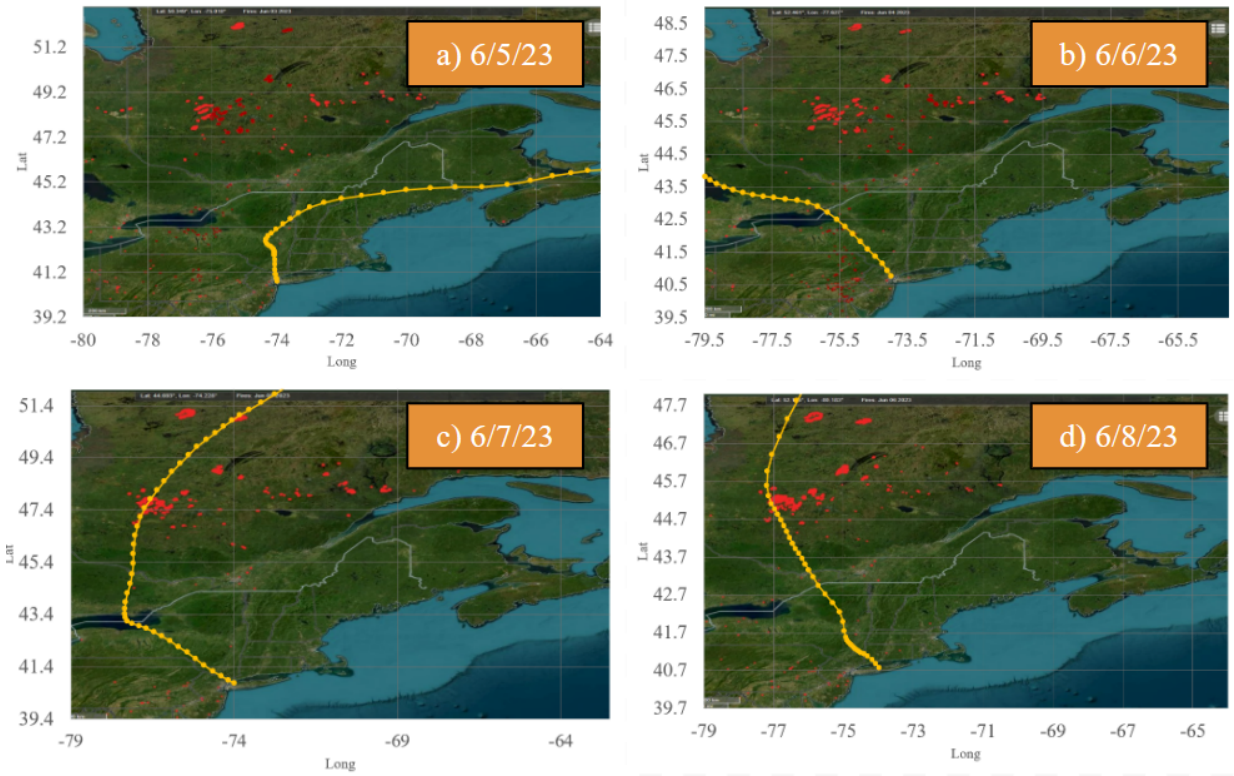


Figure 4. 48-hr back-trajectory of Canada Wildfire on June 5th (a), Junne 6th (b), June 7th (c), June 8th (d). Each trajectory endpoint indicates 1-hr air mass travel.

Using the advanced back trajectory patterns generated by HYSPLIT4, in combination with intricately detailed maps displaying pinpointed wildfire locations sourced from FIRMS US/Canada (Fire Information for Resource Management System US/Canada), it has become feasible to accurately track and identify specific pollution sources and their accompanying atmospheric compositions that influence the air quality in NYC. A crucial observation to underscore is the data from days like June 5th (Fig. 4 (a)) and June 6th (Fig. 4 (b)). During these periods, the origin of pollutant sources doesn't appear to be tied geographically to Canada. Instead, indications point towards more local or regional sources of pollution with shorter atmospheric trajectories.

These proximate sources exert a pronounced influence on the surrounding environment. The primary contributors include gases released from vehicle emissions, active construction sites, and ongoing industrial processes. Such localized sources are often underestimated but play a pivotal role in influencing air quality in urban regions.

Conversely, during peak days like June 7th (Fig. 4 (c)) and June 8th (Fig. 4 (d)), there's undeniable evidence showcasing the impact of transcontinental smoke transport originating from the Canada wildfires. This is evident in the discernible alterations of the trajectories on these dates. Specifically, the pathways on both June 7th and June 8th are oriented such that they directly intersect and traverse Canada, focusing particularly on the Ottawa Province. These trajectory shifts are not coincidental. Given that dispersion models have proven adept at predicting the direction and density of smoke dispersion, their data corroborates the notion that the atmospheric trajectories during these peak days were significantly affected by the massive wildfires in Canada. This goes to show the profound and far-reaching impacts of large-scale environmental events, transcending national borders and affecting regions thousands of miles away.

Vol. 2023 (2) 337 – 346
ISSN 2688-3651 [online]

3.3 Modeling Fire Trajectory With and Without Deposition

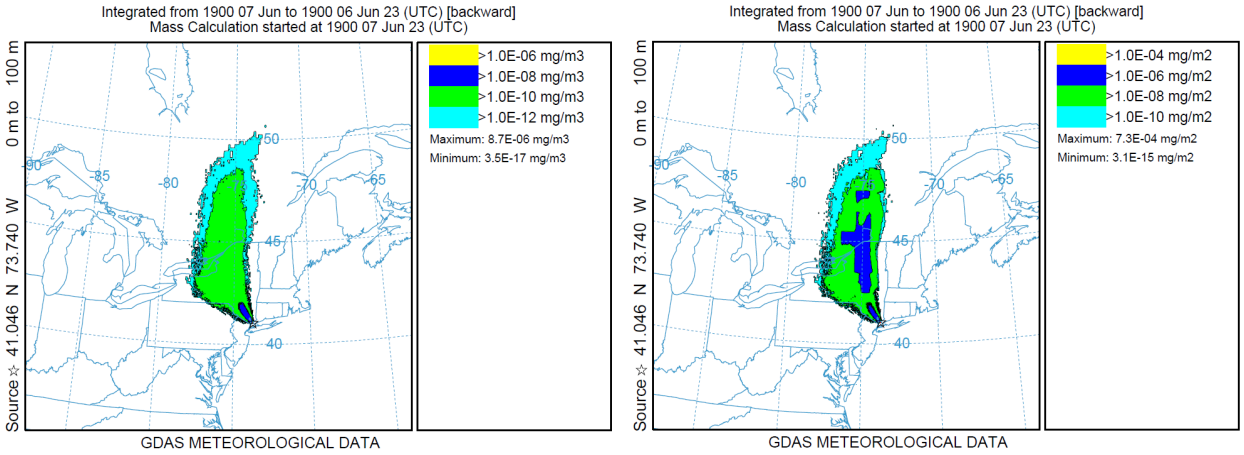


Figure 5. Simulations of 24-hr backward dispersion concentration on June 7th at NYC without deposition (Left) with deposition (Right).

Fig. 5 presents HYSPLIT4 48-hour backward dispersion concentration simulations for New York City (NYC) dated June 7th. On the left, the simulation, which does not account for deposition, stands in stark contrast to its counterpart on the right that factors in deposition effects. The former simulation reflects conditions without the influence of global warming, while the latter potentially signifies the impact of global warming (Seinfeld, 2016). The dry deposition velocity for the simulation excluding deposition was set at a default rate of 0.001 meters/second. In comparison, the simulation that accounted for deposition had a velocity of 0 meters/second. Both simulations assessed concentrations from ground level up to an altitude of 100 meters. Intriguingly, the deposition-inclusive simulation exhibited a dramatically higher concentration, with a peak value of 7.3E-04 $\mu g/m^3$, as opposed to the 8.7E-03 $\mu g/m^3$, observed in the non-deposition scenario. This underscores that limiting or omitting deposition can escalate concentrations considerably, here by a factor of 89. Such elevated concentrations, as evidenced in the right simulation, may hint at the repercussions of global warming. Temperature increase brought about by global warming can curtail deposition velocities, resulting in heightened accumulation of pollutants from wildfires in the atmosphere, subsequently deteriorating air quality.

4. Limitations and Assumptions

While the HYSPLIT4 model was pivotal in our analysis, it's worth noting certain limitations. The model provides concentration data only for the outermost contours, restricting our ability to gauge variations in concentrations within these contours, especially when comparing simulations with and without deposition. This limitation posed challenges in quantifying differences between these two scenarios. Additionally, HYSPLIT4 could benefit from enhanced flexibility in its parameter options. The current model does not easily accommodate alterations to atmospheric physics parameters, such as the secondary formation of $PM_{2.5}$. Such capabilities would have been valuable, especially during the Canada Wildfire event, allowing for a more nuanced simulation of secondary particulate formation. Furthermore, employing enhanced grid-based modeling approaches such as Community Multiscale Air Quality (CMAQ) could offer more refined data, enhancing our capacity to better quantify the impacts of global warming on atmospheric dispersion during wildfire events.

5. Conclusion

In this research, we meticulously examined specific locales within New York City and various air monitoring stations across New York state to discern the true ramifications of the Canada wildfires on air quality and their possible correlation with global warming. Leveraging the HYSPLIT4 modeling framework, we analyzed patterns in $PM_{2.5}$ and

Vol. 2023 (2) 337 – 346
ISSN 2688-3651 [online]

Ozone concentrations. Furthermore, using the Lagrangian dispersion scheme, we investigated the pathways taken by pollutants emanating from the wildfires. Our models, which included HYSPLIT4 simulations considering scenarios with and without dispersion, revealed that the Canada Wildfires had a profound effect on NY's dispersion trajectories and $PM_{2.5}$ concentrations across different urban pockets within a 32-mile radius of NYC. Remarkably, there was a staggering 1679% surge in $PM_{2.5}$ concentrations due to the wildfires, and the aftermath saw an enduring 85% uptick in $PM_{2.5}$ levels in zones like Queens College 2. Moreover, during critical days like June 7th and June 8th, the dispersion trajectories underwent a notable shift. Previously dominated by local industrial pollution sources, the trajectories realigned, suggesting pollutants would be funneled directly from the fire sites to NYC.

Coupled with these findings is the startling revelation that global warming has exacerbated NYC's air quality concerns. Specifically, our data indicates that the influence of global warming during the Canada Wildfire event has amplified NYC's air quality to levels that are 89 times higher than what might have been anticipated without such global warming.

Drawing from these extensive analyses, two salient themes emerge. First, the Canada wildfires not only directly affect the immediate vicinity but also have far-reaching repercussions on regions such as NYC. Second, the overarching shadow of global warming acts as a catalyst, exacerbating the effects of such environmental disasters. Together, these insights underscore the urgent need for heightened awareness and proactive measures, not just for areas directly hit by wildfires but for surrounding regions that bear the indirect brunt of these calamities.

References

Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. Proceedings of the National Academy of Sciences, 113(42), 11770-11775.

Brook, R. D. et al. (2010). Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. Circulation, 121(21), 2331-2378.

CAMS. (2023, August 3). *A record-breaking boreal wildfire season.* Atmosphere Monitoring Service. https://atmosphere.copernicus.eu/record-breaking-boreal-wildfire-season

Cohen, L. (2023, June 7). New York City Air becomes some of the worst in the world as Canada wildfire smoke blows in. CBS News. https://www.cbsnews.com/news/new-york-city-second-worst-air-quality-in-world-canada-wildfire-smoke/

Dockery, D. W. et al. (1993). An association between air pollution and mortality in six U.S. cities. New England Journal of Medicine, 329(24), 1753-1759.

Dominici, F. et al. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. JAMA, 295(10), 1127-1134.

Draxler, R. R., & Hess, G. D. (1997, June). *An overview of the HYSPLIT_4 modelling system for trajectories, dispersion, and deposition.* NOAA Air Resources Laboratory. https://www.arl.noaa.gov/documents/reports/MetMag.pdf

Draxler, R. R., & Hess, G. D. (1998). An overview of the HYSPLIT_4 modelling system for trajectories, dispersion, and deposition. *Australasian Physical & Engineering Sciences in Medicine*, 47(4), 295–308. https://www.arl.noaa.gov/documents/reports/MetMag.pdf

Environmental Protection Agency (EPA). (2012, December 14). Revised Air Quality Standards for Particulate Pollution and Updates to the Air Quality Index (AQI). The National Ambient Air Quality Standards for Particle Pollution. https://www.epa.gov/sites/default/files/2016-04/documents/2012_aqi_factsheet.pdf

Ford, B. et al. (2018). Future fire impacts on smoke concentrations, visibility, and health in the contiguous United States. *GeoHealth*, 2(8), 229–247. https://doi.org/10.1029/2018gh000144

Vol. 2023 (2) 337 – 346
ISSN 2688-3651 [online]

Intergovernmental Panel on Global Warming (IPCC). (2021). *Global Warming 2021: The Physical Science Basis.* Contribution of Working Group I to the Sixth Assessment Report. https://www.ipcc.ch/report/ar6/wg1/

Liu, Y. et al. (2020). *Fire behavior and smoke modeling: Model improvement and measurement needs for next-generation Smoke Research and forecasting systems.* International journal of wildland fire. 10.1071/wf18204

Liu, Y., Stanturf, J., & Goodrick, S. (2010). Trends in global wildfire potential in a changing climate. *Forest Ecology and Management,* 259(4), 685-697. https://acp.copernicus.org/articles/5/799/2005/

New York State Department of Environmental Conservation. (2023). *Air Monitoring Station Locations & Attributes Map.* New York Open Data. https://data.ny.gov/Energy-Environment/Air-Monitoring-Station-Locations-Attributes-Map/ekga-4wdn

Pope, C. A. et al. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA, 287(9), 1132-1141.

Popovich, N. (2023, July 18). How Canada's record wildfires got so bad, so fast. *The New York Times*. https://www.nytimes.com/interactive/2023/07/18/climate/canada-record-wildfires.html

Price, C. (2009). Will a drier climate result in more lightning? *Atmospheric Research*, 91(2–4). https://doi.org/10.1016/j.atmosres.2008.05.016

Reid, J. S., Koppmann, R., Eck, T. F., & Eleuterio, D. P. (2005). A review of biomass burning emissions part II: intensive physical properties of biomass burning particles. *Atmospheric Chemistry and Physics, 5.* https://acp.copernicus.org/articles/5/799/2005/

Rolph, G., Stein, A., & Stunder, B. (2017). *Real-time Environmental Applications and Display System: READY.* Environmental Modelling & Software. https://doi.org/10.1016/j.envsoft.2017.06.025

Romps, D. M. et. al. (2014). Projected increase in lightning strikes in the United States due to global warming. *Science*, 346(6211). 10.1126/science.12591

Salahieh, N., Sutton, J., & Mascarenhas, L. (2023, June 29). More than a third of the US population, from the Midwest to the East Coast, under Air Quality Alerts from Canadian wildfire smoke. CNN. https://www.cnn.com/2023/06/27/us/canada-wildfire-smoke-great-lakes/index.html

Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric Chemistry and Physics: From Air Pollution to Global Warming.* John Wiley & Sons. https://www.wiley.com/en-us/Atmospheric+Chemistry+and+Physics%3A+From+Air+Pollution+to+Climate+Change%2C+3rd+Edition-p-9781118947401

Stein, A.F., Draxler, R.R, Rolph, G.D., Stunder, B.J.B., Cohen, M.D., & Ngan, F. (2015). NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bull. Amer. Meteor. Soc.*, 96. http://dx.doi.org/10.1175/BAMS-D-14-00110.1

U.S. Environmental Protection Agency. (2023). *Download daily data.* https://www.epa.gov/outdoor-air-quality-data/download-daily-data

Journal of Research High School

Vol. 2023 (2) 347 – 352
ISSN 2688-3651 [online]

Investigate the Link Between the hMLH1 Gene and Microsatellite Instability (MSI) in Colorectal Carcinoma (CRC)

Sareena Garg,[1] * Jeru Manuel[2]

[1]Notre Dame High School, San Jose, CA, USA
[2]MENTORCONNECT
*Corresponding Author: sgarg24@ndsj.org

Advisor: Dr. Jeru Manuel, jerumm@gmail.com

Received June 19, 2023; Revised September 26, 2023; Accepted, November 1, 2023

Abstract

Colorectal carcinoma is one of the most common cancers in the world, with the primary causal factor being genomic instability. Microsatellite Instability is an indicator of an increased tendency of genome alterations which may be caused by defective mismatch repair pathway. *hMLH1, is* a critical *m*ismatch repair pathways gene known to be mutated in Colorectal Carcinoma. Aim of the study was to better understand the relationship between *hMLH1* and Microsatellite Instability in Colorectal Carcinoma using publicly available datasets. An *in silco* study was performed using a dataset retrieved from cBioportal and the Cancer genome atlas. The results demonstrated that low *hMLH1* expression had significantly higher MSI as expected. However, it interestingly showed an inverse association between *hMLH1* and fraction of genome alterations. This possibly highlights the complexity of the link between *hMLH1* and Microsatellite Instability.

Keywords: Colorectal carcinoma, Microsatellite instability, MSI, hMLH1 gene, fraction of genome altered

1. Introduction

Colorectal Carcinoma (CRC) is cancer of the colon or rectum, arising from precancerous polyps, adenomatous polyps, or serrated polyps that form over several years. It is the third most common cancer and the fourth most common cause of cancer-related deaths globally (Herman et al., 1998). CRC, like most cancers is caused by mutations in critical genes like DNA repair mechanisms resulting in alterations of the genome. Depending on the origin of the mutation, CRC is classified as sporadic, inherited, or familial (Marmol et. al., 2018). Genomic instability, the increased tendency of genome alteration during cell division, is an important feature underlying CRC. Microsatellites are short non-coding repeating sequences throughout the genome. They occur at thousands of locations and have a higher mutation rate. When instable, the specific genes that monitor genomes for errors are unable to correct thus leading to instability. Microsatellite Instability (MSI) which is caused by a hypermutable phenotype, due to loss of DNA repair mechanisms, are one of the underlying mechanisms that causes this instability (Kawakami et al., 2015). The mutations can affect non-coding regions as well as codifying microsatellites, resulting in the reading frames of oncogenes or tumor suppressor genes being altered and tumors developing. Fraction of genomic alterations (FGA) includes measuring the percentage of copy number altered chromosome regions due to cancer-driven mutations, gene fusions, amplification, deletion, and post-transitional modifications.

Mismatch repair (MMR) pathways play a vital role in identifying and repairing mismatched bases during DNA replication and genetic recombination in normal and in cancerous cells (Sameer et al., 2014). Defects in MMR are also known to cause subsequent high MSI which leads to the accumulation of a mutation load. MMR genes mutated in tumors with MSI include *hMSH2, hMLH1, hMSH6,* h*PMS1* and *hPMS2.* MLH1/MSH2 phenotype constitutes a

Vol. 2023 (2) 347 – 352
ISSN 2688-3651 [online]

pathologically and clinically distinct subtype of sporadic CRC (Richman, 2015). These markers are important to define therapeutic strategy in CRC. Several clinical trials have demonstrated that MMR deficiency or high MSI is significantly associated with long-term immunotherapy-related responses and better prognosis in CRC (Richman, 2015). *hMLH1* has an impact on the fraction of genome alterations, because the lower the range the less altered the genome is and the higher the range the more altered the genome is, therefore the *hMLH1* gene expression leads to MSI which is a biomarker for CRC (Marmol et al., 2017). The goal of this study was to understand better the expression of *hMLH1* and corresponding MSI. It was predicted that CRC patients may have a lower expression of *hMLH1* and a corresponding higher MSI. By analyzing publicly available datasets on the *hMLH1* expression, MLH1/MSH2 phenotype, the MSI status and genomic alteration in CRC patients, the link between *hMLH1* and MSI was better understood.

2. Materials And Methods

The *in silco* study was performed using the publicly available TCGA dataset: Colorectal Adenocarcinoma (cBioportal for Cancer Genomics, n.d.), licensed by National Institute of Health (NIH).

2.1 Data processing

The dataset was sorted to specifically test the hypothesis by applying criteria that made sure no patients had information missing. Only patients with the following variables were included: sex type, cancer stage, cancer type, MSI type, MLH1 silencing (an epigenetic modification that prevents the expression of *hMLH1*), and vital status. The sorted dataset was matched to the original clinical data to make sure all patient IDs were the same. Out of the 276 patients originally, 202 patients remained. All files chosen were imported to Microsoft Excel and then sorted and characterized based on frequency distribution. The frequency distributions across the variables in the dataset are shown in Table 1.

2.2 Data analysis

The file was imported into RStudio – 1.4.1106 and DATAtab for further statistical analysis (DATAtab, n.d.). Histogram and Box & whisker plots were created through Excel and cBioportal (cBioportal for Cancer Genomics, n.d.).

Shapiro-Wilk normality distribution test was conducted to check that continuous variables in the datasets followed normal distribution. T-tests were considered initially to check for the correlation between datasets, but results from Shapiro-Wilk test showed that the datasets being considered did not have normal distribution, and so the Mann-Whitney U test, a non-parametric test that does not need uniform distribution, was used to test for correlation between the continuous variable datasets. The null hypothesis for the Mann-Whitney test states that there is no difference between the datasets if the p-value > 0.05 for 5% significance, and the alternate hypothesis states that there is a significant difference between the datasets if the p-value < 0.05. Chi-square test was used to test the independence of categorical datasets. Chi-square test has the same null and alternate hypothesis as the Mann-Whitney U test.

All instruments used allowed for a graphical view of the results and indicated the difference in the means. Larger difference in means meant more significant differences between variables.

3. Results

3.1 Characteristics of the dataset

Once data sorting was completed there were 202 patients that satisfied the selection criteria, the details are shown in Table 1. MSI high was defined when two of five markers showed instability in the genome. MSI low was determined when only one MSI marker showed instability and the rest showed stability. Note that *hMLH1* refers to

Vol. 2023 (2) 347 – 352
ISSN 2688-3651 [online]

the gene, while MLH1 refers to the protein created by *hMLH1* gene. Units for MLH1 expression are RPKM or Reads per kilo million mapped reads in a library. (Biostars, n.d.)

$$RPKM\ of\ a\ gene = \frac{Number\ of\ reads\ of\ a\ gene\ \times\ 10^3\ \times\ 10^6}{Total\ number\ of\ mapped\ reads\ from\ given\ library\ \times\ gene\ length}$$

Reads in the above equation for MLH1 expression scoring refer to immunohistochemical data scored regarding staining intensity (negative, weak, moderate or strong) and fraction of stained cells (<25%, 25-75% or >75%). (The Human Protein Atlas, n.d.). Any patient with Tumor stage 1 or 2 in the dataset were put into the Cancer Stage category Low while any patient with Tumor stage 3 or 4 cancer was put into the Cancer Stage category High. The data with the variables sex type, cancer stage, cancer type, MSI type, MLH1 silencing, and vital status was grouped for certain tests in the report based on the median of the MLH1 expression.

It was observed in the data, 100% of patients had cancer with ~44% having high stage cancer, and ~58% having colon cancer, yet surprisingly only 9.4% of the patients had died. A reason for higher survival could be the stable microsatellite stability status observed in 68.8% of the patients.

Table 1. Cohort characteristics

Total Sample Size = 276. Selected Samples = 202.		
Variables (n)	No. of Patients	Frequency (%)
Sex (202)		
Female	97	48.0%
Male	105	52.0%
Cancer Stage (202)		
High	89	44.1%
Low	113	55.9%
Cancer Type (202)		
Colon	118	58.4%
Colorectal	34	16.8%
Rectal	50	24.8%
MSI Type (202)		
MSS	139	68.8%
Low MSI (MSI-L)	35	17.3%
High MSI (MSI-H)	28	13.9%
MLH1 Silencing (202)		
0 (Not Silenced)	177	87.6%
1 (Silenced)	25	12.4%
Vital Status (202)		
Alive	183	90.6%
Dead	19	9.4%

3.2 MSI type and *hMLH1* expression

MSI was seen in ~30% of the patients among which ~44% had high MSI. Figure 1, a box and whisker plot graph, shows the MLH1 expression on the y axis and the three MSI types on the x-axis. The graph shows a significant difference in MLH1 expression between the Low MSI (MSI-L) and High MSI (MSI-H) samples. MSI-L samples had a mean of 20.9 RPKM and a standard deviation of 6.7, while MSI-H samples had a much smaller mean of 8.0 RPKM but a larger standard deviation of 8.0 because of a few outliers.

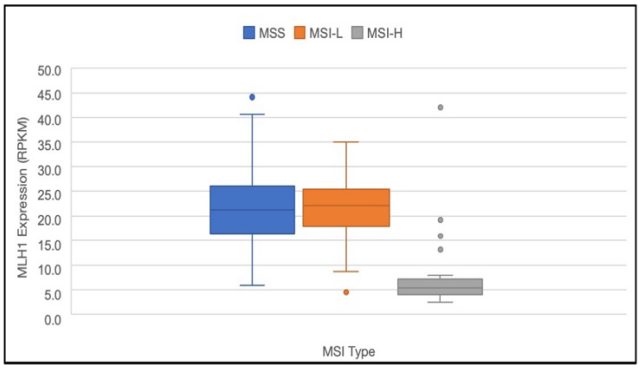


Figure 1: Compares the MLH1 expression with the three MSI types – MSS, MSI-L, and MSI-H.

To test the hypothesis statistically, further analysis was done to first verify that the MSI-H and MSI-L samples were statistically independent, and subsequently the relationship between MLH1 expression and MSI type was checked.

Running the Shapiro-Wilk normality test indicated that MSI-L dataset was normally distributed, while MSI-H dataset was not. Mann-Whitney U test, a non-parametric test, revealed that the difference between MSI-L and MSI- H with respect to MLH1 expression was highly statistically significant, p-value=<.001, r=0.7. If p-value < .05 for 5% significance, MSI-L and MSI-H are considered to be significantly different.

Vol. 2023 (2) 347 – 352
ISSN 2688-3651 [online]

The estimated median of the 206 patients with MLH1 expression was 20.03 RPKM. Based off this, the patients were characterized in to two categories of high and low MLH1 expression for further analysis. In Figure 2, Group A has low MLH1 expression values, while Group B has high MLH1 expression values. This graph shows that patients with low MLH1 expression (Group A) had significantly high MSI-H, while patients with high MLH1 expression had practically no MSI-H.

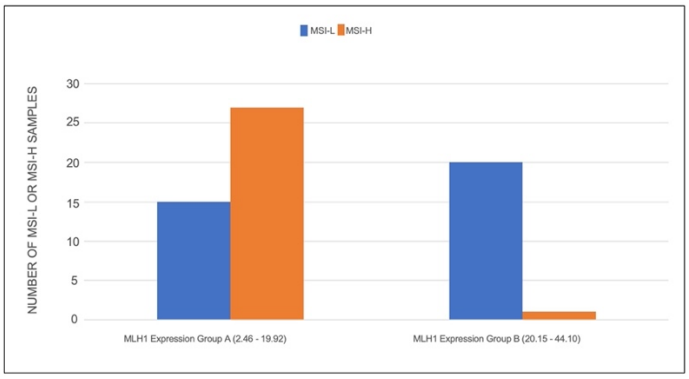


Figure 2: Comparison of number of samples with MSI-L and MSI-H between two groups categorized by MLH1 Expression.

A Chi-square test was performed between MLH1 Expression Groups and MSI Status, and a statistically significant relationship was found, p = <.001. The Chi2 test is therefore significant and the null hypothesis that the two groups are independent is rejected.

This validates the hypothesis that high MSI samples are significantly different from low MSI samples, and that there is a significant relationship between MLH1 expression and MSI.

3.3 High correlation between MLH1 expression and MLH1 Silencing

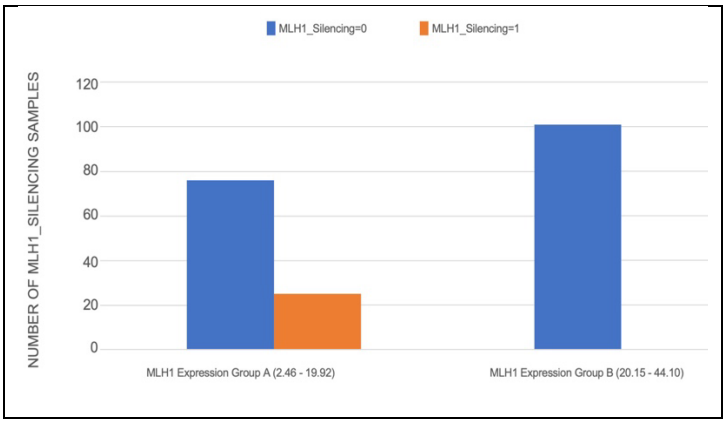


Figure 3: Compares number of samples in low and high MLH expression groups for the two cases of MLH1_Silencing.

MLH1_Silencing=0 represents patients who don't have their *hMLH1* gene silenced, while MLH1_Silencing=1 represents the patients who have their *hMLH1* gene silenced. The Relationship between MLH1 expression and MLH1_Silencing is explored in Figure 3, which shows that the presence of MLH1_Silencing corresponds to a reduction in MLH1 expression and therefore there are no samples of Group B with MLH1_Silencing=1. As expected, the samples are present in both groups when MLH1_Silencing=0.

Results of Chi-square test to test for independence between MLH1 Expression Groups and MLH1_Silencing showed that there was a statistically significant relationship between them, p = <.001. The calculated p-value of <.001 is lower than the defined significance level of 5%. The Chi-square test is therefore significant and the null hypothesis that they were independent was rejected.

This result is significant because it validates the relationship between MLH1 Silencing and *hMLH1* expression.

3.4 Patients with low *hMLH1* expression have a less altered genome

The impact of *hMLH1* on genome alterations was further analyzed. The Mann-Whitney U test showed that the difference between MLH1 Expression Group A (2.46 - 19.92) and MLH1 Expression Group B (20.15 - 44.10) with respect to FGA was statistically significant, p=.003. As demonstrated in Figure 4, it was found that lower expression of *hMLH1* gene demonstrates lower Fraction Genome Altered (FGA). Group A has a mean of 0.22 and a standard deviation of 0.18, while Group B has a higher mean of 0.31 and a higher standard deviation of 0.2. This potentially indicates an adverse impact of higher MLH1 expression and was contradictory to the hypothesis.

Vol. 2023 (2) 347 – 352
ISSN 2688-3651 [online]

4. Discussion

CRC, owing to its high incidence and mortality rate, is of big concern especially in the developed world. Based on phenotype researchers have classified it biologically into two types: 1. MSI and 2. MSS but chromosomally unstable. MSI tumors can help doctors identify the potential cause of the tumor. It also helps with determining the therapy module. The latest developments indicate that immunotherapy, a very new treatment regimen had higher efficacy when there is High MSI.

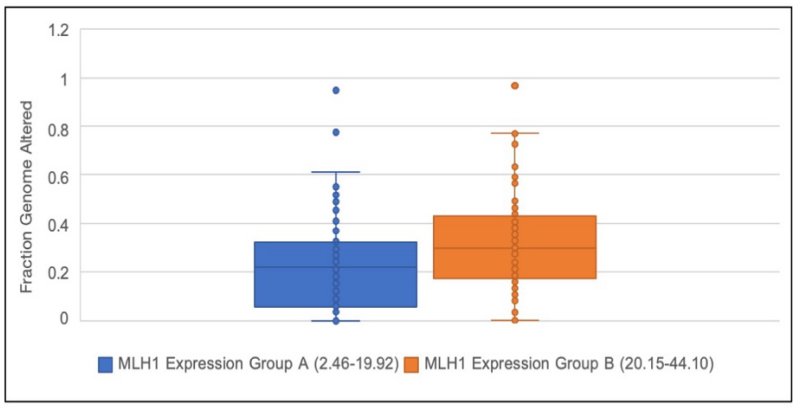


Figure 4: Comparison of Fraction of Genome Altered with MLH1 Expression Groups.

It is also known that MLH1 silencing is one of the causes for MSI. However, only about 30% of CRC tumors are seen with MSI, and among them only 50% have high MSI (Marmol et al., 2017). There is always a need to better understand the underlying molecular cause of MSI, as the complexity of CRC is quite significant. Owing to the interest in studying CRC, the aim was to explore the publicly available dataset to study the link between MSI and *hMLH1* expression and its impact on genomic alterations. The goal of this study was to analyze the impact of low *hMLH1* expression on MSI and its role in genomic alterations in CRC. The cohort was validated, and it was found that the sorted dataset demonstrated a similar frequency of MSI that aligns with the literature. On performing further analysis, it was identified that *hMLH1* expression and MLH1 silencing were significantly associated. Low *hMLH1* expression demonstrated high MSI, whereas the patients with high expression of *hMLH1* showed low MSI. These findings validated prior studies about the role of MLH1 and MSI. An important and interesting observation was the inverse association between *hMLH1* expression and FGA. This contradicts the expected findings and throws open a lot of questions which require research to better understand the implication. It however emphasis the complexity of molecular biological process and highlights the need to explore further the basic biological questions using public datasets. However, that was not within the scope of this research project. But the future does look exciting to unravel these further with advancement of better technology and improved data analysis methods.

5. Conclusion

The purpose of this study was to analyze the impact of low *hMLH1*expression on MSI in CRC. From the analysis it was concluded that patients with low *hMLH1* expression have significantly higher MSI. This led us to explore the idea that MSI-H can show a trend toward a better prognosis for CRC and reviving *hMLH1* expression could restore MMR activity which could eventually result in lower MSI. This means that less people will have their life threateningly affected by colorectal carcinoma and immunotherapy can allow patients to recover from cancer in a smaller timeframe. Looking ahead, it may be valuable to research if methylation of *hMLH1* in pre-malignant adenomatous polyps is an early event in carcinogenesis of CRC. Studying methylation is advantageous, as it could yield potential targets for treatment to combat tumor progression.

Acknowledgment

I would like to thank Skoolmentor, late Nutan Malhotra, and especially my mentor Dr. Manuel for teaching me, guiding me throughout this process and helping me with proofreading of this paper.

Vol. 2023 (2) 347 – 352
ISSN 2688-3651 [online]

References

Biostars. (n.d.). *Blog:Gene expression units explained: RPM, RPKM, FPKM and TPM.* https://www.biostars.org/p/273537/#:~:text=Gene%20expression%20units%20explained%3A%20RPM%2C%20RPKM%2C%20FPKM%20and%20TPM&text=In%20RNA%2Dseq%20gene%20expression,units%20from%20mapped%20sequence%20data.

Bloom, A. (2023) *What Is Microsatellite Instability?* MD Anderson Cancer Center. https://www.mdanderson.org/cancerwise/what-is-microsatellite-instability-MSI.h00-159617067.html

cBioportal for Cancer Genomics. (n.d.). *Colorectal Adenocarcinoma (TCGA, Nature 2012).* https://www.cbioportal.org/study/summary?id=coadread_tcga_pub

DATAtab. (n.d.). *Statistics Calculator.* https://datatab.net/statistics-calculator/charts

Fleming, M., et al. (2012). Colorectal carcinoma: Pathologic aspects. *Journal of Gastrointestinal Oncology*, 3(3):153-73. doi: 10.3978/j.issn.2078-6891.2012.030. PMID: 22943008; PMCID: PMC3418538.

Gelsomino, F., et al. (2016). The evolving role of microsatellite instability in colorectal cancer: A review. *Cancer Treatment Reviews*, 51:19-26. doi: 10.1016/j.ctrv.2016.10.005. Epub 2016 Oct 27. PMID: 27838401.

Herman, J. G., et al. (1998). Incidence and functional consequences of *hMLH1* promoter hypermethylation in colorectal carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95 (12) 6870-6875 https://doi.org/10.1073/pnas.95.12.6870

Kawakami, H., Zaanan, A., and Sinicrope, FA. (2015). Microsatellite instability testing and its role in the management of colorectal cancer. *Current Treatment Options in Oncology*, 16(7):30. doi: 10.1007/s11864-015-0348-2. PMID: 26031544; PMCID: PMC4594190.

Mármol, I., et al. (2017). Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *International Journal of Molecular Sciences*, 18(1):197. doi:10.3390/ijms18010197. PMID: 28106826; PMCID: PMC5297828.

Richman, S., (2015). Deficient mismatch repair: Read all about it (Review). *International Journal of Oncology*, 47(4):1189-202. doi:10.3892/ijo.2015.3119. PMID: 26315971; PMCID: PMC4583524.

Sameer, A. S., et al. (2014). Mismatch Repair Pathway: Molecules, Functions, and Role in Colorectal Carcinogenesis. *European Journal of Cancer Prevention*, vol. 23, no. 4, pp. 246–57. JSTOR, https://www.jstor.org/stable/48504328.

The Cancer Genome Atlas (TCGA) (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407), 330–337. https://doi.org/10.1038/nature11252

The Human Protein Atlas. (n.d.). *Help & FAQ. How is protein expression scored?* https://www.proteinatlas.org/about/help#4

Yurgelun, M. B., et al. (2017). Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. *Journal of Clinical Oncology*, 35(10):1086-1095. doi:10.1200/JCO.2016.71.0012. PMID: 28135145; PMCID: PMC5455355.

Journal of Research High School

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

The Role of Epithelial Markers in Breast Cancer Metastasis: Systematic Review and Meta-Analysis

Mihaela Tzvetkova[1] *, Nia Dimitrova[1]

[1]American College of Sofia, Sofia, Bulgaria
*Corresponding Author: m.tzvetkova24@acsbg.org

Advisor: Dr. Ganka Dineva, g.dineva@acsbg.org

Received July 24, 2023; Revised October 4, 2023; Accepted, October 30, 2023

Abstract

Epithelial-mesenchymal transition is believed to be a fundamental component of cancer metastasis. Hence, epithelial markers have emerged as potential therapeutic targets and diagnostic markers of metastatic cancers, leading to their significance in cancer research. In this review, studies on 15 different markers were identified to elucidate further the role of epithelial markers in breast cancer metastasis. Based on the studies, the respective role of the epithelial markers in metastatic breast cancer was derived. The cellular mechanisms guiding the markers' behavior were investigated by identifying and describing their associated miRNAs. The studies of 5 epithelial markers that had identified cellular mechanisms affecting breast cancer metastasis were screened for undergoing meta-analysis. Twenty-one studies in total had sufficient data to undergo meta-analysis. Based on the content of the studies and the conducted meta-analysis, the results' limitations, strengths, and implications were discussed in detail. Although, due to the limited amount of studies, definite conclusions cannot be made, the meta-analysis revealed novel inferences and confirmed inferences made by other researchers on the role of the specific epithelial marker in metastatic breast cancer. Additionally, the study provides insight into significant gaps in the field and urges greater exploration of the topic.

Keywords: Epithelial markers, Breast cancer, EMT, Metastasis, miRNA

1. Introduction

Most instances of morbidity and mortality due to malignant tumors in women are correlated with breast cancer. Most breast cancer-related deaths occur due to metastasis, or the process by which an original primary tumor evolves to a distal secondary tumor (Hagemeister et al., 1980). Metastasis is a highly complex process that requires epithelial-mesenchymal transition (EMT) (Sun et al., 2020). During EMT epithelial cells repress their epithelial characteristics and acquire mesenchymal, as a result of changes in gene expression and gene regulation mechanisms. Since the loss of epithelial markers is associated with EMT, and hence metastatic progression, revealing their status in breast cancer metastasis can potentially reveal new therapeutic targets and biomarkers for diagnosis of metastatic breast cancer (MBC) (Tyler & Tirosh, 2021). The objective of this study is to provide a comprehensive review and meta-analysis on the role of epithelial markers in MBC, and to explore their transcriptional regulation via miRNAs. We hypothesize that specific epithelial markers play a crucial role in MBC and are influenced by distinct miRNAs, which can also serve as potential diagnostic markers and therapeutic targets.

As per the guidelines proposed by the EMT International Association (TEMTIA), we acknowledge the need of a distinct description of the cellular mechanisms guiding the role of the epithelial markers in metastasis and the contribution of genetic alterations, due to the complexity of EMT and its context dependent nature (J. Yang et al., 2020). As a result, all the epithelial markers identified in the review are associated with microRNA(miRNA), a type of transcription regulator. Not only is the inclusion of miRNAs in accordance with the guidelines proposed by

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

TEMTIA, but since miRNA dysregulation has been detected in multiple metastatic cancers, including breast cancer, identification of miRNA sequences and their roles can contribute to novel cancer detection techniques through their utilization as biomarkers and components of new treatments through gene editing techniques (J. Yang et al., 2020). It has been shown that certain miRNA changes can be corrected using miRNA mimics or antagomirs, normalizing the signaling pathways and the gene regulatory network, and reversing the phenotype in malignant cells (O'Bryan et al., 2017). Due to the emergence of precise gene-editing techniques such as CRISPR-Cas9, miRNAs have increased potential in cancer treatment and should be a focus of research (Godden et al., 2022). Following the TEMTIA guidelines, a necessary criterion for the markers, whose studies were subjected to meta-analysis, was clear cellular processes through which they influence MBC. Fundamental characteristics acknowledged in the analysis of the results are the molecular (luminal A/B, Triple-negative, and HER2+ enriched) and/or histological breast cancer subtypes of the samples included in the study, their tumor progression stage, and microenvironment (Q. Liu et al., 2017). The epithelial markers selected for the meta-analysis portion of the study due to their defined cellular mechanisms were B-Catenin, Nectin-4, MUC1, JAM-A, and CD44.

B-catenin is a multifunctional membrane protein that's a key component of cell-cell adhesion machinery as an intracellular signal inducer in the Wnt pathway (Shang et al., 2017). The Wnt/B-catenin signaling pathway has been shown to have a regulative role in multiple cell processes including cell motility, making its disruption a causative factor for multiple pathologies, including MBC (Komiya & Habas, 2008). In normal cells the absence of Wnt leads to the phosphorylation of cytoplasmic β-catenin by GSK3β and casein kinase Iα (CK Iα), which in turn prevents nuclear accumulation of β-catenin, allowing its ubiquitination and subsequent degradation by the ubiquitin/proteasome system (Shang et al., 2017). Nevertheless, when Wnt binds to Frizzled (FZD), it activates Disheveled (Dsh), whose activation inhibits GSK3β (Zeng et al., 2008). As a result, B-catenin is not degraded and accumulates in the cytoplasm and nucleus. There, it interacts with transcription coregulators like T cell factor/lymphocyte enhancer factor (Tcf/Lef), forming a B-catenin/Lef/Tcf complex. This complex transactivates the gene that encodes cyclin D1, leading to overgrowth of cells in the lobules and ducts inside the breast (Buechel et al., 2021). Nuclear accumulation of B-catenin also results in the loss of E-cadherin and consequent loss of cell polarity and adhesion, promoting the process of EMT, and therefore metastasis (Buechel et al., 2021).

Nectins are members of the immunoglobulin superfamily (IgSF) and are components of E-cadherin-based adherens junctions in epithelial cells, thereby having a vital role in the enhancement of cellular viability and movement ability (Mandai et al., 2015). Generally, studies agree that Nectin-4 is not expressed in normal epithelium, which contributes to their increased potential to act as a biomarker or treatment in MBC. Nectin-4 has been shown to affect metastasis by modulating the CXCR4/CXCL12-LYVE-1- axis (Sethy et al., 2021). Nectin-4 overexpression leads to an increase in CXCR4 expression and LYVE-1-lymphatic vessel density (LVD). Upregulation of LVD has been associated with increased invasive abilities and poor prognosis in patients (Ramani et al., 2012). CXCR4-expressing cancer cells are attracted by CXCL12- expressing organs, thereby initiating metastasis to distant organs (Guo et al., 2016). Additionally, ADAM-17, whose expression is driven by cancer stem cells, sheds the Nectin-4 ectodomain, which interacts with endothelial Integrin-B4. This interaction promotes metastasis in breast cancer stem cells by activating the Src-PI3K-AKT-iNOS axis (Siddharth et al., 2018). In particular, Nectin-4 has been shown to promote breast cancer stem cell metastasis via the Pi3k/Akt axis through WNT/β-Catenin signaling (Siddharth et al., 2017).

JAM-A is an immunoglobulin-like molecule that acts as a tight junction protein, and as such has a role in tumor cell adhesion, polarity, invasion and migration (Severson & Parkos, 2009). The cellular mechanisms through which JAM-A affects metastasis indicate that JAM-A operates differently in tissue- and cell- specific contexts. For example, by inhibiting the Akt/B-catenin signaling pathway, JAM-A disrupts Akt-mediated phosphorylation of B-catenin, thereby preventing its accumulation in the nucleus and therefore metastasis (Nava et al., 2011). In contrast, in HER2-positive breast cancer, a type of cancer that has increased proliferation ability, increased JAM-A expression promotes HER2 expression by causing the binding of FOXA1 to the HER2 gene promoter (Cruz et al., 2022). HER2 has also been shown to activate the PI3K/Akt pathway, where PI3K phosphorylation leads to Akt2 phosphorylation, whose amplification has been associated with MBC (Milella et al., 2015). JAM-A has also been shown to activate Rap1 GTPase and β1-integrin, both of which lead to increased metastatic potential of breast tumors. Rap1 activation prohibits metastasis in other types of cancers (Yi-Lei et al., 2017).

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

MUC-1 is a transmembrane membrane glycoprotein associated with the protection of the epithelial layer by providing lubrication of luminal epithelial surfaces, thereby promoting motility (W. Chen et al., 2021). By interacting with ICAM-1, an adhesion receptor, glycosylated MUC-1, facilitates the interaction between epithelial and endothelial cells. This process enables adhesion of circulating cancer cells to the inner lining of the blood vessel, directly or as a result of a precedent interaction with E–selectin (Hayashi et al., 2001). Glycosylated MUC1 also interacts with Src, a non-receptor tyrosine kinases that says a key role in signal transduction pathways, thereby inducing pro-migratory Rac1- and Cdc42-dependent actin reorganization at sites of contact with endothelial cells, which promotes an invasive phenotype in the tumor cell (Shen et al., 2008). MUC1 can also drive tumor angiogenesis by upregulating vascular-endothelial growth factor (VEGF), thereby promoting endothelial migration and tube formation (Khodabakhsh et al., 2021). Epidermal growth factor receptor (EGFR) stimulates growth of cancer cells, and activates STAT1 and STAT3 in breast cancer, which promote cell survival and motility. MUC1 and EGFR have a positive feedback relationship in breast cancer, resulting in dependence of EGFR prolongation on MUC1. Hence, STAT3 induces the expression of Twist one, which forms a complex with MUC-1 that results in its expression in an auto-indicative loop, accounting for its upregulation in breast cancer (Bitler et al., 2010). A subunit of MUC-1, MUC1-C, can also induce EMT and thus metastasis by activating the inflammatory NF-κB p65 pathway, which induces the transcription of ZEB1 and B-cell lymphoma 2-related protein A1 (BCL2A1) (Ahmad et al., 2009).

CD44 is a cell-surface glycoprotein involved in cell-cell interactions, adhesion, and motility, and CD44 has been used as a surface marker for breast cancer stem cells (CSCs) (Thapa & Wilson, 2016). Breast CSCs that exhibit CD44+/CD24- are potentially one of the main factors contributing to relapse of triple negative breast cancer (TNBC) due to their exacerbated self-renewal and differentiation abilities (X. Qiao et al., 2021). CD44 expression activates Rho GTPases and PI3K/AKT and MAPK-Ras, thereby promoting cytoskeletal remodeling and invasion. CD44 promotes cleavage of hyaluronan, resulting in modifications of the tumor microenvironment and essentially tumor progression. Expression of CD44 promotes docking of collagen specific MMP9. When MMP9 is found in the edges of migratory cells it promotes collagen degradation, thereby leading to an invasive phenotype, and cleavage of TGFB which also promotes invasion (Louderbough & Schroeder, 2011). Under specific conditions the ECM component hyaluronate stimulates CD44 to bind with merlin, a tumor suppressor protein, thereby conferring growth arrest in tumor cells (Herrlich et al., 2006). Other ways through which CD44 can prevent metastasis is by activating caspase-3 and hence promoting apoptosis of tumor cells, or by inhibiting PI3K activation/AKT phosphorylation (Ghatak et al., 2002).

The present meta-analysis distinguishes itself by integrating all studied epithelial markers linked to MBC and discussing their transcriptional regulation via miRNAs. We anticipated to encounter a correlation between epithelial markers and breast cancer metastasis, as well as miRNAs influence. Both miRNAs and markers could be used as therapeutic markers for targeted therapy. This approach results in a uniquely structured review, offering a more detailed depiction of the process than previously seen in other systematic reviews on the subject.

2. Materials and Methods

Marker selection: All known epithelial cell markers were identified via the Bio-techne database (Epithelial Cell Markers and Intracellular Molecules, n.d.), and were assessed for role in cell adhesion through the National Library of Medicine database. To identify which of the remaining molecules are the subject of breast cancer studies, advanced search was performed via marker name, boolean operator “and”, and “breast cancer metastasis”. Review articles were identified through keywords: epithelial-mesenchymal transition (EMT), tumor marker, cancer metastasis, name of epithelial cell marker and were manually searched for more references on the topic. Markers implicated in more than five breast cancer metastasis studies proceeded to the next selection stage. These markers were then examined for studies exploring their connection with miRNAs. Table 1 presents the markers that play a role in MBC, their associated miRNAs, and studies discussing their involvement in MBC. The studies of markers that had identified cellular mechanisms in the context of MBC and associated miRNAs, were discussed in detail and analyzed for meta-analysis eligibility.

Meta-analysis: Odds ratio (OR) was used to examine the association between the expression of epithelial markers

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

and their prevalence in MBC. OR represents the likelihood of an event occurring when exposed to a specific factor, in contrast to the likelihood of the event without that exposure. In the current context, the OR offers insight into the likelihood of the occurrence of MBC in the presence of an epithelial marker, in contrast to the likelihood of MBC without the epithelial marker. A 2x2 contingency table was set up, with one axis indicating the presence or absence of MBC and the other indicating the expression or non-expression of the epithelial marker. The calculated OR was derived using the formula: [OR = $\frac{ad}{bc}$], where 'a' denotes individuals with both the epithelial marker and MBC, 'b' signifies those with the marker but without MBC, 'c' represents those without the marker but with MBC, and 'd' identifies those without either the marker or MBC. In order for studies to be eligible for the meta-analysis, they should have reported their results such that marker transcription acts as a dependent variable, and variables measuring metastasis act as independent variables. +/- metastasis or +/- lymph node involvement would be seen as variables measuring metastasis. An OR value of 1 would suggest no association between the epithelial marker and MBC. In contrast, an OR greater than 1 would indicate an increased likelihood of MBC in the presence of the marker, whereas an OR less than 1 would suggest a decreased likelihood. The 95% confidence interval (CI) provides an estimation of the accuracy of the OR. A wide range of CI suggests that the OR's accuracy is low, while a narrow CI suggests greater accuracy. The 95% CI doesn't reflect statistical significance in the same manner as the p-value. However, if the 95% CI doesn't cross the null value (e.g., OR=1), it's often interpreted as evidence of statistical significance. Heterogeneity refers to the variability or differences in study outcomes. Analyzing heterogeneity offers insights into the influence of varying methodologies and conditions on the study outcomes. The heterogeneity of the data was assessed using tau square and chi square to make sure that all studies are evaluating the same effect. The tau squared (Tau^2) represents between-study variance, with elevated values indicating substantial inter-study variability. The chi-square (Chi^2) tests the hypothesis that the studies are evaluating the same effect, with a low p-value (typically < 0.05) suggesting that heterogeneity is present beyond chance. The p-value represents the probability of observing the given data, or more extreme data, under the null hypothesis of no effect. A p-value less than 0.05 is conventionally deemed indicative of statistical significance. Furthermore, the I^2 statistic provides a quantification of the proportion of total variation across studies that's attributable to heterogeneity rather than chance. In the context of this research, an I^2 value below 50% was interpreted as indicating satisfactory homogeneity. All statistical analyses were conducted using RevMan Version 5.

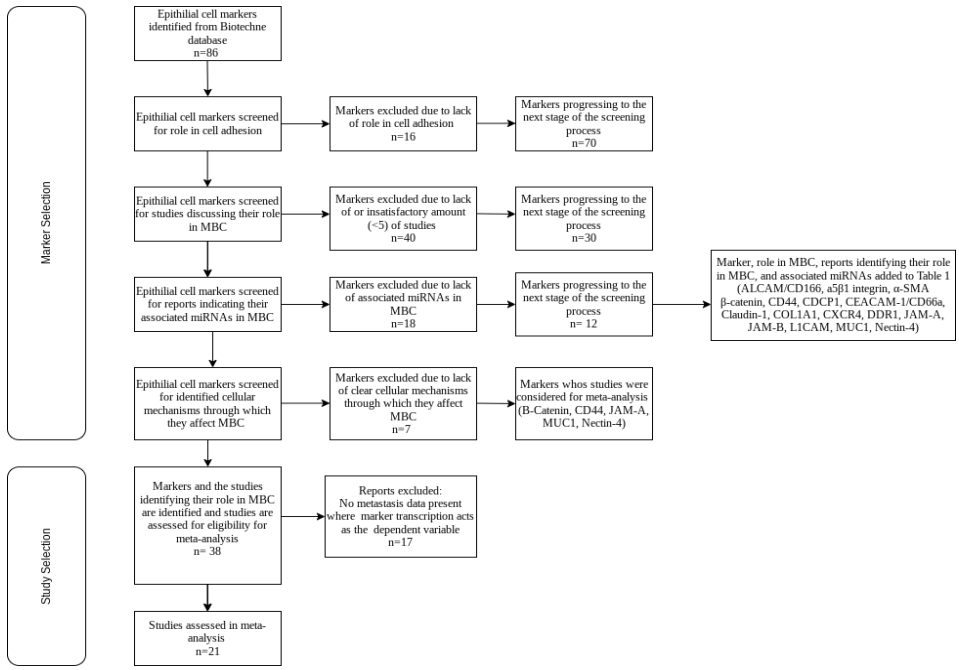


Figure 1. Flow chart representing selection procedure for epithelial cell markers and study selection for meta-analysis based on studies listed in Table 1.

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

3. Results

12 epithelial markers were shown to have a relationship with MBC in more than 5 studies, and an association with a type of miRNA, which is illustrated in Table 1. The relationship between miRNAs and metastasis was portrayed in Figure 2 and forest plots were produced to depict the statistical analysis undergone by the eligible studies of the 5 selected markers.

Table 1 Epithelial markers associated with MBC, their role in MBC, the studies discussing their role in MBC, and their associated miRNAs

Name of Marker	Role of Marker in Metastasis	Associated miRNAs	Studies discussing role of marker in metastasis
ALCAM/CD166	tumor suppressor	miR-125	(Davies et al., 2008), (Akamn et al., 2015)
α-SMA	oncoprotein	miR-200c	(Tang et al., 2015), (Mierke et al., 2011)
Integrin α5β1	oncoprotein	miR-31,-149	(Wang, Yanfang, et al. 2011), (Chan, S. 2014) , (Augoff, K., et al 2011)
β-catenin	oncoprotein	miR-200c,-29,-125b, -1229 - 141	(Z. Wang et al 2015), (Nie, J. et al., 2019), (Kwon, J. J. et al., 2019), (Liu, B. et al., 2018), (Tan, Z. et al., 2016), (Si, W. et al., 2016)
CD44	mixed	miR-205,-34a	(Ouhtit et al., 2007) (Tse, 2005) (Zhang, Lu. et al., 2020), (Ahir, M. et al., 2020)
CDCP1	oncoprotein	miR-198	(Wright, H. J. et al 2017) (Hu, Y. et al., 2017)
CEACAM-1/CD66a	tumor suppressor	miR-342	(Weng, C. et al., 2016) (C. Yang et al., 2017)
Claudin-1	oncoprotein	miR-155	(Zhou, B. et al., 2015) (Chiang et al., 2019)
COL1A1	oncoprotein	miR-196b-5p	(Zhu, X. et al.,2008), (Jiang, Y. et al., 2022) (W. Wu & Zheng, 2022)
CXCR4	oncoprotein	miR-9,-139	(Liu, Y. et al., 2021), (Cheng, C.-W. et al., 2021) (J. Li et al., 2021)
DDR1	oncoprotein	miR-199b-p	(Wu, A. et al., 2018) (Baltes et al., 2020)
JAM-A	mixed	miR-495, -145	(Cao, M. et al., 2014), (Ye, D. et al., 2019), (Naik et al., 2018), (Murakami et al., 2011)
JAM-B	tumor suppressor	miR-374	(Li, W. et al., 2019) (Bhan et al., 2013)
L1CAM	oncoprotein	miR-21-3p	(Doberstein, K. et al., 2014)
MUC1	oncoprotein	miR-200c, -141, -1226,	(Rajabi, H. et al., 2013), (Gao, Y. et al., 2016), (Kufe et al., 2010),
Nectin-4	mixed	miR-520c-3p	(Liu, Y. et al., (2022)) (Zeindler et al., 2019), (Sethy et al., 2018)

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

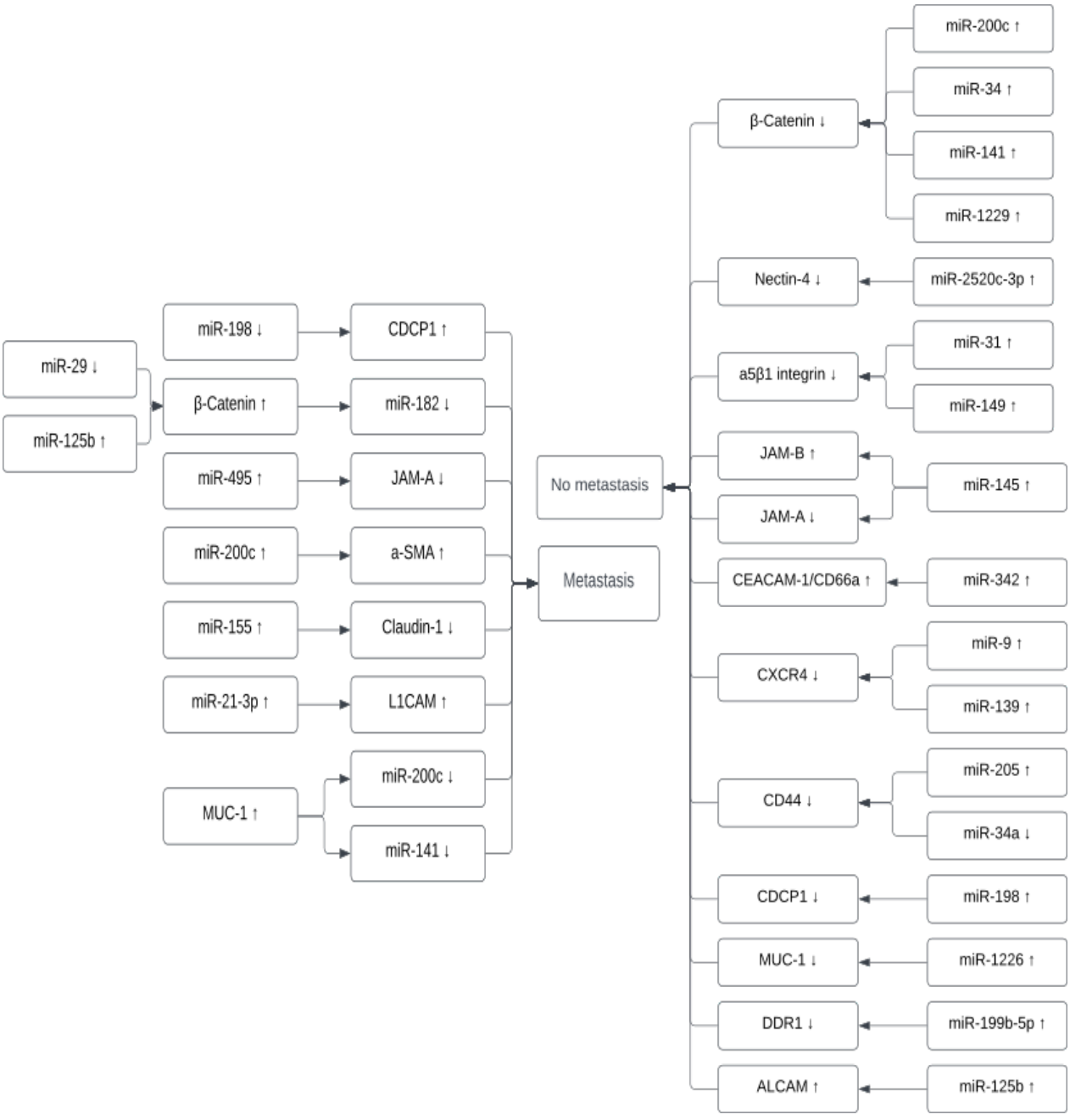


Figure 2. Description of relationship between miRNAs and epithelial markers in MBC

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

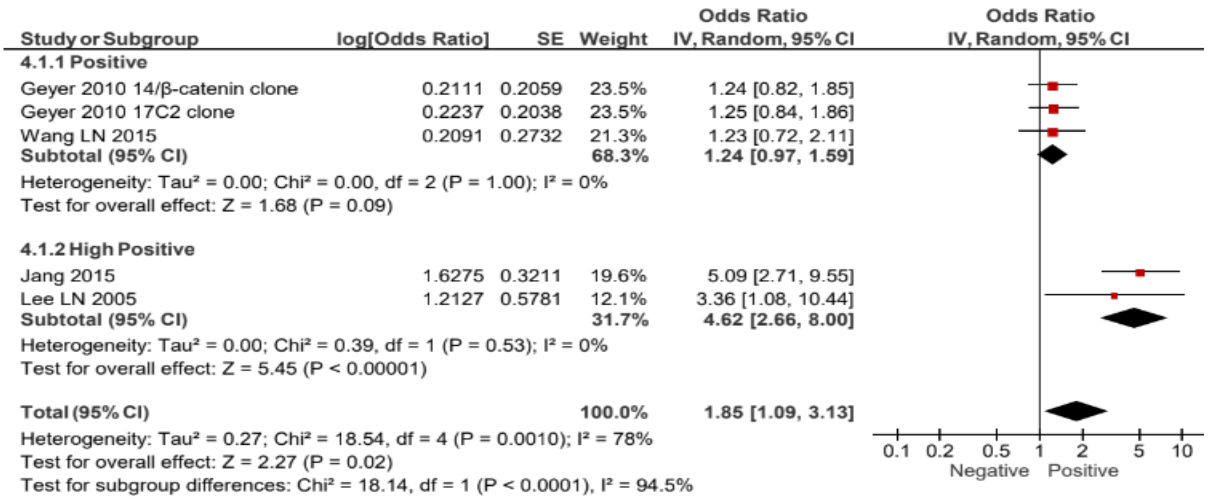


Figure 3. Meta-analysis on four studies assessing the association of β-catenin expression with the occurrence of MBC. The pooled OR was 1.85 (95% CI: 1.09-3.13; Z=2.27; P=0.02) with heterogeneity (I2 78% P=0.0010). Two studies showed mild positive correlation between B-catenin and the occurrence of MBC: Geyer (Geyer et al., 2011) and Wang (Z. Wang et al., 2015). The subtotal OR for the two studies was 1.24 (95 % Cl: 0.97-1.59; Z= 1.68(P= 0.09) without heterogeneity(I2 0% P=1). It should be noted that data from Geyer was used twice as he utilized two different types of B-catenin antibodies to prohibit metastasis. Two studies showed a high positive correlation between B-Catenin expression and occurrence of MBC: Jang (Jang et al., 2015) and Lee (Won-Lee, 2005). The subtotal OR was 4.62 (Cl% 95 2.66-8.00, Z= 2.27, P= 0.02) without heterogeneity (I2 0% P=0.53).

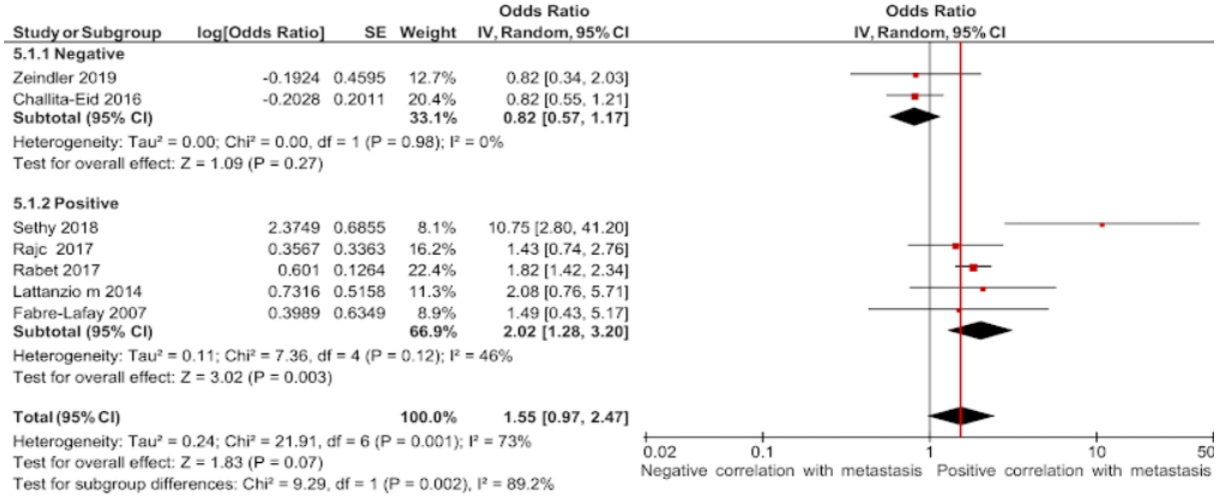


Figure 4. Meta-analysis on six studies assessing the association of Nectin-4 expression with the occurrence of MBC. The pooled OR was 1.55 (95% Cl: 0.97-2.47; Z= 1.83; P=0.07) with heterogeneity (I2 73% P= 0.002). Five studies showed positive correlation between Nectin-4 expression and the occurrence of MBC : Fabre-Lafay (Fabre-Lafay et al., 2007), Rabbit (M-Rabet et al., 2017), Lattanzio (Lattanzio et al., 2014), Sethy (Sethy et al., 2018), and Rajc (Rajc et al., 2017). The subtotal OR for the studies was 2.02 (95% Cl 1.28- 3.20 Z= 2.10; P= 0.04) with heterogeneity (I2=46%, P= 0.12). Two studies showed a negative correlation between Nectin-4 expression and the occurrence of MBC: Zeindler (Zeindler et al., 2019) and Chalita (Challita-Eid et al., 2016). The subtotal OR for those studies was 0.82 (95% Cl 0.82 Z= 1.09; P= 0.27) without heterogeneity.

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

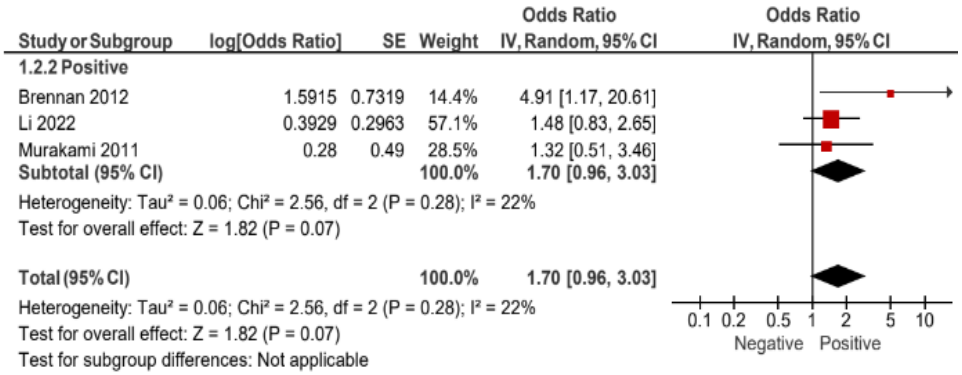


Figure 5. Meta-analysis on three studies assessing the association of JAM-A expression with the occurrence of MBC: Murakami (Murakami et al., 2011), ,Li (C.-H. Li et al., 2022), and Brennan (Brennan et al., 2013). The pooled OR was 1.70 (95% CI: 0.96-3.03; Z=1.82; P=0.28) with heterogeneity of 22% P=0.28. All three studies showed positive correlation between JAM-A and the occurrence of MBC.

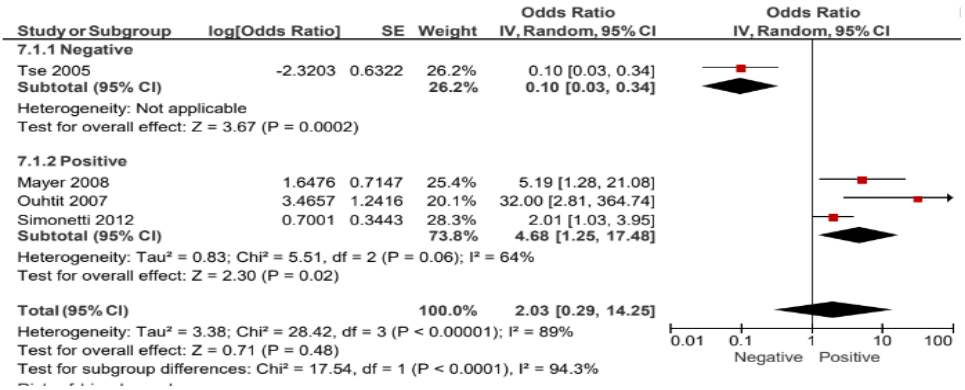


Figure 6. Meta-analysis on three studies assessing the association of CD44 expression with MBC. The pooled OR was 2.03 (95% Cl: 0.29 - 14.25; Z = 0.71; P=0.48) with heterogeneity (I2 89% P<0.00001). Three studies showed positive correlation between CD44 expression and the occurrence of MBC: Mayer (Mayer et al., 2008), Simonetti (Simonetti et al., 2012), and Ouhtit (Ouhtit et al., 2007). The subtotal OR for those studies was 4.68 (95% Cl 1.25 - 17.48; Z=2.30; P = 0.02) with heterogeneity 64%, P = 0.06. One study showed a negative correlation between CD44 and the occurrence of MBC: Tse (Tse, 2005). The subtotal OR for the study was 0.10 (95% Cl 0.10 Z= 3.67, P= 0.0002).

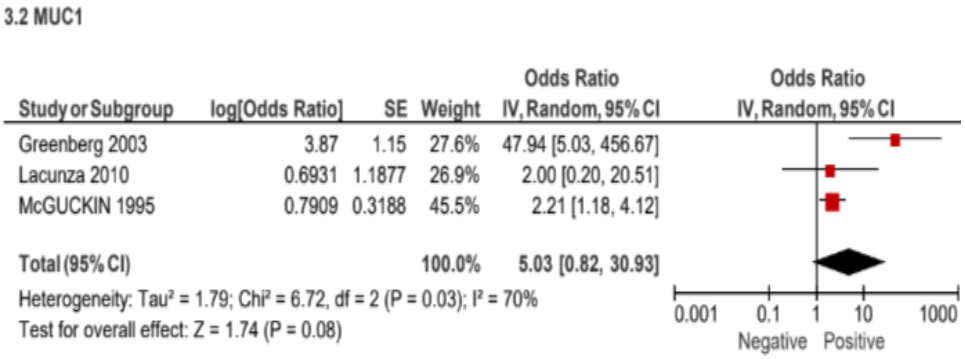


Figure 7. Meta-analysis on three studies assessing the association of MUC1 expression with MBC. The pooled OR was 5.03(95% Cl: 0.82 - 30.93; Z= 1.74; P= 0.08 with heterogeneity 70% P= 0.03. Three studies showed a positive correlation between MUC1 expression and the occurrence of MBC: McGuckin (Mcguckin, 1995), Greenberg (Greenberg et al., 2003), and Lacunza (Lacunza et al., 2010).

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

3 Discussion

Studies are unanimous in the notion that β-catenin expression is positively correlated with MBC and that β-catenin is expressed in the nucleus and/or cytoplasm of breast cancer cells. Lack of consensus occurs regarding how apparent the correlation between β-catenin and MBC is. As the diamond representing the subtotal OR ratio for mild positive correlation crosses the horizontal line representing the 95% interval, it's likely that Geyer and Wang do not present a statistically significant result (Fig 3). This would be in accordance with the results reported by Wang, where he stated that although high expression of B-catenin is correlated with poor patient outcome, no statistically significant correlations was noticed between B-catenin expression and metastasis (Z. Wang et al., 2015). Nevertheless, Geyer, who received similar results to Wang claimed that aberrant nuclear B-catenin expression was significantly associated with lymph node metastasis, which we failed to show in our statistical analysis, as the odds ratio for Geyer crosses the 95% interval line (Fig 3) (Geyer et al., 2011). Possible reasons include that the study only included enough data to develop an odds ratio in the context of lymph node metastasis, and not lymph vascular invasion. The lack of heterogeneity between the two studies can be attributed to both Geyer and Wang reporting their results in the context of lymph node metastasis. Factors that could have limited the results of Geyer's study include that Wang collected data in a short period of time (2006-2007), which may not be long enough to observe metastasis in patients, and that the patients were treated with anthracycline-based chemotherapy. Future longitudinal investigations could provide invaluable insights into the temporal dynamics of β-catenin expression and its implications for metastasis. The administration of anthracycline-based chemotherapy among patients underscores the need to rigorously examine the potential influence of such treatments on β-catenin expression and its subsequent association with MBC. For Lee and Jang, which indicate high positive correlation between B-catenin expression and MBC, the diamond representing the subtotal OR didn't cross the 95% interval, leading to the conclusion that the results are statistically significant (Fig 3). A potential limitation of both studies would have been the small sample sizes used. However, since the results are statistically significant, one could conclude that the small sample size doesn't significantly undermine the results. Even though the studies exhibit no heterogeneity, the subtypes of breast cancer used in the two studies are different (Jang - Sca-1 positive and Lee - ductal breast carcinoma) (Jang et al., 2015), (Won-Lee, 2005). Lack of heterogeneity despite this factor may suggest a lack of significant correlation between the role of B-catenin in MBC, and breast cancer subtype.Studying the influence of β-catenin expression on different breast cancer subtypes could clarify its role in metastatic potentials. Studies not included in the meta-analysis due to lack of numerical data necessary to do an odds ratio and the cellular mechanisms through which B-catenin affects cancer described in the introduction, also support the notion that B-Catenin overexpression in the nucleus and/or cytoplasm correlates with MBC (Quinn et al., 2021), (Lin et al., 2000), (De et al., 2016).

The role of Nectin-4 in MBC has been a controversial topic in research with some studies suggesting that overexpression of Nectin-4 is negatively correlated with MBC, and others proposing that it is positively correlated with MBC (Fig 4). The subtotal and individual odds ratios of the studies which suggest negative correlation between Nectin-4 expression and MBC, all cross the 95% confidence interval line, suggesting that the results are not statistically significant (Fig 4). Despite the lack of heterogeneity between the results of the two studies (Fig 4), they discuss MBC in the context of different breast cancer subtypes (Zeindler - TNBC and Chelita - ductal and lobular) and used different antibodies to locate Nectin-4 (Zeindler- AGS-22M6, ASG-22C and Chelita - M22-244b3). Since prior studies on Nectin-4 have demonstrated its sensitivity to different types of antibodies, the homogeneity between the two studies was unexpected (Lattanzio et al., 2014). A comprehensive study focusing on how various antibodies impact the detection and quantification of Nectin-4 expression would be valuable.The subtotal OR ratio of the studies representing positive correlation between Nectin-4 and MBC didn't cross the 95% confidence interval line, indicating that the results from all the studies portrayed a statistically significant positive correlation between Nectin-4 expression and MBC (Fig 4). As indicated by the crossing of the subtotal OR ratio of Rajc with the 95% confidence interval line, the study failed to report a statistically significant result (Fig. 5). Such a conclusion would be consistent with the results explicitly stated by Rajc that MBC in HER2 negative breast cancer and Nectin-4 expression are not significantly correlated with one another (Rajc et al., 2017). Even though Fabre-Lafay and Lattanzio both reported significant correlation between Nectin-4 expression and MBC in TNBC and luminal A breast cancer, respectively,

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

they both cross the 95% confidence interval line, indicating a statistically insignificant relationship (Fig 4)(Lattanzio et al., 2014), (Fabre-Lafay et al., 2007). Possible reasons for the difference between the results reported by them and those demonstrated by the statistical analysis include the limited amount of numerical data reported in Fabre-Lafey, and hence used in the statistical analysis, and the varied treatment the patients were subjected to in Lattanzio that could have additionally influenced metastasis. Rabet is a strong study, illustrated through its lack of cross with the 95% confidence interval line and high weight, due to its large amount of data (Fig 4). Sethy is also a strong study, with the only limitation being its heterogeneity from the other studies, likely resulting from its smaller dataset and assessment of ductal carcinomas, without specification of the molecular subtypes (Sethy et al., 2018). Except for Frabe Lafay who didn't specify the type of breast cancer carcinoma, all the other studies investigated Nectin-4 correlation in the context of a molecular subtype (Rabet - TNBC, Lattanzio and Rajc - luminal A). The inconsistency in data presentation suggests a need for standardized data collection and reporting methods to ensure comparability across studies.The study not included in the meta-analysis and the cellular processes outlined in the introduction also agree that Nectin-4 expression positively correlates with MBC (Shao et al., 2022).

The studies included in the meta-analysis all show JAM-A expression as having positive correlation with MBC. The pooled OR ratio crosses the 95% confidence interval leading to the conclusion that no statistically significant correlation between JAM-A expression and MBC can be observed (Fig 5). Murakami has a very low OR ratio meaning that the results reported aren't statistically significant, which is supported by inference made in the study (Murakami et al., 2011). Although Li reports similar results to Murakami, he states that JAM-A plays a role in several processes related to cell motility and is predominantly expressed in TNBC cells which are often associated with increased metastatic potential (Li et al., 2022). Since Li crosses the 95% confidence interval, the statistical analysis fails to reflect the reported results (Fig 5). Possible reasons include that in the study HER2 signaling and positive ER was perceived as a sign of metastasis, due to their causative relationship with TNBC. Nevertheless, to maintain homogeneity, the statistical analysis only considered the lymph node metastasis variable. A weakness of Li is the heterogeneity seen in the data pool with some patients being diagnosed in 1991. Brenan showcases a statistically significant positive correlation between JAM-A and MBC in Figure 5, supported by their own inferences in the study (Brennan et al., 2013). The heterogeneity of the data is in the acceptable range, as all the studies measure metastasis through lymph node involvement (Fig 5). Although different subtypes are used, limiting homogeneity, there are common subtypes used. For example, both Li and Brennan assess JAM-A expression in luminal A, luminal B, HER2 positive, and basal subtypes. Like Murakami, Li also assessed JAM-A expression in TNBC. The results are generally consistent with one another apart from Li who reported no correlation between JAM-A expression and MBC in basal breast cancer, whereas Brennan reported correlation between them. Brenan reported low expression of JAM-A in Luminal A breast cancer metastasis. However, Li united luminal A and B, and deduced positive correlation between JAM-A and metastasis in the subtype. More studies are needed to elucidate the role of JAM-A and metastasis in the context of luminal breast cancer subtypes. Overall, even though the three studies do not produce a statistically significant result, there are many more studies showcasing how JAM-A expression can lead to metastasis (Yang Wang and Lui, 2012), (McSherry et al., 2011) which weren't included due to lack of sufficient data to form OR ratios. There is also a study that shows how JAM-A expression can be negatively correlated with MBC (Naik et al., 2008) which we couldn't include due to the same reason. As there are explanations based on cellular mechanisms supporting both roles of JAM-A, inclusion of both types of studies would have yielded results with greater implications.

The role of CD44 has been disputed, with studies suggesting both its negative and positive correlation with MBC. The subtotal OR for the studies indicating positive correlation between MBC and CD44 doesn't cross the 95% confidence interval, meaning that the results are significant (Fig 6). However, the heterogeneity of the studies exceeds the expected range. Possible factors contributing to the heterogeneity of the results are the different ways through which MBC was measured through (Mayer - lymph node metastasis), (Ouhtit - metastasis to the liver), (Simonetti-number of invasive ductal/ micropapillary carcinomas) (Mayer et al., 2008) (Ouhtit et al., 2007) (Simonetti et al., 2012). The status of Ouhtit as an outlier can be attributed to them examining heterogeneity in vivo in mice, whereas the other two studies examined metastasis in patients. The heterogeneity among them suggests the sensitivity of CD44 to different tumor microenvironments and its context dependent nature, discussed in other studies, as well (Louderbough & Schroeder, 2011). To reduce heterogeneity, future studies could adopt a standardized method of

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

measuring MBC. The OR ratio of the study that reported negative correlation between CD44 and MBC, didn't cross the 95% confidence interval indicating a statistically significant result (Fig 6). A weakness of the study is the difference between the age and tumor size in the control group with no observed metastasis, and the ones with, indicating that the two variables could have affected metastasis in addition to CD44 expression. Future research should ensure that control groups are matched carefully based on factors like age, tumor size, and other relevant parameters. This would provide a more accurate assessment of CD44's role without potential confounding variables.Tse only tested for MBC in the context of standard CD44, whereas the other three studies also tested for MBC in the context of CD44 variants such as CD44v5 and CD44v6, thereby suggesting a potential role of variants in the dual nature of CD44 in MBC.Future research should ensure that control groups are matched carefully based on factors like age, tumor size, and other relevant parameters. This would provide a more accurate assessment of CD44's role without potential confounding variables. A deeper investigation into the different CD44 variants (e.g., CD44v5 and CD44v6) and their individual or combined roles in MBC could distinguish whether certain variants have more pronounced effects on MBC than the standard CD44.More studies portraying negative correlation between CD44 would have strengthened the results, however (Lopez et al., 2005) didn't include sufficient numerical data.

Studies are unanimous in the notion that MUC1 expression positively correlates with MBC. However, as indicated by the subtotal OR ratio crossing the 95% confidence interval, the studies do not provide enough data for a statistically significant correlation (Fig 7). Although there were a total of 10 studies discussing the role of MUC1 in MBC, all leading to the conclusion stated above, we were able to find sufficient data for an OR ratio only in 3. Not only are 3 studies insufficient, but their individual sample sizes were also very small, further contributing to the statistical insignificance of the results. The data has heterogeneity in the acceptable range. However, it was unexpected that McGuckin and Lacunza have greater similarity between their results than Greenberg and Lacunza, since both Lacunza and Greenberg measured MUC1 expression in vivo, whereas McGuckin measured it in vitro (Greenberg et al., 2003) (Lacunza et al., 2010).

The data encompassed in this study facilitated the contextual interpretation of both the meta-analysis outcomes and the identification of epithelial markers implicated in MBC, highlighting the need for marker comparisons. The meta-analysis findings not only suggest new research avenues to refine our understanding of the epithelial markers' role in MBC, but also offer a critical evaluation of current literature.

Acknowledgment

We would like to thank Dr. Dineva for her guidance and support, which have contributed significantly to the development and completion of this review.

References

Ahmad, R., et al. (2009). MUC1-C oncoprotein functions as a direct activator of the nuclear factor-κB p65 transcription factor. Cancer Research, 69(17), 7013–7021. https://doi.org/10.1158/0008-5472.can-09-0523

Ahir, M., et al. (2020). Delivery of dual MIRNA through CD44-targeted mesoporous silica nanoparticles for enhanced and effective triple-negative breast cancer therapy. Biomaterials Science, 8(10), 2939–2954. https://doi.org/10.1039/d0bm00015a

Akman, H. B., et al. (2015). ALCAM is indirectly modulated by miR-125b in MCF7 cells. Tumour Biology: The Journal of the International Society for Oncodevelopmental Biology and Medicine, 36(5), 3511–3520. https://doi.org/10.1007/s13277-014-2987-5

Al-Mrabt, N., et al. (2010). Abstract P6-08-11: The potential role of junctional adhesion molecule (JAM)-2 in breast cancer cells and the expression of JAM2 in ductal mammary carcinoma. Cancer Research, 70(24_Supplement), P6-08-11-P6-08–11. https://doi.org/10.1158/0008-5472.sabcs10-p6-08-11

Augoff, K., et al. (2011). miR-31 Is a Broad Regulator of β1-Integrin Expression and Function in Cancer Cells. Molecular Cancer Research, 9(11), 1500–1508. https://doi.org/10.1158/1541-7786.mcr-11-0311

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

Baltes, F., et al. (2020). Targeting Discoidin Domain Receptor 1 (DDR1) signaling and its crosstalk with β1-integrin emerges as a key factor for breast cancer chemosensitization upon collagen type 1 binding. International Journal of Molecular Sciences, 21(14), 4956. https://doi.org/10.3390/ijms21144956

Bhan, A., et al.. (2013). Antisense transcript long noncoding RNA (lncRNA) HOTAIR is transcriptionally induced by estradiol. Journal of Molecular Biology, 425(19), 3707–3722. https://doi.org/10.1016/j.jmb.2013.01.022

Bitler, B. G., Goverdhan, A., & Schroeder, J. A. (2010). MUC1 regulates nuclear localization and function of the epidermal growth factor receptor. Journal of Cell Science, 123(10), 1716–1723. https://doi.org/10.1242/jcs.062661

Brennan, K., et al. 2013). Junctional adhesion molecule-A is co-expressed with HER2 in breast tumors and acts as a novel regulator of HER2 protein degradation and signaling. Oncogene, 32(22), 2799–2804. https://doi.org/10.1038/onc.2012.276

Buechel, D., et al. (2021). Parsing β-catenin's cell adhesion and Wnt signaling functions in malignant mammary tumor progression. Proceedings of the National Academy of Sciences of the United States of America, 118(34). https://doi.org/10.1073/pnas.2020227118

Burandt, E., et al. (2014). Loss of ALCAM expression is linked to adverse phenotype and poor prognosis in breast cancer: A TMA-based immunohistochemical study on 2,197 breast cancer patients. Oncology Reports, 32(6), 2628–2634. https://doi.org/10.3892/or.2014.3523

Burkhardt, M. (2006). Cytoplasmic overexpression of ALCAM is prognostic of disease progression in breast cancer. Journal of Clinical Pathology, 59(4), 403–409. https://doi.org/10.1136/jcp.2005.028209

Cao, M., Nie, et al. (2014). MicroRNA-495 induces breast cancer cell migration by targeting JAM-A. Protein & Cell, 5(11), 862–872. https://doi.org/10.1007/s13238-014-0088-2

Challita-Eid, P. M., et al. (2016). Enfortumab vedotin antibody–drug conjugate targeting nectin-4 is a highly potent therapeutic agent in multiple preclinical cancer models. Cancer Research, 76(10), 3003–3013. https://doi.org/10.1158/0008-5472.can-15-1313

Chan, S., et al. (2014). MicroRNA-149 targets GIT1 to suppress integrin signaling and breast cancer metastasis. Oncogene, 33(36), 4496–4507. https://doi.org/10.1038/onc.2014.10

Chen, M.-J., et al. (2017). MiR-148a and miR-152 reduce tamoxifen resistance in ER+ breast cancer via downregulating ALCAM. Biochemical and Biophysical Research Communications, 483(2), 840–846. https://doi.org/10.1016/j.bbrc.2017.01.012

Chen, W., et al. (2021). MUC1: Structure, function, and clinic application in epithelial cancers. International Journal of Molecular Sciences, 22(12), 6567. https://doi.org/10.3390/ijms22126567

Cheng, C. W., et al. (2021). Mir-139 modulates cancer stem cell function of human breast cancer through targeting CXCR4. Cancers, 13(11), 2582. https://doi.org/10.3390/cancers13112582

Chiang, C. H., Hou, M. F., & Hung, W. C. (2013). Up-regulation of miR-182 by β-catenin in breast cancer increases tumorigenicity and invasiveness by targeting the matrix metalloproteinase inhibitor RECK. Biochimica et Biophysica Acta. General Subjects, 1830(4), 3067–3076. https://doi.org/10.1016/j.bbagen.2013.01.009

Chiang, SK., et al. (2019). DOCK1 regulates growth and motility through the RRP1B-claudin-1 pathway in claudin-low breast cancer cells. Cancers, 11(11), 1762. https://doi.org/10.3390/cancers11111762

Cruz, R. G. B., et al.. (2022). A transcriptional link between HER2, JAM-A and FOXA1 in breast cancer. Cells (Basel, Switzerland), 11(4), 735. https://doi.org/10.3390/cells11040735

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

Damonte, P., et al. (2007). EMT tumorigenesis in the mouse mammary gland. Laboratory Investigation; a Journal of Technical Methods and Pathology, 87(12), 1218–1226. https://doi.org/10.1038/labinvest.3700683

Davies, S. R., et al. (2008). Expression of the cell to cell adhesion molecule, ALCAM, in breast cancer patients and the potential link with skeletal metastasis. Oncology Reports, 19(2), 555–561. https://doi.org/10.3892/or.19.2.555

De, P., et al. (2016). Wnt-beta-catenin pathway signals metastasis-associated tumor cell phenotypes in triple negative breast cancers. Oncotarget, 7(28), 43124–43149. https://doi.org/10.18632/oncotarget.8988
Doberstein, K., et al. (2014). miR-21-3p is a positive regulator of L1CAM in several human carcinomas. Cancer Letters, 354(2), 455–466. https://doi.org/10.1016/j.canlet.2014.08.020

Doberstein, K., et al. (2014). L1CAM is expressed in triple-negative breast cancers and is inversely correlated with Androgen receptor. BMC Cancer, 14(1). https://doi.org/10.1186/1471-2407-14-958

Epithelial Cell Markers and Intracellular Molecules. (n.d.). Www.rndsystems.com. Retrieved August 1, 2023, from https://www.rndsystems.com/research-area/epithelial-cell-markers-and-intracellular-molecules

Fabre-Lafay, S., et al. (2007). Nectin-4 is a new histological and serological tumor associated marker for breast cancer. BMC Cancer, 7(1). https://doi.org/10.1186/1471-2407-7-73

Felding-Habermann, B., et al. (2001). Integrin activation controls metastasis in human breast cancer. Proceedings of the National Academy of Sciences of the United States of America, 98(4), 1853–1858. https://doi.org/10.1073/pnas.98.4.1853

Garcia-Recio, S., et al. (2022). Multiomics in primary and metastatic breast tumors from the AURORA US network finds microenvironment and epigenetic drivers of metastasis. Nature Cancer. https://doi.org/10.1038/s43018-022-00491-x
Gao, Y., et al. (2016). The roles of microrna-141 in human cancers: From diagnosis to treatment. Cellular Physiology and Biochemistry, 38(2), 427–448. https://doi.org/10.1159/000438641

Geyer, F. C., et al. (2011). β-Catenin pathway activation in breast cancer is associated with triple-negative phenotype but not with CTNNB1 mutation. Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc, 24(2), 209–231. https://doi.org/10.1038/modpathol.2010.205

Ghatak, S., Misra, S., & Toole, B. P. (2002). Hyaluronan oligosaccharides inhibit anchorage-independent growth of tumor cells by suppressing the phosphoinositide 3-kinase/akt cell survival pathway. The Journal of Biological Chemistry, 277(41), 38013–38020. https://doi.org/10.1074/jbc.m202404200

Godden, A. M., et al. (2022). An efficient miRNA knockout approach using CRISPR-Cas9 in Xenopus. Developmental Biology, 483, 66–75. https://doi.org/10.1016/j.ydbio.2021.12.015

Greenberg, R., et al. (2003). Detection of hepatocyte growth factor/scatter factor receptor (c-Met) in axillary drainage after operations for breast cancer using reverse transcriptase–polymerase chain reaction. Breast Cancer Research: BCR, 5(3). https://doi.org/10.1186/bcr588

Guo, F., et al. (2016). CXCL12/CXCR4: a symbiotic bridge linking cancer cells and their stromal neighbors in oncogenic communication networks. Oncogene, 35(7), 816–826. https://doi.org/10.1038/onc.2015.139

Hagemeister, F. B., et al. (1980). Causes of death in breast cancer a clinicopathologic study. Cancer, 46(1), 162–167. https://doi.org/10.1002/1097-0142(19800701)46:1<162::aid-cncr2820460127>3.0.co;2-b

Hattrup, C. L., & Gendler, S. J. (2006). MUC1 alters oncogenic events and transcription in human breast cancer cells. Breast Cancer Research: BCR, 8(4). https://doi.org/10.1186/bcr1515

Hayashi, T., et al.. (2001). MUC1 mucin core protein binds to the domain 1 of ICAM-1. Digestion, 63(Suppl. 1), 87–92. https://doi.org/10.1159/000051917

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

Herrlich, P., et al. (2006). CD44 acts both as a growth- and invasiveness-promoting molecule and as a tumor-suppressing cofactor. Annals of the New York Academy of Sciences, 910(1), 106–120. https://doi.org/10.1111/j.1749-6632.2000.tb06704.x

Hou, J., et al. (2020). The roles of integrin α5β1 in human cancer. OncoTargets and Therapy, 13, 13329–13344. https://doi.org/10.2147/ott.s273803

Hu, Y., et al. (2017). miR-198 functions as a tumor suppressor in breast cancer by targeting CUB domain-containing protein 1. Oncology Letters, 13(3), 1753–1760. https://doi.org/10.3892/ol.2017.5673
Ihnen, M., et al. (2010). Expression levels of Activated Leukocyte Cell Adhesion Molecule (ALCAM/CD166) in primary breast carcinoma and distant breast cancer metastases. Disease Markers, 28(2), 71–78. https://doi.org/10.3233/DMA-2010-0685

Jang, G.-B., et al. (2015). Blockade of Wnt/β-catenin signaling suppresses breast cancer metastasis by inhibiting CSC-like phenotype. Scientific Reports, 5(1). https://doi.org/10.1038/srep12465

Jassem, J., et al. (2011). The neuronal cell adhesion molecule L1CAM as a therapeutic targe in breast cancer. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 29(15_suppl), e21138–e21138. https://doi.org/10.1200/jco.2011.29.15_suppl.e21138

Jiang, Y., et al. (2022). Collagen fiber features and COL1A1: are they associated with elastic parameters in breast lesions, and can COL1A1 predict axillary lymph node metastasis? BMC Cancer, 22(1). https://doi.org/10.1186/s12885-022-10092-7

Jing, X., et al. (2018). Overexpression of MUC1 predicts poor prognosis in patients with breast cancer. Oncology Reports. https://doi.org/10.3892/or.2018.6887

Kar, S., et al. (2020). Wnt/β-catenin signaling pathway regulates osteogenesis for breast cancer bone metastasis: Experiments in an in vitro nanoclay scaffold cancer testbed. ACS Biomaterials Science & Engineering, 6(5), 2600–2611. https://doi.org/10.1021/acsbiomaterials.9b00923

Karpathiou, G., et al. (2021). Immunohistochemical analysis of L1 cell adhesion molecule and high endothelial venules in breast cancer brain metastasis. Pathology, Research and Practice, 223(153484), 153484. https://doi.org/10.1016/j.prp.2021.153484

Khodabakhsh, F., et al. (2021). Crosstalk between MUC1 and VEGF in angiogenesis and metastasis: a review highlighting roles of the MUC1 with an emphasis on metastatic and angiogenic signaling. Cancer Cell International, 21(1). https://doi.org/10.1186/s12935-021-01899-8

Kim, M. J., et al. (2020). Novel antibodies targeting MUC1-C showed anti-metastasis and growth-inhibitory effects on human breast cancer cells. International Journal of Molecular Sciences, 21(9), 3258. https://doi.org/10.3390/ijms21093258

King, J. A., et al. (2004). Activated leukocyte cell adhesion molecule in breast cancer: prognostic indicator. Breast Cancer Research: BCR, 6(5). https://doi.org/10.1186/bcr815

Komiya, Y., & Habas, R. (2008). Wnt signal transduction pathways. Organogenesis, 4(2), 68–75. https://doi.org/10.4161/org.4.2.5851

Kufe. (2010). miR-1226 targets expression of the mucin 1 oncoprotein and induces cell death. International Journal of Oncology, 37(1). https://doi.org/10.3892/ijo_00000653

Kwon, J. J., et al. (2019). A systematic review of Mir-29 in cancer. Molecular Therapy - Oncolytics, 12, 173–194. https://doi.org/10.1016/j.omto.2018.12.011

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

Lacunza, E., et al. (2010). MUC1 oncogene amplification correlates with protein overexpression in invasive breast carcinoma cells. Cancer Genetics and Cytogenetics, 201(2), 102–110. https://doi.org/10.1016/j.cancergencyto.2010.05.015

Lattanzio, R., et al. (2014). Membranous Nectin-4 expression is a risk factor for distant relapse of T1-T2, N0 luminal-A early breast cancer. Oncogenesis, 3(9), e118–e118. https://doi.org/10.1038/oncsis.2014.32

Lavrsen, K., et al. (2013). Aberrantly glycosylated MUC1 is expressed on the surface of breast cancer cells and a target for antibody-dependent cell-mediated cytotoxicity. Glycoconjugate Journal, 30(3), 227–236. https://doi.org/10.1007/s10719-012-9437-7

Li, C.-H., et al. (2022). The activation of EP300 by F11R leads to EMT and acts as a prognostic factor in triple-negative breast cancers. In Research Square. https://doi.org/10.21203/rs.3.rs-1454643/v1
Li, J., et al. (2021). Vitamin D regulates CXCL12/CXCR4 and epithelial-to-mesenchymal transition in a model of breast cancer metastasis to lung. Endocrinology, 162(7). https://doi.org/10.1210/endocr/bqab049

Li, W., et al. (2022). Novel insights into the roles and therapeutic implications of MUC1 oncoprotein via regulating proteins and non-coding RNAs in cancer. Theranostics, 12(3), 999–1011. https://doi.org/10.7150/thno.63654

Li, W., et al. (2021). MicroRNA-34a: Potent tumor suppressor, cancer stem cell inhibitor, and potential anticancer therapeutic. Frontiers in Cell and Developmental Biology, 9. https://doi.org/10.3389/fcell.2021.640587

Li, Y., & Galileo, D. S. (2010). Soluble L1CAM promotes breast cancer cell adhesion and migration in vitro, but not invasion. Cancer Cell International, 10(1), 34. https://doi.org/10.1186/1475-2867-10-34

Lin, S.-Y., et al. (2000). β-Catenin, a novel prognostic marker for breast cancer: Its roles in cyclin D1 expression and cancer progression. Proceedings of the National Academy of Sciences of the United States of America, 97(8), 4262–4266. https://doi.org/10.1073/pnas.060025397

Liu, Q., et al. (2017). Factors involved in cancer metastasis: a better understanding to “seed and soil” hypothesis. Molecular Cancer, 16(1). https://doi.org/10.1186/s12943-017-0742-4

Liu, B., et al. (2018). Mir-200c/141 regulates breast cancer stem cell heterogeneity via targeting HIPK1/β-catenin axis. Theranostics, 8(21), 5801–5813. https://doi.org/10.7150/thno.29380

Liu, Y., Zhao, Q., Xi, T., Zheng, L., & Li, X. (2021). MicroRNA-9 as a paradoxical but critical regulator of cancer metastasis: Implications in personalized medicine. Genes & Diseases, 8(6), 759–768. https://doi.org/10.1016/j.gendis.2020.10.005

Liu, Y., et al. (2022). Nectin-4 promotes osteosarcoma progression and metastasis through activating PI3K/AKT/NF-κB signaling by down-regulation of miR-520c-3p. Cancer Cell International, 22(1). https://doi.org/10.1186/s12935-022-02669-w

Lopez, J. I., et al. (2005). CD44 attenuates Metastatic invasion during breast cancer progression. Cancer Research, 65(15), 6755–6763. https://doi.org/10.1158/0008-5472.can-05-0863

Louderbough, J. M. V., & Schroeder, J. A. (2011). Understanding the dual nature of CD44 in breast cancer progression. Molecular Cancer Research: MCR, 9(12), 1573–1586. https://doi.org/10.1158/1541-7786.mcr-11-0156

Lu, L., et al. (2015). Circulating tumor cell clusters-associated gene plakoglobin and breast cancer survival. Breast Cancer Research and Treatment, 151(3), 491–500. https://doi.org/10.1007/s10549-015-3416-1

Machado-Pineda, Y., et al. (2018). CD9 controls integrin α5β1-mediated cell adhesion by modulating its association with the metalloproteinase ADAM17. Frontiers in Immunology, 9. https://doi.org/10.3389/fimmu.2018.02474

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

Mandai, K., et al. (2015). Nectins and nectin-like molecules in development and disease. In Current Topics in Developmental Biology (pp. 197–231). Elsevier.

Mayer, S., et al. (2008). Increased soluble CD44 concentrations are associated with larger tumor size and lymph node metastasis in breast cancer patients. Journal of Cancer Research and Clinical Oncology, 134(11), 1229–1235. https://doi.org/10.1007/s00432-008-0397-z

McFarlane, S., et al. (2015). CD44 increases the efficiency of distant metastasis of breast cancer. Oncotarget, 6(13), 11465–11476. https://doi.org/10.18632/oncotarget.3410

Mcguckin, M. (1995). Prognostic significance of muc1 epithelial mucin expression in breast cancer*1. Human Pathology, 26(4), 432–439. https://doi.org/10.1016/0046-8177(95)90146-9

McSherry, E. A., et al. (2011). Breast cancer cell migration is regulated through junctional adhesion molecule-A-mediated activation of Rap1 GTPase. Breast Cancer Research: BCR, 13(2). https://doi.org/10.1186/bcr2853

Mierke, C. T., et al. (2011). Integrin α5β1 facilitates cancer cell invasion through enhanced contractile forces. Journal of Cell Science, 124(3), 369–383. https://doi.org/10.1242/jcs.071985

Milella, M., et al.. (2015). PTEN: Multiple Functions in Human Malignant Tumors. Frontiers in Oncology, 5. https://doi.org/10.3389/fonc.2015.00024

Moisini, I., et al. (2021). L1CAM expression in recurrent estrogen positive/HER2 negative breast cancer: A novel biomarker worth considering. Applied Immunohistochemistry & Molecular Morphology, 29(4), 287–292. https://doi.org/10.1097/pai.0000000000000909

Morozevich, G. E., et al. (2011). Implication of integrin α5β1 in human breast carcinoma apoptosis and drug resistance. Biomeditsinskaia khimiia, 57(1), 77–84. https://doi.org/10.18097/pbmc20115701077

M-Rabet, M., et al. (2017). Nectin-4: a new prognostic biomarker for efficient therapeutic targeting of primary and metastatic triple-negative breast cancer. Annals of Oncology, 28(4), 769–776. https://doi.org/10.1093/annonc/mdw678

Murakami, M., et al. (2011). Abrogation of junctional adhesion molecule-A expression induces cell apoptosis and reduces breast cancer progression. PloS One, 6(6), e21242. https://doi.org/10.1371/journal.pone.0021242

Naik, M. U., et al. (2008). Attenuation of junctional adhesion molecule-A is a contributing factor for breast cancer cell invasion. Cancer Research, 68(7), 2194–2203. https://doi.org/10.1158/0008-5472.can-07-3057

Nava, P., et al. (2011). JAM-A regulates epithelial proliferation through Akt/β-catenin signalling. EMBO Reports, 12(4), 314–320. https://doi.org/10.1038/embor.2011.16

Nie, J., et al. (2019). MiR-125b regulates the proliferation and metastasis of triple negative breast cancer cells via the Wnt/β-catenin pathway and EMT. Bioscience, Biotechnology, and Biochemistry, 83(6), 1062–1071. https://doi.org/10.1080/09168451.2019.1584521

O'Bryan, S., et al. (2017). The roles of oncogenic miRNAs and their therapeutic importance in breast cancer. European Journal of Cancer (Oxford, England: 1990), 72, 1–11. https://doi.org/10.1016/j.ejca.2016.11.004

Ouhtit, A., et al. (2007). In vivo evidence for the role of CD44s in promoting breast cancer metastasis to the liver. The American Journal of Pathology, 171(6), 2033–2039. https://doi.org/10.2353/ajpath.2007.070535

Peng, Y., et al. (2020). High expression of JAM2 indicates better prognosis and immunotherapy response in breast cancer. In bioRxiv. https://doi.org/10.1101/2020.12.11.421081

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

Piao, D., Jiang, T., Liu, G., Wang, B., Xu, J., & Zhu, A. (2012). Clinical implications of activated leukocyte cell adhesion molecule expression in breast cancer. Molecular Biology Reports, 39(1), 661–668. https://doi.org/10.1007/s11033-011-0783-5

Qiao, X., et al. (2021). Association of human breast cancer CD44-/CD24- cells with delayed distant metastasis. ELife, 10. https://doi.org/10.7554/elife.65418

Quinn, H. M., et al. (2021). YAP and β-catenin cooperate to drive oncogenesis in basal breast cancer. Cancer Research, 81(8), 2116–2127. https://doi.org/10.1158/0008-5472.can-20-2801

Rahn, J. J., et al. (2005). MUC1 mediates transendothelial migration in vitro by ligating endothelial cell ICAM-1. Clinical & Experimental Metastasis, 22(6), 475–483. https://doi.org/10.1007/s10585-005-3098-x

Rajabi, H., et al. (2013). Muc1-C oncoprotein activates the zeb1/mir-200c regulatory loop and epithelial–mesenchymal transition. Oncogene, 33(13), 1680–1689. https://doi.org/10.1038/onc.2013.114

Rajc, J., et al. (2017). Prognostic role of Nectin-4 expression in luminal B (HER2 negative) breast cancer. Pathology, Research and Practice, 213(9), 1102–1108. https://doi.org/10.1016/j.prp.2017.07.019

Ramani, P., Dungwa, J. V., & May, M. T. (2012). LYVE-1 upregulation and lymphatic invasion correlate with adverse prognostic factors and lymph node metastasis in neuroblastoma. Virchows Archiv: An International Journal of Pathology, 460(2), 183–191. https://doi.org/10.1007/s00428-011-1190-y

Sánchez-Cid, et al. (2017). MicroRNA-200, associated with metastatic breast cancer, promotes traits of mammary luminal progenitor cells. Oncotarget, 8(48), 83384–83406. https://doi.org/10.18632/oncotarget.20698

Schroeder, J. A., et al. (2003). MUC1 alters β-catenin-dependent tumor formation and promotes cellular invasion. Oncogene, 22(9), 1324–1332. https://doi.org/10.1038/sj.onc.1206291
Sethy, C., et al. (2021). Nectin-4 promotes lymphangiogenesis and lymphatic metastasis in breast cancer by regulating CXCR4-LYVE-1 axis. Vascular Pharmacology, 140(106865), 106865. https://doi.org/10.1016/j.vph.2021.106865

Sethy, C., et al. (2018). Clinical significance of Nectin-4 expression in metastasis and angiogenesis for tumor relapse. In Preprints. https://doi.org/10.20944/preprints201810.0154.v1

Severson, E. A., & Parkos, C. A. (2009). Structural determinants of Junctional Adhesion Molecule A (JAM-A) function and mechanisms of intracellular signaling. Current Opinion in Cell Biology, 21(5), 701–707. https://doi.org/10.1016/j.ceb.2009.06.005

Shang, S., Hua, F., & Hu, Z.-W. (2017). The regulation of β-catenin activity and function in cancer: therapeutic opportunities. Oncotarget, 8(20), 33972–33989. https://doi.org/10.18632/oncotarget.15687

Shao, F., et al. (2022). Nectin-4-targeted immunoSPECT/CT imaging and photothermal therapy of triple-negative breast cancer. Journal of Nanobiotechnology, 20(1). https://doi.org/10.1186/s12951-022-01444-3

Shen, Q., et al. (2008). MUC1 initiates Src-CrkL-Rac1/Cdc42–mediated actin cytoskeletal protrusive motility after ligating intercellular adhesion molecule-1. Molecular Cancer Research: MCR, 6(4), 555–567. https://doi.org/10.1158/1541-7786.mcr-07-2033

Siddharth, S., et al. (2017). Nectin-4 is a breast cancer stem cell marker that induces WNT/β-catenin signaling via Pi3k/Akt axis. The International Journal of Biochemistry & Cell Biology, 89, 85–94. https://doi.org/10.1016/j.biocel.2017.06.007

Siddharth, S., et al. (2018). The soluble nectin-4 ecto-domain promotes breast cancer induced angiogenesis via endothelial Integrin-β4. The International Journal of Biochemistry & Cell Biology, 102, 151–160. https://doi.org/10.1016/j.biocel.2018.07.011

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

Simonetti, S., et al. (2012). Immunophenotyping analysis in invasive micropapillary carcinoma of the breast: Role of CD24 and CD44 isoforms expression. Breast (Edinburgh, Scotland), 21(2), 165–170. https://doi.org/10.1016/j.breast.2011.09.004

Si, W., et al. (2016). Mir-34a inhibits breast cancer proliferation and progression by targeting WNT1 in Wnt/β-catenin signaling pathway. The American Journal of the Medical Sciences, 352(2), 191–199. https://doi.org/10.1016/j.amjms.2016.05.002

Sun, X., et al. (2020). Exploring the metabolic vulnerabilities of epithelial–mesenchymal transition in breast cancer. Frontiers in Cell and Developmental Biology, 8. https://doi.org/10.3389/fcell.2020.00655

Tan, Z., et al. (2016). MicroRNA-1229 overexpression promotes cell proliferation and tumorigenicity and activates Wnt/β-catenin signaling in breast cancer. Oncotarget, 7(17), 24076–24087. https://doi.org/10.18632/oncotarget.8119

Tang, X., et al. (2015). Stromal mir-200s contribute to breast cancer cell invasion through CAF activation and ECM Remodeling. Cell Death & Differentiation, 23(1), 132–145. https://doi.org/10.1038/cdd.2015.78

Thapa, R., & Wilson, G. D. (2016). The importance of CD44 as a stem cell biomarker and therapeutic target in cancer. Stem Cells International, 2016, 1–15. https://doi.org/10.1155/2016/2087204

Tse, G. M. K. (2005). CD44s is useful in the differentiation of benign and malignant papillary lesions of the breast. Journal of Clinical Pathology, 58(11), 1185–1188. https://doi.org/10.1136/jcp.2005.026906

Tyler, M., & Tirosh, I. (2021). Decoupling epithelial-mesenchymal transitions from stromal profiles by integrative expression analysis. Nature Communications, 12(1). https://doi.org/10.1038/s41467-021-22800-1

Vadhan, A., et al.. (2022). CD44 promotes breast cancer metastasis through AKT-mediated downregulation of nuclear FOXA2. Biomedicines, 10(10), 2488. https://doi.org/10.3390/biomedicines10102488

Wang, Yanfang, et al. (2011). Integrin subunits alpha5 and alpha6 regulate cell cycle by modulating the chk1 and Rb/E2F pathways to affect breast cancer metastasis. Molecular Cancer, 10(1). https://doi.org/10.1186/1476-4598-10-84

Wang, Yang, & Lui, W.-Y. (2012). Transforming growth factor-β1 attenuates junctional adhesion molecule-A and contributes to breast cancer cell invasion. European Journal of Cancer (Oxford, England: 1990), 48(18), 3475–3487. https://doi.org/10.1016/j.ejca.2012.04.016

Wang, Yujun, Yu, L., & Wang, T. (2018). MicroRNA-374b inhibits the tumor growth and promotes apoptosis in non-small cell lung cancer tissue through the p38/ERK signaling pathway by targeting JAM-2. Journal of Thoracic Disease, 10(9), 5489–5498. https://doi.org/10.21037/jtd.2018.09.93

Wang, Z., et al. (2015). Clinical implications of β-catenin protein expression in breast cancer. International Journal of Clinical and Experimental Pathology, 8(11), 14989–14994.

Weng, C., Nguyen, T., & Shively, J. E. (2016). MiRNA-342 regulates CEACAM1-induced lumen formation in a three-dimensional model of mammary gland morphogenesis. The Journal of Biological Chemistry, 291(32), 16777–16786. https://doi.org/10.1074/jbc.m115.710152

Williams, K., et al. (2013). CD44 integrates signaling in normal stem cell, cancer stem cell and (pre)metastatic niches. Experimental Biology and Medicine (Maywood, N.J.), 238(3), 324–338. https://doi.org/10.1177/1535370213480714

Won-Lee, L. E. E. (2005). Prognostic Significance of Abnormal beta - catenin Expression in Breast Carcinoma. Korean Journal of Pathology, 114–119. https://pesquisa.bvsalud.org/portal/resource/pt/wpr-147993

Vol. 2023 (2) 353 – 371
ISSN 2688-3651 [online]

Wright, H. J., Hou, J., Xu, B., Cortez, M., Potma, E. O., Tromberg, B. J., & Razorenova, O. V. (2017). CDCP1 drives triple-negative breast cancer metastasis through reduction of lipid-droplet abundance and stimulation of fatty acid oxidation. Proceedings of the National Academy of Sciences of the United States of America, 114(32). https://doi.org/10.1073/pnas.1703791114

Wu, A., et al. (2018). miR-199b-5p inhibits triple negative breast cancer cell proliferation, migration and invasion by targeting DDR1. Oncology Letters. https://doi.org/10.3892/ol.2018.9255

Wu, W., & Zheng, L. (2022). Comprehensive analysis identifies COL1A1, COL3A1, and POSTN as key genes associated with brain metastasis in patients with breast cancer. Evidence-Based Complementary and Alternative Medicine: ECAM, 2022, 1–7. https://doi.org/10.1155/2022/7812218

Yamashita, M., et al. (2012). Role of stromal myofibroblasts in invasive breast cancer: stromal expression of alpha-smooth muscle actin correlates with worse clinical outcome. Breast Cancer (Tokyo, Japan), 19(2), 170–176. https://doi.org/10.1007/s12282-010-0234-5

Yang, C., et al. (2017). Inhibition of cell invasion and migration by CEACAM1-4S in breast cancer. Oncology Letters, 14(4), 4758–4766. https://doi.org/10.3892/ol.2017.6791

Yang, F., et al. (2019). Inhibition of Dipeptidyl peptidase-4 accelerates epithelial–mesenchymal transition and breast cancer metastasis via the CXCL12/CXCR4/mTOR axis. Cancer Research, 79(4), 735–746. https://doi.org/10.1158/0008-5472.can-18-0620

Yang, J., et al. … On behalf of the EMT International Association (TEMTIA). (2020). Guidelines and definitions for research on epithelial–mesenchymal transition. Nature Reviews. Molecular Cell Biology, 21(6), 341–352. https://doi.org/10.1038/s41580-020-0237-9

Yao, Q., et al. (2015). CXCR4 in breast cancer: oncogenic role and therapeutic targeting. Drug Design, Development and Therapy, 4953. https://doi.org/10.2147/dddt.s84932

Ye, D., Shen, Z., & Zhou, S. (2019). function of microrna-145 and mechanisms underlying its role in malignant tumor diagnosis and treatment. Cancer Management and Research, Volume 11, 969–979. https://doi.org/10.2147/cmar.s191696

Yi-Lei, Z., et al. (2017). Roles of Rap1 signaling in tumor cell migration and invasion. Cancer Biology & Medicine, 14(1), 90–99. https://doi.org/10.20892/j.issn.2095-3941.2016.0086

Zeindler, J., et al. (2019). Nectin-4 expression is an independent prognostic biomarker and associated with better survival in triple-negative breast cancer. Frontiers in Medicine, 6. https://doi.org/10.3389/fmed.2019.00200

Zeng, X., et al. (2008). Initiation of Wnt signaling: control of Wnt coreceptor Lrp6 phosphorylation/activation via frizzled, dishevelled and axin functions. Development (Cambridge, England), 135(2), 367–375. https://doi.org/10.1242/dev.013540

Zhang, Lu, et al. (2020). miR-205/RunX2 axis negatively regulates CD44+/CD24- breast cancer stem cell activity. American Journal of Cancer Research, 10(6), 1871–1887.

Zhou, B., et al. (2015). Claudin 1 in breast cancer: New insights. Journal of Clinical Medicine, 4(12), 1960–1976. https://doi.org/10.3390/jcm4121952

Zhu, X., et al. (2018). Downregulation of MiR-196b-5p impedes cell proliferation and metastasis in breast cancer through regulating COL1A1. American Journal of Translational Research, 10(10), 3122–3132.