Autism Spectrum Disorder: A multifactorial analysis of genetics, early diagnosis and treatment to achieve the optimal outcome

Jayashree Balaraman^{1*}

¹ East Brunswick High School, East Brunswick, NJ USA

Received May 26, 2021; Revised November 14, 2021; Accepted, November 24, 2021

Abstract

Autism Spectrum Disorder (ASD) is a neurodevelopmental disease that impacts 1 in 54 children in the United States. Generally, the associated symptoms of ASD include social and communication impairment, intellectual disability, and repeated rigid movements. ASD encompasses many different characteristics and different standards of diagnosis. It is complicated because it is characterized as a spectrum disorder - it could range from very severe cases to mild cases of this disorder. Along with Attention Deficit Hyperactivity Disorder, ASD is ranked as the most common neurodevelopmental disorder. What complicates this disorder even further is that there are over 100 genes associated with ASD, although these genes differ for each patient. Moreover, because of genetic variability, there is a lack of standardized diagnosing protocol for ASD. The life-long symptoms of ASD makes those with this disorder suffer greatly because in most cases the symptoms do not get reduced and ASD patients do not "grow out of it". However, there is a primary goal for those with ASD to have reduced symptoms or no symptoms at all which is known as optimal outcome - the ultimate goal for those with ASD. This phenomenon is directly linked to earlier intervention, because of the greater plasticity of the brain at a younger age, which is linked to diagnosis. There are two primary routes of diagnosis - quantitative and qualitative measures. The Social Response Scale (SRS) is a commonly used and validated quantitative test to diagnose ASD. The qualitative measure is from the perspective of a parent or a physician where they assess the symptoms associated with ASD - lack of communication skills, lower intellectual ability, and repeated rigid movements. Although the genetic component of ASD is very complex, finding a genetic mechanism would make early diagnosis easier and the subsequent initiation of early specific intervention could lead to optimal outcome for ASD patients.

Keywords: Autism Spectrum Disorder, Attention Deficit Hyperactivity Disorder, Optimal Outcome, Social Response Scale, Quantitative, Qualitative, Early Intervention, Genetics, Diagnosis, Plasticity

1. Introduction

Autism Spectrum Disorder, a neurodevelopmental disease consisting of a group of complex neurodevelopmental disorders that are represented by social impairment and repetitive behaviors. Though the time when ASD symptoms become visible vary,

* Corresponding Author jdb.balaraman@gmail.com there are some cases where the child develops normally but starts to show symptoms at age 2 or 3 years old (*Autism Spectrum Disorder Fact Sheet*, n.d.). While there are disorders related to ASD ranging from Fragile X Syndrome to tuberous sclerosis, there is no definitive known root cause to

Advisor: Michelle Chin mchin88@knights.UCF.edu ASD. In past years, there have been many developments with regards to increasing the quality of life of those with ASD. Though ASD is often considered a lifelong disorder, there is a group of individuals who lose some or all of their symptoms. This characteristic is classified as the optimal outcome. This characteristic is measured by utilizing a scale to rate qualitative characteristics ranging from visual reception to communication and socialization. The main objective was that with a comprehensive behavioral and developmental plan at an earlier age, consisting of intervention from trained therapists and clinicians, would enable the symptoms of children with ASD to improve and increase their chances of becoming healthier or achieving the optimal outcome. The two possible outcomes of the randomized control trial that was conducted to test the efficiency of the ESDM, was an improvement in symptoms or a decline in symptoms which was run utilizing screening examinations, to find a qualitative measure and compared to their initial scores. This review aims to investigate the correlations of genetics. diagnosis and comprehensive early intervention on the possibility of achieving the optimal outcome.

2. Genetics of ASD

Genetics plays an imperative role in ASD. There are around 100 genes associated with ASD and it differs in each patient, demonstrating how difficult it is to diagnose ASD definitively. Moreover, ASD is not associated with a single gene. It is not only the number of genes that cause this issue, but also that these genes only account for some of the total ASD cases, but not all of them. As a spectrum-disorder, ASD has a varying level of severity due to different amounts of genes with mutations. This paper will highlight two genes, *Phosphatase and tensin homolog* (PTEN) and *tuberous sclerosis complex 1 and 2* (TSC1/TSC2) that have shown to be associated with ASD.

According to the Brain Facts Book, PTEN is a tumor suppressor enzyme that controls cell division and prevents uncontrolled cellular proliferation. It also activates the mTOR pathway through its repression of several intermediate proteins (Figure 1).

The *mTOR pathway* is a significant part of cell proliferation and growth - it monitors "the availability of nutrients, mitogenic signals and cellular energy and oxygen levels" (Zarogoulidis, et al., 2014). If PTEN mutations, such as PTEN haploinsufficiency, are what primarily cause ASD, a potential solution is inhibiting the mTOR pathway (Sato, 2016). To test this theory, scientists have conducted studies with mice who have PTEN mutations. PTEN-mutant mice had altered sociability, anxiety, and repetitive behaviors. Scientists used drugs that inhibit the mTOR pathway (Wnuk, et al., 2018). The mice's anxiety, social. and communication skills markedly increased after treatment, thus resulting in an improvement in quality of life (Sato, 2016). These results indicate the potential cause and treatment for those with ASD. Taken together, these results suggest that inhibiting the mTOR pathway and its rapid cell proliferation and growth could be a possible factor in improving ASD symptoms.



Figure 1. *P13K-PTEN-mTOR Pathway*. The figure shows how PTEN and TSC1/2 are upstream modulators of the mTOR pathway. Activation of the mTOR pathway leads to effects downstream in cell growth and proliferation (Lasarge & Danzer, 2014).

In addition, TSC1 and TSC2 are also associated genes with ASD. TSC1 and TSC2 are genes that code for proteins that control cell growth and size. A mutation on TSC1 and TSC2 causes tuberous sclerosis, which is characterized by uncontrollable cell growth, which could lead to numerous benign tumors in the body. (*Tuberous Sclerosis Complex*, 2020). Interestingly, 40% of patients with tuberous sclerosis also have ASD, demonstrating the possibility of a similar genetic cause. Contributing to the mTOR pathway, these genes have a major role in terms of cell growth, metabolism and proliferation (Wnuk, et al., 2018). This once again shows the correlation between the mTOR pathway and ASD, which can provide a different outlook on the cause of autism. If a genetic characteristic of autism was discovered, early diagnosis would become possible similar to the means of diagnosing disorders like Down-syndrome, which is associated with Trisomy 21. The identification of a trisomy 21 immediately that Down syndrome dictates is present, demonstrating how it could be diagnosed easily.

Although PTEN and TSC1/TSC2 are important for the development of ASD, twin studies now show that genetics may not be the only factor in developing ASD. Some people are more prone to ASD than others. If one twin were diagnosed with ASD, the other twin is only 36-95% likely to be diagnosed as well (Autism Spectrum Disorder Fact Sheet, n.d.). Identical twins have the same genetic makeup so one would assume the genetic heritability is 100%, which is not the case here. This suggests that there may be other epigenetic factors that contribute to ASD development. These questions can only be answered with further research in this field. Because the genetic makeup of identical twins are the same, the fact that the other twin is not guaranteed to have ASD shows that genetics is not the only factor in play, which once again describes the complexity of diagnosis.

3. Quantitative and Qualitative Diagnostic Measures

ASD is diagnosed with both qualitative and quantitative measures. The qualitative measure is usually done by the parents or pediatricians who screen for the associated symptoms of ASD, including lack of communication skills, lower intellectual ability, and repeated rigid movements. There are three types of tests that screen for ASD in a quantitative manner. The Social Communication Questionnaire (SCQ) and the Social Responsiveness Scales (SRS) are easy and inexpensive ways to screen for ASD routinely. The Autism Diagnostic Observation Schedule (ADOS) is another quantitative test that is conducted by a trained-physician.

The SCQ is a questionnaire type test. It was created for children 4 years and older, who have a mental age above 2. Mental age is defined as a combination of factors including responsibility, IQ, and social skills at a specific time. The time duration assigned to complete the assessment is 10 minutes. The SCQ is to be completed by either the parent or caregiver and consists of 40 yes-or-no questions. An example of a question is "gets hyperactive, angry, screams, yells often" and the parents would have to answer true or false for this question (Social Communication Questionnaire: Questions, General Description, Uses and Limitations, 2019). The assessment measures the communication and social skills from a parent/caregiver standpoint. This questionnaire utilizes the qualitative symptoms observed by parents to provide the predicted severity of each symptom, thus providing a diagnosis if the cutoff score meets the limit needed for ASD. There are two versions of this specific assessment -Lifetime and Current (Rutter, 2003). The Lifetime SCQ assessment utilizes the child's complete developmental history. The cutoff score for a probable diagnosis of ASD on this type is 15 where the possible score can range from 0-40. The Current SCQ assessment utilizes the child's developmental history for the past 3 months and is used to determine improvements or worsening of ASD symptoms.

The SRS is also a questionnaire type test. The second edition SRS assessment consists of 4 different types of forms - a school age form, a preschool form, an adult (relative/other report) form, and an adult (self-report) form (Social Responsiveness Scale, Second Edition (SRS-2), n.d.). The school age form is created for children of ages 4 to 18 and should be completed by a parent/caretaker. The preschool form is created for children of ages 2.5 to 4.5 and should be completed by a parents/caretaker as well. The adult (relative/other report) form is for ages 19 and up and should be completed by the adult themselves. The time duration to complete the assessment is 15 to 20

minutes, which is longer than the SCQ assessment. The SRS reveals aspects of social skills that are usually missed by parents/caregivers - the yes-or-no questions are key for this principle (Moody, et al., 2017). Social awareness, social cognition, social communication, social motivation, and restricted interests and repetitive behaviors are the five subscales in which the severity and social deficits are portrayed by the total score (Social Responsiveness Scale, Second Edition (SRS-2), n.d.). The aggregate of all five subscales determines the severity of ASD symptoms. The score for the SRS ranges from 15 to 60. 30 is the cut-off score on this assessment where a score in the range 30-36 results in the diagnosis of mild/moderate autism, a score greater than 37 is classified as severe autism and a score below 30 is classified as no autism (Park, et al., 2018). Some limitations of self report tests are confirmation and observer bias. For example, if a person is filling out the SRS questionnaire for themselves or for their child and firmly believes that they have ASD, due to confirmation bias, they are more likely to answer yes to some questions. Confirmation bias is the tendency to interpret evidence or answer questions as a way to confirm one's existing beliefs (Nickerson, 1998). Observer bias is a discrepancy that is caused during the process of observing or recording information (Mahtani, et al., 2018). Confirmation and observer bias could potentially skew the results of the questionnaire, thus affecting the patient's official score and diagnosis.

Another diagnostic measure that is more accurate than SCQ and SRS is the ADOS test. It is an activity-based assessment done by trained clinicians who utilize controlled scenarios and makes direct observations about the patient's behaviors. Clinicians are paired with individuals who they have no relationship with. (Dreison, 2019). The ADOS test can be done for 12-month infants and older and usually takes 30-60 minutes to administer. It consists of direct observation of the symptoms present as well as interviews with the individual and possibly the individual's parents (Dreison, 2019). Confirmation and observer bias is reduced because a team of trained clinicians review a video recording to make a diagnosis. It also takes around 1-2 weeks to receive the results. Although the ADOS test takes more time, there is less bias because it is conducted by trained clinicians.

These quantitative methods - SCQ, SRS, and ADOS - use subjective observations of symptoms in children to make a diagnosis. However, there is currently no pure quantitative test that does not revolve around the symptomatic basis of ASD. If such a source for diagnosis were available, this would mean a more definitive diagnosis for ASD, which would reduce the bias sometimes produced by SCQ and SRS. SRS, SCQ, and ADOS assessments remain an effective mechanism to screen for ASD, however, there are limitations. Therefore, if there existed a quantitative diagnostic, that would be more beneficial for early intervention, which would yield a higher probability for an optimal outcome.

4. The Benefits of Early Intervention and Connection to Optimal Outcome

Some studies have demonstrated that early intervention is primarily what increases the chances of achieving an optimal outcome. One such early intervention that was used in a randomizedcontrolled trial was the Early Start Denver Model (ESDM). The ESDM was a comprehensive developmental behavioral intervention that was delivered by trained therapists and parents for 2 years. (Dawson, et al., 2010). 48 children with ASD who were 18-30 months old were randomly assigned to one of two groups, the one that used the ESDM, or the assess and monitor group (A/M). In the A/M the child was referred to a local provider that provided intervention that is commonly provided in the community. (Dawson, et al., 2010). The A/M group got yearly assessments with recommendations on possible interventions as well as referrals for intervention from community providers in the region. The ESDM group received yearly assessments, 20 hours a week of the ESDM intervention from clinicians, parent training, and parent administration of the intervention for 5 or more hours per week (Dawson, et al., 2010).

The results of these trials after 2 years were significant improvements in IQ, adaptive behavior, and autism diagnosis for the ESDM group when compared to the A/M group. In addition, the ESDM

group displayed a 17.6 point improvement in their ADOS score collected at both the beginning and end of the trial, while those in the A/M group displayed an average of 7.0 point improvement. This demonstrates that when compared to "average" intervention, the ESDM group had more of a positive impact for those with ASD proving that it is more effective in this trial. The study also highlighted the importance of an early and comprehensive developmental behavioral intervention, as evidenced by improvements in IQ, adaptive behavior, and ASD diagnosis after the 2 years of the trial. The A/M community intervention represents a normal intervention program that most people diagnosed with ASD would likely get. The difference in effectiveness between the ESDM and the community intervention shows that initiating interventions at an earlier stage that is more specific to the symptoms exhibited seems to reduce the severity of ASD and produce long-term results. It brings them closer to the ultimate goal of optimal outcome.

There primarily seems to be a correlation between early, focused intervention and optimal outcome. When there is early intervention, this is caused directly because of early diagnosis. If intervention is started early, the child's brain is more "plastic' or changeable" giving children the best chances to develop to a greater level. This characteristic makes it easier to improve based on treatment plans like behavioral and speech therapy at younger ages, relating to the direct correlation between early intervention and optimal outcome (Early Intervention for Autism, n.d.). In some cases, this means losing all the symptoms associated with ASD and having a tremendous improvement in both IQ and adaptive behavior as well as social skills like communication. However, most often, an improvement in IQ, social skills, adaptive behavior, and ASD diagnosis is associated with reduced symptoms translating to an optimal outcome (Fein, et al., 2013). This demonstrates the associated relationship between early, focused interventions and an increased chance of achieving an optimal outcome.

5. Conclusion

There are many diagnostic measures and

interventions present for diagnosing and treating the symptoms of ASD. The current diagnostic methods are usually based on qualitative measures as well as quantitative measures that primarily analyze the symptoms present. As discussed in this paper, the genetics behind ASD is very complex and there are over 100 genes associated with the disease. Finding a genetic mechanism instead of a gene that contributes to ASD would facilitate the creation of a definitive diagnostic measure. Through analysis of different interventions such as the ESDM intervention that was composed of highly specific behavioral therapy, it was observed that early specific intervention tends to increase the chances for optimal outcome. The combination of definitive diagnosis as well as early specific intervention correlates with optimal outcome in those with ASD.

Overall, there are many areas for research in ASD. One of the main things that researchers should focus on is finding a genetic mechanism or an "identifying factor" that correlates to ASD. Identifying a mechanism would contribute to a faster and more definitive form of ASD diagnosis. Diagnosing ASD at an earlier stage leads to earlier intervention for children, thus increasing the chances for optimal outcome. Furthermore, another area which requires development is the creation of patient-specific treatment for those diagnosed with ASD. By doing so, there would be a patient-specific treatment plan that would be followed which would help reduce the severity of the symptoms and increase the chances of optimal outcome. Throughout the years, the quality of life of those with ASD has improved greatly and with continued advancement more developments are to come for those with ASD.

References

Autism Spectrum Disorder Fact Sheet. (n.d.). National Institute of Neurological Disorders and Stroke. Retrieved January 20, 2021, from https://www.ninds.nih.gov/Disorders/Patient-Caregiv er-Education/Fact-Sheets/Autism-Spectrum-Disorder -Fact-Sheet.

Dawson, G., Rogers, S., Munson, J., et. al. (2010). "Randomized, Controlled Trial of an Intervention for Toddlers With Autism: The Early Start Denver Model." *PEDIATRICS*, 125(1), e17–e23. https://doi.org/10.1542/peds.2009-0958

Dreison, K. (2019, October 27). A Brief Overview of the ADOS-2: An Assessment for Autism Spectrum Disorder. Children's Resource Group. Retrieved, March 25, 2021, from https://www.childrensresourcegroup.com/a-brief-ove

rview-of-the-ados-2-an-assessment-for-autism-spectr um-disorder/.

Early Intervention for Autism. (n.d.). Eunice Kennedy Shriver National Institute of Child Health and Human Development. Retrieved November 2, 2021, from

https://www.nichd.nih.gov/health/topics/autism/conditioninfo/treatments/early-intervention/.

Fein, D., Barton, M., Eigsti, I., et. al. (2013). "Optimal Outcome in Individuals with a History of Autism." *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 54(2), 195–205. https://doi.org/10.1111/jcpp.12037

Lasarge, C., Danzer, S. (2014). *Mechanisms* regulating neuronal excitability and seizure development following mTOR pathway hyperactivation. Frontiers in Molecular Neuroscience. 7(18). https://doi.org/10.3389/fnmol.2014.00018

Mahtani, K., Spencer, E. A., Brassey, J., et. al. (2018, February 1). *Catalogue of bias: observer bias*. BMJ Evidence-Based Medicine, 23(1), 23–24. https://doi.org/10.1136/ebmed-2017-110884

Moody, E. J., Reyes, N., Ledbetter, C., et. al. (2017, November). Screening for Autism with the SRS and SCQ: Variations across Demographic, Developmental and Behavioral Factors in Preschool Children. Journal of Autism and Developmental Disorders, 47(11), 3550–3561. https://doi.org/10.1007/s10803-017-3255-5

Nickerson, R. S. (1998). *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*. Review of General Psychology, 2(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Park, H. S., Yi, S. Y., Yoon, S. A., et. al. (2018, October 1). Comparison of the Autism Diagnostic Observation Schedule and Childhood Autism Rating Scale in the Diagnosis of Autism Spectrum Disorder: *A Preliminary Study*. Journal of the Korean Academy of Child and Adolescent Psychiatry, *29*(4), 172–177. https://doi.org/10.5765/jkacap.180015

Rutter, M. (2003). *(SCQ) Social Communication Questionnaire*. WPS: Educational & Psychological Assessments for Clinicians & Educators. Retrieved March 10, 2021, from https://www.wpspublish.com/scq-social-communicati on-questionnaire.

Sato A. (2016). *mTOR*, a Potential Target to Treat Autism Spectrum Disorder. CNS & neurological disorders drug targets, 15(5), 533–543. https://doi.org/10.2174/187152731566616041312063 8

Social Communication Questionnaire: Questions, General Description, Uses and Limitations (2019, February 6). QuestionPro. Retrieved March 17, 2021, from

https://www.questionpro.com/blog/social-communica tion-questionnaire/

Social Responsiveness Scale, Second Edition (SRS-2) (n.d.). Pearson Assessment. Retrieved March 25, 2021, from https://www.pearsonclinical.com.au/products/view/5 12.

Tuberous Sclerosis Complex. (2020, August 18). MedlinePlus. Retrieved April 5, 2021, from https://medlineplus.gov/genetics/condition/tuberous-s clerosis-complex/.

Wnuk, A., Davis, A., Parks, C., et. al. (2018). *Brain Facts: a Primer on the Brain and Nervous System*. Society for Neuroscience.

Zarogoulidis, P., Lampaki, S., Turner, J. F., et. al. (2014). *mTOR pathway: A current, up-to-date mini-review (Review)*. Oncology Letters, 8(6), 2367–2370. https://doi.org/10.3892/ol.2014.2608



Utilization of Thermoacoustics in Developing Energy-Efficient Cooling Technology

Andrew Kim^{1*}, Caroline Sheffet¹

¹ Montclair Kimberley Academy, Montclair, New Jersey USA

Received June 11, 2021; Revised February 3, 2022; Accepted, February 7, 2022

Abstract

The considerable amount of energy required to run modern-day air conditioning systems results in the emission of potent greenhouse gases, which poses various potential environmental and health risks. To provide a solution to the disadvantages of standard air conditioning, this paper observes the harmful effects standard air conditioning has on the global environment and describes the development of a possible solution, thermoacoustic technology, which would provide a cleaner method of cooling. It was determined that thermoacoustic cooling technology utilizes the motion of sound waves in a contained environment to allow for the transfer of heat within the device. This paper provides a thorough description of the design and development of thermoacoustic cooling technology, including an explanation of what would physically occur within the system as it operates by analyzing the sound waves and heat transfer; the materials needed to build the structure; and how the materials would be incorporated to create this environmentally clean technology.

Keywords: Thermoacoustics, Air Conditioning, Thermoacoustic Cooling

1. Introduction

1.1 Health Risks Associated with High Temperatures

Heat waves are defined as sustained periods of abnormally excessive heat. Heat waves are considered by the World Health Organization to be one of the most hazardous natural dangers because they pose many threats to human health such as blood clots, acute cerebrovascular accidents, severe dehydration, heat exhaustion, and heat stroke. Heat stroke symptoms include faintness, headaches, swelling, heat rash, lack of energy and weakness, cramps, irritability, and dry, warm skin. Overheating can even lead to death. The World Health Organization reported more than 166,000 deaths globally due to heat waves between 1998 and 2017 (Heatwaves, n.d.). According to the Centers for Disease Control and Prevention (CDC), from 2004 through 2018, heat-related conditions caused 10,527 deaths in the United States (Vaidyanathan et al., 2020, pp. 729-734). Statistical approaches approximate that extreme heat causes more than 1,300 deaths in the United States per year (Climate Change, 2020). A heat wave that occurred in Europe in 2003 led to more than 70,000 deaths alone. Individuals younger than four or older than sixty-five years old, as well as those who routinely consume prescriptions, are most vulnerable to heat waves and have higher chances of dying or experiencing health complications (Heatwaves, n.d.).

Advisor: Dr. Jacquelyn Blum jblum@mka.org

^{*} Corresponding Author andrewjkim2020@gmail.com

High temperatures pose immediately apparent health risks, such as heat stroke and heat-related illnesses; however, other less obvious health risks are often overlooked. For example, heat increases the risk of triggering heart conditions, such as heart attacks and respiratory diseases, like pneumonia. As a result, scientists predict that death reports vastly underestimate those related to excess heat exposure (Climate Change, 2020). Several studies suggest that many "heat-related" deaths are not being recorded as such and thus are missing from the calculation, making it difficult to fully interpret data. Data that is available, however, demonstrates that a substantial increase in the occurrence of hot temperatures and heat waves is closely linked to the vast rise in heat-related deaths (Vaidyanathan et al., 2020, pp. 729-734).

1.2 Environmental Effects of Air Conditioning

As a result of global warming, heat exposure is increasing in frequency, duration, and magnitude globally. From 2000 to 2016, population exposure to heat waves increased by around 125 million people (Heatwaves, n.d.). As displayed in the graph, 2006 was one of the hottest years on record in the United States with frequent heat waves, correlates with many more heat-related deaths (Vaidyanathan et al., 2020, pp. 729-734).



Figure 1. Data Source: Climate Change Indicators: Heat-Related Deaths. (2020, November 2). United States Environmental Protection Agency. Retrieved April 21, 2021, from https://www.epa.gov/climate-indicators/climate-chan ge-indicators-heat-related-deaths#:~:text=Some%20st atistical%20approaches%20estimate%20that,set%20s hown%20in%20Figure%201.

From the 1960s to the 2010s, across 50 metropolitan areas, the mean number of heat waves increased from 2 to 6 per year, and the heat wave season increased by 47 days; 92% of these 50 cities saw a rise in the occurrence of heat waves, and 88% experienced elongated seasons (U.S. Environmental Protection Agency, 1961-2018). These effects are more obvious in metropolitan areas; however, per the urban heat island effect, non-urban communities are also harmed by unusually high temperatures. For instance, heat waves place a strain on energy, transportation, water and food supplies, human health, and emergency services (Heatwaves, n.d.). The Earth continuously warms and heat waves become more frequent. Throughout this cycle, air conditioning remains the standard cooling technology to accommodate rising temperatures, but it also poses a serious threat to the environment by contributing to global warming.

Air conditioners often emit harmful by-product gasses because they run on electricity, which is commonly generated by burning fossil fuels. By 2050, some studies predict, a quarter of global warming will be as a result of this process air conditioning operates on (Mize, n.d.). According to Daikin, the largest selling air conditioning company in the world, air conditioning units require large amounts of energy; the refrigerants destroy the ozone layer and worsen global warming (Response to Climate, n.d.). Before the year 2010, air conditioners constantly emitted hydrochlorofluorocarbons (HCFCs), which are proven to deplete the ozone layer. In 2010, the production, use, and import of HCFCs was banned, with the exception of necessary use for the servicing of existing equipment. As a result, since 2010, Hydrofluorocarbons (HFCs), which do not break down the ozone layer, are the primary refrigerant that air conditioning systems require to operate. Although HFCs do not harm the ozone layer, they are still much more potent greenhouse gases than CO₂ (Phasing out HCFC, HCFC-22 (CHCIF₂) and 2015). HFC-134a (CH₂FCF₃) are two very potent greenhouse gases that have been a very significant part of air conditioning systems worldwide. According to a study done by the Proceedings of the National Academy of Science of the United States of America (PNAS), the global

excretion of these harmful gases has drastically increased over the last 20 years. Since 2000, HFC-134a emission levels consistently surpass the United Nations Framework Convention on Climate Change, even by 60% from the years 2009 to 2012 (Xiang et al., 2014). While HCFCs have been phased out over the past three decades, older air conditioning units are still leaking HCFC gas, and modern units that run on HFCs are still contributing to high energy use and greenhouse gas emissions (Phasing out HCFC, 2015). The banning of HCFCs marks progress towards a cleaner Earth, but HFCs are still easily and commonly leaked at any point in the air conditioning process from manufacturing, to installation, to disposal. R-407C and R-410A are among the most commonly used HFCs for air conditioners. Over a thousand times more potent than CO₂, the disastrous potential HFCs have on Earth's climate is often overlooked (Ospina, 2018).

One solution to the environmental problems resulting from current air conditioning methods is the use of thermoacoustics to provide a cleaner and safer alternative to standard cooling. The thermoacoustic system discussed in this paper requires no moving mechanical parts, no CO₂, no precious materials, and no refrigerants (Blain, 2019), thus serving as a more environmentally safe method of cooling.

2. Development of Thermoacoustic Cooling Technology

This study adopts the design of a standing-wave thermoacoustic refrigerator. Some benefits of thermoacoustic refrigeration include the use of environmentally friendly fluids (mediums) and relatively few moving components in development (Design and simulation of small capacity thermoacoustic refrigerator, 2019). The key parts of the device include heat exchanges, stack, resonator, and a source of energy to drive the device and enable for the transferring of heat energy.

2.1 Stand Waves and Wave Formation

Stand waves are the pattern result of a reflected wave interacting with waves sent from the energy source. This process occurs when both waves cross through a shared medium, the substance the wave is carried by (Formation of Standing Waves, n.d.). Two types of points exist within the stand wave pattern. The first is the node (Figure 2), a location with destructive wave repetitive interference (no displacement), and the second is the antinode, a with repetitive constructive location wave interference (maximum displacement). The properties of stand wave patterns allow for their formation in pipes and tubes. Due to the enclosed tubular design of the thermoacoustic cooling device, stand waves are able to be formed due to the continuous formation and reflection of waves to and from the direct source.

The consistent formation of waves inside the thermoacoustic device are dependent on the medium used and an energy/wave source. The vibration of an object or medium is the source of wave formation. It forces the particles of the medium to vibrate, and the resonance of such vibrations (ability of vibrating systems to force surrounding systems to vibrate) enables the formation of sound waves. As particles continue to vibrate, the waves are transmitted (Sound Waves in Air, n.d.).



Figure 2. Visual Representation of stand wave and respective node/antinode points; nodes are regions of destructive interface as antinodes are regions of constructive interface (Stand waves review, n.d.).

2.2 Waves and Heat Transfer

The properties of sound waves allow for the transfer of heat within various components of the thermoacoustic cooling device. In air, sound waves are longitudinal waves where the displacement of the medium is parallel to the movement of the wave. Sound waves are also capable of forming pressure

allowing compression variations, for the (constructive stand wave interface), rarefaction (destructive interference) and a back-and-forth motion of particles in the medium with respect to the direction of the sound (Sound Waves in Air, n.d.). This process of particle motion is critical for the thermoacoustic capabilities of the device design. When a gas medium is compressed and makes contact with a solid surface, heat is transferred from the gas to the surface. On the other hand, heat is released back to the gas molecules during rarefaction (An Overview of Stack Design for a Thermoacoustic Refrigerator, 2015). This oscillatory pattern is what drives the continuous withdrawal and replacement of heat within the solid surfaces of the thermoacoustic device, allowing for the transfer of thermal energy.



Figure 3. Representation of the compression and rarefaction of sound waves with respect to variations in pressure (Sound Waves in Air, n.d.).

2.3 Visualization of the theoretical product

The figure below (Figure 4) illustrates a sample design of a thermoacoustic cooler that was adapted for the purposes of the study.





2.4 Resonator

The resonator or resonance tube is the location where wave oscillation occurs, specifically a standing wave pattern. Ultimately, it is the outermost component or shell of the thermoacoustic cooling system. The frequency and wave amplitude are each dependent on the design (geometry and dimensions) and material composition of the resonator. The resonance tube should be constructed with light yet durable material to resist the variation in pressure caused by sound wave oscillation (Measuring the performance of different stack materials in thermoacoustic device, 2015). In addition, the length of the resonator often depends on the wavelength of the acoustic wave and can influence the amount of energy lost within the system (Design and simulation of a small capacity thermoacoustic refrigerator, 2015). The resonator is filled with the gas medium and surrounds the stack and heat exchangers, which will be described in the subsequent sections.

2.5 Stack

The stack is responsible for the movement of heat. Both a stand wave pattern and the design of the stack are critical to ensure the transfer of heat from the cold end of the stack to the hot end of the stack (The Reality of a Small Household Thermoacoustic Refrigerator, 1996). Stacks can vary in geometries, such as spiral or pin-array type, but a commonly used design is parallel plate stacks. In the most favorable circumstances, the stack would be produced from material with a low thermal conductivity, compared to the used medium, to avoid heat being conducted back to the cold end of the stack. Literature states that Mylar, a polyester film, is a widely used material used for thermoacoustics due to its low thermal conductivity (An Overview of Stack Design for a Thermoacoustic Refrigerator, 2015). The material of the stack is the solid surface where heat exchange occurs with the compression and rarefaction of waves, as described previously.

When the medium and stack surface make contact, the thermal penetration depth must also be considered to determine an optimal spacing between stacks. Thermal penetration is the depth at which heat can diffuse through a gas. Research concludes that an effective plate spacing length is about 0.3 mm or roughly three times the thermal penetration depth (The optimal stack spacing for thermoacoustic refrigeration, 2002).

2.6 Heat exchangers

There are two heat exchangers in thermoacoustic cooling technology, a hot and cold exchange. They are placed on opposite ends of the stack and are responsible for moving heat in and out of the stack and its surroundings. Heat is pumped from the cold exchanger to the hotter one through the stack. Relative to the design, the cold exchanger is to be set adjacent to the resonator. To ensure the hot changer does not absorb too much heat, it should be able to release excess heat through water cooling (Design of thermoacoustic refrigerators, 2001).

2.7 Source of sound

A driver is required to create the waves inside the resonator medium and allow for the formation of stand waves and, thus, the transfer of heat through the stack. One source would be a loudspeaker. To mimic the process of refrigeration, the end of the loudspeaker should be sealed to the device to avoid leakage (An Overview of Stack Design for a Thermoacoustic Refrigerator, 2015). A standard practice in acoustic technology is implementing a quarter wavelength.

2.8 Choice of Medium

Optimal conditions for the medium gas inside the resonator of which the sound waves pass through include high sound velocity and thermal conductivity to allow for a smooth flow of heat transfer. Helium is effective medium commonly used for an thermoacoustic devices because its low density correlates to a higher thermal conductivity and allows sounds to pass through quicker. Air is less effective than helium but still serves well as a significantly cheaper and accessible substitute that can be mixed with other non-flammable inert gases like argon (Measuring the performance of different stack materials in thermoacoustic device, 2015).

3. Discussion

Upon researching the necessary components of a thermoacoustic refrigerating device, a theoretical model for a thermoacoustic cooler would be composed of the following materials. A tubular resonator that fits the disk shape of a loudspeaker would serve as the outermost layer of the cooler. A stack composed of a series of evenly spaced, parallel Mylar plates would fit inside the resonator. Two cylindrical steel lids with steel fins would serve as effective hot and cold heat exchangers, and would be inserted at both ends of the stack system within the resonator (Development of a thermoacoustic heat pump for distillation column, 2017).

We recognize the development and implementation of an actual thermoacoustic cooling device is a costly procedure and have, therefore, deemed it not yet suitable for domestic and personal use. We believe that the cooling technology will serve best in corporations or large infrastructures with regards to affordability and increased energy efficiency, targeting those that require a substantial amount of energy for daily operations and/or are located in hot climate regions. Given our preliminary design of a thermoacoustic cooler, our future goals involve analyzing more optimal conditions for this device including seeking less expensive alternatives to helium and determining the optimal stack spacing given the thermal penetration depth of various gas mediums. Although, this type of technology is not widely affordable at its current state, it has potential to become a standard household application while maintaining the same cooling properties as standard air conditioning. Thermoacoustics is a growing field of science and much is yet to be uncovered; we are hopeful that the development of a thermoacoustic device will revolutionize the cooling industry and provide an alternative and cleaner method of cooling.

Acknowledgements

We would like to thank Gianna Zou for helping with our research and Dr. Jacquelyn Blum, MKA, Montclair for being our advisor.

References

Bartos, D., Kurzbach, S., Filipiak, Z., & Petro, M. (2015). Measuring the performance of different stack materials in thermoacoustic device. Roskilde University.

Bhansali, P. S., Patunkar, P. P., Gorade, S. V., Adhav, S. S., & Botre, S. S. (2015). An Overview of Stack Design for a Thermoacoustic Refrigerator, International Journal of Research in Engineering and Technology, 04(06), 68-72. https://ijret.org/volumes/2015v04/i06/IJRET2015040 6010.pdf

Blain, L. (2019, January 25). Electricity-free air con: Thermoacoustic device turns waste heat into cold using no additional power. New Atlas. Retrieved April 25, 2021, from

https://newatlas.com/soundenergy-thermoacoustic-co oling/58169/

Climate Change Indicators: Heat-Related Deaths. (2020, November 2). United States Environmental Protection Agency. Retrieved April 21, 2021, from https://www.epa.gov/climate-indicators/. climate-change-indicators-heat-related-deaths#:~:text =Some%20statistical%20approaches%20estimate%2 0that,set%20shown%20in%20Figure%201

Formation of Standing Waves. (n.d.). Physics Classroom.

https://www.physicsclassroom.com/class/waves/Less on-4/Formation-of-Standing-Waves

Heatwaves. (n.d.). World Health Organization. Retrieved April 25, 2021, from https://www.who.int/health-topics/heatwaves#tab=tab _1

Kajurek, J., Rusowicz, A., & Grzebielec, A. (2019). Design and simulation of a small capacity thermoacoustic refrigerator. SN Applied Science. https://link.springer.com/content/pdf/10.1007/s42452 -019-0569-2.pdf

Mize, D. (n.d.). Do Air Conditioners Cause Any Harm to the Environment? Masters Heating and Cooling Inc. Retrieved April 25, 2021, from https://mastersheatcool.com/blog/do-air-conditionerscause-any-harm-to-the-environment/ Nave, R. (n.d.). Sound Waves in Air. HyperPhysics. http://hyperphysics.phy-astr.gsu.edu/hbase/Sound/tral on.html

Ospina, C. (2018, August 7). Cooling Your Home but Warming the Planet: How We Can Stop Air Conditioning from Worsening Climate Change. Climate Institute. Retrieved April 25, 2021, from https://climate.org/cooling-your-home-but-warming-t he-planet-how-we-can-stop-air-conditioning-from-wo rsening-climate-change/

Phasing out HCFC refrigerants to protect the ozone layer. (2015, July). United States Environmental Protection Agency. Retrieved April 27, 2021, from https://www.epa.gov/sites/production/files/2015-07/d ocuments/phasing_out_hcfc_refrigerants_to_protect_ the_ozone_layer.pdf

Response to Climate Change. (n.d.). Daikin Global. Retrieved April 25, 2021, from https://www.daikin.com/csr/environment/climatechan ge/ #:~:text=Air%20 conditioners%20consumer%20 large%20amounts,and%20contribute%20to%20globa l%20warming

Review of investigation in eco - friendly thermoacoustic refrigeration system. (2017). A. Raut, U. Wankhede. Retrieved from https://www.semanticscholar.org/paper/Review-of-in vestigations-in-eco-friendly-system-Raut-Wankhede/ 93a58b81734b90012ee9d9627dbc6927760b5f59

Starr, R., Bansal, P.K., Jones, R. W., & Mace, B. R. (1996). The Reality of a Small Household Thermoacoustic Refrigerator. Purdue e-Pubs. https://docs.lib.purdue.edu/cgi/viewcontent.cgi?articl e=1343&context=iracc

Tijani, H., Lycklama á Nijeholt, J.A., & Spoelstra, S. (2017). Development of a thermoacoustic heat pump for distillation column. IEA. http://hpc2017.org/wp-content/uploads/2017/05/O.3. 7.5-Development-of-a-thermoacoustic-heat-pump-for -distillation-column.pdf

Tijani, H., Zeegers, J., & De Waele, A. (2001). Design of thermoacoustic refrigerators. Cryogenics

Tijani, H., Zeegers, J., & De Waele, A. (2002). The optimal stack spacing for thermoacoustic refrigeration. The Journal of the Acoustical Society of America, 112(1), 128 - 133. DOI:10.1121/1.1487842

Standing waves review. (n.d.). Retrieved June 15, 2021, from

https://www.khanacademy.org/science/high-school-p hysics/waves-and-sound/standing-waves-2/a/standing -waves-review-ap

U.S. Environmental Protection Agency. (1961-2018). U.S. heat wave frequency and length are increasing. U.S. Global Change Research Program. Retrieved April 25, 2021, from https://www.globalchange.gov/browse/indicators/us-h eat-waves#:~:text=1.,per%20year%20during%20the %202010s

Vaidyanathan, A., Malilay, J., Schramm, P., & Saha, S. (2020). Heat-Related Deaths — United States, 2004–2018. Morbidity and Mortality Weekly Report, 729-734.

http://dx.doi.org/10.15585/mmwr.mm6924a1

Xiang, B., et al. (2014). Global Emissions of Refrigerants HCFC-22 and HFC-134a: Unforeseen Seasonal Contributions. National Academy of Sciences, 111(49), 17379-17384. https://doi.org/10.1073/pnas.1417372111



Future Possibilities for the use of CRISPR on Mutations in Three Eye Disorders

Anika Ranjan^{1*}

¹ Novi High School, Novi, MI USA

Received August 31, 2021; Revised January 16, 2022; Accepted, January 20, 2022

Abstract

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) technology is a useful tool to insert, delete, and substitute DNA in the genome. This is done by separating DNA via double-stranded breaks and can be performed through two different mechanisms: Homology-directed Repair (HDR) and Non-Homologous End Jointing (NHEJ). Double strand breaks would be used for a tracrRNA:crRNA to guide the enzyme cas9 to insert, delete, or substitute the desired DNA. CRISPR has limitations on which DNA sequences it can work with. Other concerns include ethical questions and base-pair limitations. X-linked Congenital Night Blindness, Snowflake Vitreoretinal Degeneration, and Cataract Microcornea syndrome are three genetic diseases caused by mutations in the genes CACNA1F, KCNJ13, and ABCA3 respectively. This research paper will discuss the possibilities of CRISPR and decipher specific substitutions for all three eye disorders.

Keywords: X-Linked Congenital Night Blindness, Snowflake Vitreoretinal Degeneration, Microcornea Cataract Syndrome, CRISPR

1. Introduction

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) gene-editing technology in a form of gene therapy in which the scientist inserts an artificial guide-RNA (gRNA) that attracts CRISPR-Cas9 to a desired piece of DNA. Double strand breaks separate the DNA strand for the artificial gRNA, while Homology-directed Repair (HDR) and Non-Homologous End Jointing (NHEJ) correct the genome. HDR is usually the most preferred method.

CRISPR has advantages that make it a higher preference than other technologies. While CRISPR can substitute DNA in a genome and insert/delete base pairs, one crucial benefit is that CRISPR is higher affordability (Redman, et al.,2016). Other methods can cost upward of \$1,000, but CRISPR is relatively more affordable, costing less than \$100 per treatment for one patient (Doudna and Charpentier, 2014). Before 2007, it would take three months to administrate CRISPR gene therapy, but that time has now been shortened to 1 to 2 hours (Doudna and Charpentier, 2014). These benefits allow CRISPR treatments that are currently in trials to cure a variety of diseases including Sickle Cell Disease (SCD) and cancer (Cong and Zhang, et al., 2021). The deadly inherited SCD has already been treated in mice by using CRISPR (Newby, et al., 2021). A single letter of DNA in red blood cells caused the formation of a specific non-pathogenic variant, which conveniently multiplies healthy copies of itself (Newby, et al., 2021)! The initial mutation, caused by the SCD

Advisor: Kevin DAmico kdamico@princeton.edu

^{*} Corresponding Author anikaranj@gmail.com

allele, reduced the red blood cells' ability to carry oxygen and curve into a crescent/sickle shape. The only current treatment is bone marrow transplant, but the treatment is difficult to obtain with specific conditions that might be deadly (Newby, et al., 2021).

The injection treatment in trials focuses on using an adenine base editor to target specific gene sequences for targeting specific sequences. On a more molecular level, the A-T base pairs are converted to G-C base pairs, as the SCD mutation is a consequence of the A base pair altering to a T (Siliezar, 2021). Previous human trials with the CRISPR treatment have determined that a maximum of 80% of the dangerous mutation were edited, providing hope for patients across the world (Siliezar, 2021). Researchers at the University of Pennsylvania are focusing on genetically modifying cancerous immune cells to fight a variety of different cancers via CRISPR. Currently, trials have only resulted in 10% success rate, but no harmful off-target effects have arisen. Before the drug can be introduced to the market, long-term effects need to be monitored: one such limitation of CRISPR.

Some of CRISPR's limitations include off-target effects, short base pair lengths, and potential unethical practices. Off-target effects include base pair changes that occur outside the targeted region on DNA; these base-pair changes can be transcribed into different proteins, creating more mutations in the genome (Plumer, et al., 2018). CRISPR technology has an off-target frequency higher than 50%, creating a need for engineering more CRISPR/Cas9 variants (Plumer, et al., 2018). One way to reduce off-target effects is to optimize the guide RNA (Plumer, et al., 2018). This, however, can now be performed using current technologies including E-CRISPR, CRISPR-design, and others (Heigwer, et al., 2014). Another limitation is the number of base pairs that CRISPR/Cas 9 can cut (Plumer, et al., 2018). The Cas9 enzyme is capable of cutting~ 20 base pairs in length (Zhang, et al., 2015). This property can be useful for diseases and disorders that have insertions and deletions that are under this limitation. X-Linked Congenital Night Blindness, Snowflake Vitreoretinal Degeneration, and Cataract Microcornea Syndrome all include mutations that are single substitutions. While ethical questions are raised on how much gene

editing should we do, CRISPR has a limit of 20 base pairs, and therefore cannot be used for every disease or disorder.

By collecting data of mutations from different sources for X-linked Congenital Night Blindness, Snowflake Vitreoretinal Degeneration, and Cataract Microcornea Syndrome, I have analyzed past research papers for substitutions that can potentially be fixed by CRISPR for researchers to target with future treatments. This paper discusses, the impact of the potential corrected mutation on the disorder as a whole and addresses ethical questions that may arise.

2. Findings

2.1 The KCNJ13 and ABCA3 genes with mutations defined



Figure 1. KCNJ13 and ABCA3 genes with mutations defined. The figure shows the R162W mutant located on the KCNJ13 gene. The R162W mutant is the result of a C to G base pair transition. The ABCA3 gene has multiple single substitution mutations: L39V-115C>G, V931-277G>A, and D1465N-4393G>A. All three mutations for the ABCA3 gene result in amino acid changes (Hejtmancik, et. al., 2008).

2.2 X-Linked Congenital Night Blindness

X-Linked Congenital Night Blindness is a recessive disorder, which is characterized by mutations in the retina, which helps to detect light and color (Boycott, et, al., 2001). Symptoms of X-Linked Congenital Night Blindness include night blindness, reduced acuity, myopia, and nystagmus (Boycott, et, al., 2001). The gene CACNA1F, responsible for the disease phenotype, is located on the retina on the surface of photoreceptors cells (Boycott, et, al., 2001). CACNA1F is responsible for encoding a retina-specific voltage-gated L-type calcium channel al-subunit, that functions as the pore and voltage sensor (Beck-Hansen and

Pearce, 1993). Both CACNA1F and another gene, NYX, make proteins that help in the process of passing visual signals from rods to cones in bipolar cells which helps with the transmission of visual from the eyes information to the brain (Beck-Hansen, 1998). In the CACNA1f gene, 14 mutations have been discovered in 36 patients; six of which caused premature protein truncation (Boycott, et, al., 2001). Studies have concluded that mutations can cause amino acid substitutions/deletions and premature protein truncation (Boycott, et, al., 2001). CACNA1F has 48 exons and a predicted amino acid length of 1977 (b.Beck-Hansen, 1998) (Figure 2).



Figure 2. The process of inserting a genetically modified strand using CRISPR. CRISPR RNA (crRNA) and trans-encoded small crRNA (tracrRNA) direct the Cas9 nuclease with the help of the PAM sequence (Plumer, et. al., 2018).

2.3 Snowflake Vitreoretinal Degeneration

Snowflake Vitreoretinal Degeneration is characterized by fibrillar degeneration of the vitreous humor, early-onset cataract and minute crystalline deposits among others (Edwards and Robertson, 2006). A mutation in the KCNJ13 gene is one of the causes. In a previous study, the KCNJ13 gene was analyzed and showed a c.484C > T transition, which changed the CGG codon to a UGG codon resulting in the R162W mutant (Hejtmancik, et al., 2008). In the 210 unaffected family members present in this trial, none of them contained this R162W mutant, providing further evidence that the R162W mutant could be the cause of Snowflake Vitreoretinal Degeneration. The R162W mutant then causes the production of the transmembrane protein Kir7.1 to cause major structural damage (Hejtmancik, et al., 2008). Kir7.1 is a Kþ-selective inward-rectified channel, located between 20 genes sequenced from D2S2158 to D2S2202 (Hejtmancik, et al., 2008). The R162W mutant is located in the Kir7.1 gene (Hirose, 1974).

2.4 Cataract Microcornea Syndrome

Cataract Microcornea Syndrome is characterized by a corneal diameter below 10 mm in both eyes and inherited cataracts (Chen, et al. 2014). Cataracts are defined as blurred or dim vision, which can develop into a total cataract when the eyes reach visual maturity (Chen, et al. 2014). Myopia, iris coloboma, and sclerocornea are also possible additional symptoms (Chen, et al. 2014). Corneal dystrophy and Corneal Opacity are other common eye disorders that may occur with Cataract Microcornea Syndrome (Klintworth, et al., 2009).

Before the discovery of the impact of the ABCA3 gene, 9 genes led to Cataract Microcornea Syndrome: some being CRYAA, CRYBA4, CRYBB1, CRYBB2, CRYGC, and CRYGD (a.Hansen, et al., 2019). This study also concluded that the gene ABCA3 is another cause of Cataract Microcornea Syndrome (Chen, et al. 2014). In two Chinese families, the missense mutations were c.115C>G, c.277G>A, c.4393G>A, and c.2408C>T (Chen, et al. 2014). There were also two splice-site mutations, c.4053+2T>C, c.2765-1G>T later identified (Chen, et al. 2014). The ABCA3 gene spans over 80,000 nucleotides and can be transcribed into 6500 base pair mRNA (Mulugeta, et al.,2002). This synthesizes a 1704 amino acid protein (Mulugeta, et al., 2002). ABCA3 is also predicted to be a glycoprotein that could hydrolyze ATP to provide energy for substrate transport involved in eye development (Chen, et al. 2014). Therefore, a mutation in ABCA3 might cause an impact on eye development.

3. Discussion

3.2 CRISPR

HDR can only be used on multiplying cells. Cells like the liver, neuron, and muscle which do not multiply, cannot perform HDR double-strand breaks (Uddin, et al., 2020). While HDR is the most preferred method, it is less prone to mistakes (Uddin, et al., 2020). Double strand breaks enable editing for deletions, insertions, and substitutions (Uddin, et al., 2020). The CRISPR/Cas9 system includes guide RNA (gRNA) which directs a special Cas9 nuclease to create double-strand breaks in the desired segment of DNA (Uddin, et al., 2020). CRISPR RNA (crRNA) and trans-encoded small crRNA (tracrRNA) are used to direct the Cas9 nuclease to target precise locations (Uddin, et al., 2020). This is called the tracrRNA:crRNA complex (Uddin, et al., 2020). The PAM sequence is located immediately after the target sequence and helps the Cas9 nuclease cut the sequence (Uddin, et al., 2020). The steps mentioned before are how a genetically modified strand of DNA is substituted in the DNA strand. The steps below are how bacteria integrate spacer DNA - a memory system of past infections.

There are three stages in which the CRISPR-Cas9 system is divided into Spacer integrations, processing of the primary transcript of pre-crRNA, and DNA/RNA interference (Makarova and Koonin, et al., 2015). In the first step, Cas 1 and Cas 2 enzymes insert the spacer DNA in between DNA repeats (Bolotin, et al., 2005). Both these proteins form a complex, where Cas 1 integrates the Spacer DNA, and Cas 2 performs a non-enzymic function (Makarova and Koonin, et al., 2015). In the second step, pre-crRNA is processed into guide-crRNA via RNA endonuclease complex or RNase III (Makarova and Koonin, et al., 2015). Then the now mature crRNA can be bound by Cas proteins resulting in either type I, type II, or type III effector complex (Chylinski, et al., 2014). In the third step, the effector complex is used to target DNA or RNA known as Cascade (CRISPR-associated complex for antiviral defense) (Wang, et al., 2018) (Figure 1).

3.2 X-Linked Congenital Night Blindness

CRISPR can be used to fix certain base pairs. 14 other mutations were found from another 16 patients tested who had incomplete CSNB. With a total of 20 mutations identified, 14 were predicted to cause protein transaction, and 6 were predicted to cause amino acid substitution/deletion (Boycott, et, al., 2001). One example of such is truncating mutation, located in the 3' splice site intron 40 (Boycott, et, al., 2001). One of the mutations was an A to G transition which was predicted to cause the loss of a splice site mutation, which resulted in the loss of exon 41 (Boycott, et, al., 2001). This then resulted in a premature stop codon in exon 42 and premature truncation of the a1F protein (Boycott, et, al., 2001). The A-G transition would be one base pair in length, so CRISPR technology could be used to fix this mutation (Boycott, et, al., 2001). CRISPR could also be used to fix the 6 amino acid substitutions and deletions (Boycott, et, al., 2001). It would depend on each protein mutations for the other 14 mutations identified. Since, however, there were 20 mutations identified in 16 patients, it is reasonable to conclude that there could be more mutations that exist but that have not been documented. Since CRISPR can be used to correct 6 out of the 20 mutations because they are substitutions, it is worth brainstorming on the idea of experimenting with CRISPR on the CACNA1F gene.

3.3 Snowflake Vitreoretinal Degeneration

The KCNJ13 gene mutation was a result of the CGG codon changing to a UGG codon (Hejtmancik, et al., 2008). CRISPR technology can be used to fix single mutations, so the UGG codon could be fixed. This results in the Arginine amino acid is changed into a Tryptophan amino acid. Tryptophan is a stop codon that results in a nonsense mutation leading to the production of the R162W mutant. The R162W mutant modifies the channel selectivity of the Kir7.1 protein making it permeable to Nab ions. CRISPR technology would be beneficial as it would create significant results by correcting the sequence for the R162W mutant. Therefore, CRISPR technology could be used to treat Snowflake Vitreoretinal Degeneration.

3.4 Cataract Microcornea Syndrome

The gene ABCA3 has multiple missense mutations and two splice-site mutations. The missense mutations are c.115C>G, c.277G>A, c.4393G>A, and c.2408C>T, and the splice-site mutations being c.4053+2T>C, c.2765-1G>T.

CRISPR could be used to treat these genes; however, it is important to note that not all genes for Cataract Microcornea Syndrome are known (Chen, et al. 2014). Therefore, complete treatment of all symptoms might not be accomplished. Another reason to keep in mind is that CRISPR might not be a feasible treatment for every patient. Depending on their ethnicity, patients will have different mutations.

4. Conclusions

While it is important to understand the concerns, not every disease can have significant effects with using CRISPR. Out of all three diseases examined in this paper, Snowflake Vitreoretinal Degeneration with the KCNJ13 gene has the highest potential for creating a feasible treatment for this disorder using CRISPR compared to X-Linked Congenital Night Blindness and Cataract Microcornea Syndrome. Since, however, each patient has a different set of mutations, the predictability of the usefulness of CRISPR is variable. CRISPR technology has its benefits and its concerns. CRISPR is relatively cheap, easy to use, and one of the most popular forms of gene editing. Some limitations are set, including the length of nucleotide bases, and having off-target effects as mentioned above (Uddin, et al., 2020). Ethical questions that may be raised mainly stem from embryonic editing compared to somatic editing because it happens after the offspring is born (Uddin, et al., 2020). While consent can be taken from the individual, it cannot be taken in Embryonic editing (Uddin, et al., 2020). Embryonic editing can also cause permanent problems for future generations as well as result in premature high off-target rates or death from the procedure (Uddin, et al., 2020). Although CRISPR can make progress in all three eye diseases (X-Linked Congenital Night Blindness, Snowflake Vitreoretinal Degeneration, and Cataract Microcornea Syndrome), future studies should heavily consider the side-effects and ethical concerns.

Acknowledgment

I would like to thank my mentor, Kevin DAmico, for his support and guidance.

References

Bech-Hansen, N. T., et al. (1998a). Loss-of-function mutations in a calcium-channel alpha1-subunit gene in Xp11.23 cause incomplete X-linked congenital stationary night blindness. *Nature genetics*, *19*(3), 264–267. https://doi.org/10.1038/947

Bech-Hansen, N. T., et al. (1998b). Localization of a gene for incomplete X-linked congenital stationary night blindness to the interval between DXS6849 and DXS8023 in Xp11.23. *Human genetics*, *103*(2), 124–130. https://doi.org/10.1007/s004390050794

Bech-Hansen, N. T., & Pearce, W. G. (1993). Manifestations of X-linked congenital stationary night blindness in three daughters of an affected male: demonstration of homozygosity. *American journal of human genetics*, *52*(1), 71–77.

Bolotin, A., et al. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* (*Reading, England*), 151(Pt 8), 2551–2561. https://doi.org/10.1099/mic.0.28048-0

Boycott, K. M., et, al. (2001). A summary of 20 CACNA1F mutations identified in 36 families with incomplete X-linked congenital stationary night blindness, and characterization of splice variants. Human genetics, 108(2), 91–97. https://doi.org/10.1007/s004390100461

Chen, P., et al. (2014). Mutations in the ABCA3 gene are associated with cataract-microcornea syndrome. *Investigative ophthalmology & visual science*, 55(12), 8031–8043. https://doi.org/10.1167/iovs.14-14098

Chylinski, K., et al. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic acids research*, *42*(10), 6091–6105. https://doi.org/10.1093/nar/gku241

Cong, L., & Zhang, F. (2015). Genome engineering using CRISPR-Cas9 system. *Methods in molecular biology (Clifton, N.J.), 1239,* 197–217. https://doi.org/10.1007/978-1-4939-1862-1_10

Doudna, J. A., & Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)*, *346*(6213), 1258096. https://doi.org/10.1126/science.1258096

Hansen, L., et al. (2007). Novel MAF mutation in a family with congenital cataract-microcornea syndrome. *Molecular vision*, *13*, 2019–2022.

He, Y., Wang, et al. (2018). Cas1 and Cas2 From the Type II-C CRISPR-Cas System of *Riemerella anatipestifer* Are Required for Spacer Acquisition. *Frontiers in cellular and infection microbiology*, 8 (195). https://doi.org/10.3389/fcimb.2018.00195

Heigwer, F., et al. (2014). M. E-CRISP: fast CRISPR target site identification. *Nat Methods*, 11, 122–123. https://doi.org/10.1038/nmeth.2812

Hejtmancik, J. F., et al. (2008). Mutations in KCNJ13 cause autosomal-dominant snowflake vitreoretinal degeneration. *American journal of human genetics*, *82*(1), 174–180. https://doi.org/10.1016/j.ajhg.2007.08.002

Klintworth G. K. (2009). Corneal dystrophies. *Orphanet journal of rare diseases*, *4*(7). https://doi.org/10.1186/1750-1172-4-7

Lee, M. M., et al. (2003). Snowflake vitreoretinal degeneration: follow-up of the original family. *Ophthalmology*, *110*(12), 2418–2426. https://doi.org/10.1016/S0161-6420(03)00828-5

Makarova, K. S., & Koonin, E. V. (2015). Annotation and Classification of CRISPR-Cas Systems. *Methods in molecular biology (Clifton, N.J.), 1311,* 47–75. https://doi.org/10.1007/978-1-4939-2687-9_4

Mulugeta, S., et al. (2002). Identification of LBM180, a lamellar body limiting membrane protein of alveolar type II cells, as the ABC transporter protein ABCA3. *The Journal of biological chemistry*, 277(25), 22147–22155. https://doi.org/10.1074/jbc.M201812200

Newby, G. A., et al. (2021). Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature*, *595*(7866), 295–302. https://doi.org/10.1038/s41586-021-03609-w Plumer, B., Barclay, E., Belluz, J., Irfan, U. (2018, December 27). Vox. A simple guide to CRISPR, one of the biggest science stories of the decade. https://www.vox.com/2018/7/23/17594864/crispr-cas 9-gene-editing

Redman, M., et al. (2016). What is CRISPR/Cas9?. *Archives of disease in childhood. Education and practice edition*, *101*(4), 213–215. https://doi.org/10.1136/archdischild-2016-310459

Uddin, F., Rudin, C. M., & Sen, T. (2020). CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future. *Frontiers in oncology*, *10*, 1387. https://doi.org/10.3389/fonc.2020.01387

Zhang, X. H., et al. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular therapy. Nucleic acids*, *4*(11), e264. https://doi.org/10.1038/mtna.2015.37



Impact of Demographics on Shifts in the Electorate since 2000

Andrew Wang^{1*}

¹ Panther Creek High School, Cary, NC USA

Received September 30, 2021; Revised February 3, 2022; Accepted, February 16, 2022

Abstract

Political analysts have long studied demographic shifts and their relation to the electoral change in the United States. This study conducts a qualitative analysis of 19 case-study counties to determine the limitations and accuracy of using race and age as indicators of presidential election margin shift from 2000 to 2020. Post analysis, four types were created to explain the overarching shift themes: normal, urban, Rust Belt, and group status. The normal type followed initial assumptions on age and race, that older populations voted more Republican and more diverse populations voted more Democratic. The urban-type highlighted the larger magnitude of Democratic shift seen in diverse urban areas. The Rust Belt type characterized the majority-white and aging populations of the Rust Belt undergoing Republican shifts. Finally, the group status type showed the limitation of racial diversification being positive for Democratic candidates. The social psychology concept of group status is applied to show a backlash effect of diversifying counties, whereby counties undergoing significant racial shifts become more Republican.

Keywords: Demographics, Race, Age, Political Science, Electoral shift

1. Introduction

The United States has undergone profound demographic change in the last two decades. Through this significant shift, it is vital to maintain an understanding of the relationship between demographic change and electoral shift. Electoral changes have been subject to much analysis between past presidential elections because of the insight they provide on the United States population and what they show about culture and ideas across the United States (Bacon, 2018; Brown, et al., 2011; Frey, et al., 2020; Garnham, 2020; Hudak, 2016; Jacobson, 2016; Major and Major, 2018; Pew Research Center, 2015; The Economist, 2019). As the United States population continues to change, a few leading demographic factors have been proposed as reasonable indications of electoral shift.

The role of minorities has become more pronounced in recent decades, as the United States population has diversified and voting rates among minorities have increased. Traditionally, minority voters have tended to favor Democratic candidates, and since the year 2000, the Democratic vote percentage among minority voters has risen (Hudak, 2016). In the 2000 presidential election, around 57% of the Asian Americans and Pacific Islanders (AAPI) electorate voted Democratic, rising to 73% by the presidential election of 2012 (Hudak, 2016). This trend is similarly reflected for Black voters, shifting from 90% to 93% Democratic between 2000 and 2012, and for Hispanic voters, shifting from 65% to 73% Democratic over the same period (Hudak, 2016). Though minority populations have all

Advisor: Amanda McCubbins arm4zz@virginia.edu

^{1.1} Background Literature

increased over the last few decades, the Hispanic population, in particular, has seen significant growth projected to double by 2050 (Budiman and Igielnik, 2020). This increase in Hispanic voters has been forecasted to make Southwestern states more competitive for Democrats (The Economist, 2015). For instance, many politicians predict Texas will become increasingly competitive for Democrats because of this diversification (The Economist, 2019). However, this increase in diversity has been observed to create an opposite effect too. This was highlighted in 2016, where Republican presidential candidate Donald Trump targeted older White conservatives, rallying Republican voters with his anti-immigration rhetoric (Jacobson, 2016). Though this pushed away young and nonwhite voters, more White voters voted Republican, a "backlash" effect on the country's increasing diversity (Jacobson, 2016; Maggio, 2020). That being said, increasing diversity has tended to be positive for Democratic candidates. Based on past trends, as the role and size of minority populations increases, votes for Democratic candidates are also expected to increase.

Age has become another prevalent factor in analyzing voter shift, as political views shift between generations. Democrats have a 24% margin with Millennials, compared to only a 5% margin with Generation X. For generations older than Generation X, such as Baby Boomers, Republicans win the margin (Frey, et al., 2020). Though the Baby Boomers were the largest generation, the role of younger generations is projected to increase with time. Millennials and Generation Z are predicted to compose over 50% of the electorate by 2032 (Frey, et al., 2020). Though there is some conservatizing effect as voters age, possibly due to life milestones such as marriage and homeownership, the consistency of these shifts is uncertain (Frey, et al., 2020). For instance, there has not been a significant conservative shift among older Millennials (Frey, et al., 2020). As citizens get older, this constancy of political views may indicate increasing affective polarization, which limits movement across party lines (Iyengar, et al., 2012; Iyengar, et al., 2019). Though the complexities of generational shift remain to be seen at large, for this study, the general political leanings of current generations are appropriate.

Finally, the urban-rural divide has become an increasingly important factor in considering demographic and electoral shifts, as urban areas tend to vote Democratic while rural areas vote Republican. This divide is also widening. In 2012, Democrats had a 5% margin in urban areas, rising to 17% by 2018. Opposingly, the Republican margin grew from 29% to 38% in rural areas in the same period (Milligan, 2019). Though this idea is reflected in news and literature, there is no consensus on why urban areas are more Democratic (Maxwell, 2019; Niskanen Center, 2019). Some leading reasons include educational differences, higher diversity seen in urban areas, socioeconomic differences between urban and rural areas, and simply the policy stances of each party regarding spending (Brown et al., 2018). The significant socioeconomic differences between rural and urban areas, such as education and wealth, are likely one of the leading factors for Democratic urban areas. With more tech-centered and globalized urban areas, people in urban areas are more likely to have a favorable view of immigration and economic shifts (DelReal and Clement, 2017). According to the Pew Research Center, Democrats lead by 22% among adults with post-graduate degrees and 7% among adults with college degrees (Pew Research Center, 2015). Additionally, there is some resentment among rural communities of the "liberal elite," the idea that those living in cities look down on residents of rural areas. This idea often creates the perception that government and government services are inherently anti-rural, and therefore many rural communities are against larger governments (Cramer, 2016; DelReal and Clement, 2017; Hudson, 2019). As cities across America become more urban, current voting trends suggest that traditionally Republican-dominated states will become more competitive for Democrats (Bacon, 2018; Garnham, 2020).

1.2 Study Scope

This study conducts a qualitative analysis of 19 case-study counties over the last 20 years using yearly demographic data from the United States Census Bureau and presidential election data from MIT's election data lab (U.S. Census Bureau, Population Division, 2012; MIT Election Data and Science Lab, 2018). All counties' demographic factors of race and age are analyzed for their indicative ability of electoral shift. In counties where electoral change did not match hypothesized demographic trends, other factors, such as the local economy (unemployment, significant industries, etc.) and local politics (percent citizenry, campaign spending, etc.), were considered.

2. Materials and Methods

This study compiled census data from various United States Census Bureau datasets. In finding county distributions for age and race over 20 years, multiple sources had to be assembled and cleaned for a consistent final dataset. From 2010 to 2020, a dataset called "Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2010 to July 1, 2019" was used, providing both age and race data for the period (U.S. Census Bureau, Population Division, 2020). Two other datasets, "Intercensal Estimates of the Resident Population by Sex, Race, and Hispanic Origin for Counties: April 1, 2000 to July 1, 2010" and "Intercensal Estimates of the Resident Population by Five-Year Age Groups and Sex for Counties: April 1, 2000 to July 1, 2010" were combined to provide age and race data for the period from 2000-2010 (U.S. Census Bureau, Population Division, 2012; U.S. Census Bureau, Population Division, 2012). Finally, a data set called "County Presidential Election Returns 2000-2020" was retrieved from the MIT Election Data and Science Lab to model election margins by count over the last 20 years (MIT Election Data and

Science Lab, 2018).

2.1 Data Preparation and Cleaning

This study used six groupings for race: White, Black, Native American, Asian American, Pacific Islander, and Hispanic populations, reported as a percentage of the total county population and monitored over 20 years. The final dataset for race was 18,828 rows by 24 columns, accounting for about 3,138 of the 3,142 counties in the United States. For age, this study measured residents from ages 15 and up, with age demographics in 10-year increments. Furthermore, all residents above 75 were grouped into one category, called "75+". This number was converted into a percentage for better comparisons between counties. However, because only residents ages 15 and up were recorded, the percentage composition of the counties tends to fluctuate more because it does not have residents ages 0-14. The final dataset for age was 21,946 rows by 24 columns, covering roughly the same number of counties as the age dataset.

Following data cleaning, 19 counties were selected for a qualitative analysis and case study. Counties were chosen to give the widest variety of electoral and demographic shifts, representing anomalies where demographic trends do not match electoral shifts. Some factors used in county selection included significant margin shifts, frequent party flipping, areas of sizable demographic change, and counties that went against the state's norm. A table with the reasoning for selecting each county is below:

State	County	Selection reason
DELAWARE	KENT	Frequent electoral swings
FLORIDA	PINELLAS	Frequent electoral swings
GEORGIA	FULTON	Location of Atlanta
KENTUCKY	ELLIOTT	Traditionally Democratic county which flipped hard in 2016
LOUISIANA	ASSUMPTION	Democratic county which went Republican in 2008
LOUISIANA	POINTE COUPEE	Democratic county which went Republican in 2008
MAINE	KENNEBEC	Only county in Maine which flipped back to Democratic in 2020 after flipping to Republican in 2016
MAINE	PISCATAQUIS	In a traditionally Democratic state this county has voted Republican in every election

Table 1. Selection reasoning for case study counties.

MICHIGAN	LAKE	Traditionally Democratic county which flipped Republican in 2016, voted Democratic in 2000 and 2004 in a state which largely voted Republican in those elections
MICHIGAN	SAGINAW	A Democratic county which flipped in 2016 and was the only county in Michigan to flip back to Democratic in 2020.
MINNESOTA	BIG STONE	A county with a large Republican shift that voted Democratic in 2000 and 2004
MINNESOTA	NOBLES	Notable Republican party which has shifted largely due to demographic influences, mentioned in media as well
MINNESOTA	PINE	County with a large Republican shift in 2016 but previously slim margins between Democratic and Republican
MINNESOTA	RAMSEY	A county where the margin has been moving Democratic though most counties have been shifting Republican. Anomalous county
MISSISSIPPI	JEFFERSON	County with huge margins for Democrats, one of the largest in the country
PENNSYLVANI A	ERIE	A notable 2016 county and a bellwether county
TEXAS	BEXAR	Location of San Antonio, the county has seen large demographic shift
TEXAS	DALLAS	Location of Dallas, another county with large demographic shift
WASHINGTON	KING	Location of Seattle, urban and rapidly changing area

Appendix A is a map of the United States, with case study counties in red. Additionally, Appendix B holds the presidential election margins over the 20 years for selected counties. Frequent electoral swings are defined as a county having switched parties at least 3 times in presidential elections between 2000 and 2020.

2.2 Control Variables

This study used the initial demographic assumptions that racial diversification would favor Democratic candidates and a greying population would favor Republican candidates. These hypotheses are proven correct through control counties for both demographic factors. The racial hypothesis that diversification favors Democratic candidates is seen in Ramsey County, MN, Dallas County, TX, and King County, WA. These counties all had very slight age shifts (less than 2%) with sizeable racial diversification (greater than 14%) and matched the hypothesis of increased Democratic shift. The hypothesis that greying populations favor Republican candidates is seen in many Midwestern counties, such as Piscataguis County, ME, Lake County, MI, and Big Stone and Pine counties, MN, with low racial diversification (less than 4%) with a large age shift (greater than 7%).

Following selection, counties were evaluated qualitatively for factors that would explain their electoral shift. Counties were analyzed with the initial assumptions that a decrease in the percentage of White people in a county would support Democratic candidates, and a greying county, where the county has an increasing proportion of older voters (greater than 55 years), would support Republican candidates. Based on whether or not these hypotheses tracked for the given county and electoral shift, further analysis was conducted to determine additional influential factors.

3. Results

Following data consolidation and analysis, a final dataset was produced. Below is a table indicating the race and age classification for all 19 counties studied that summarizes age and race shift into one of three categories: slight, moderate, and substantial.

Table 2.	Race	and age	change	for all	selected	counties.

State	County	Racial diversification*	Population aging*
DELAWARE	KENT	Moderate increase	Slight aging

FLORIDA	PINELLAS	Moderate increase	Moderate aging
GEORGIA	FULTON	Moderate increase	Slight aging
KENTUCKY	ELLIOTT	Slight increase	Moderate aging
LOUISIANA	ASSUMPTION	Slight increase	Substantial aging
LOUISIANA	POINTE COUPEE	Slight increase	Substantial aging
MAINE	KENNEBEC	Slight increase	Moderate aging
MAINE	PISCATAQUIS	Slight increase	Substantial aging
MICHIGAN	LAKE	Slight decrease	Substantial aging
MICHIGAN	SAGINAW	Slight increase	Moderate aging
MINNESOTA	BIG STONE	Slight increase	Moderate aging
MINNESOTA	NOBLES	Substantial increase	Slight aging
MINNESOTA	PINE	Slight increase	Substantial aging
MINNESOTA	RAMSEY	Moderate increase	Moderate aging
MISSISSIPPI	JEFFERSON	Slight increase	Substantial aging
PENNSYLVANIA	ERIE	Slight increase	Moderate aging
TEXAS	BEXAR	Moderate increase	Slight aging
TEXAS	DALLAS	Substantial increase	Slight aging
WASHINGTON	KING	Substantial increase	Slight aging

*Definitions for each kind of shift classification are in Appendix C.

4. Discussion

Through the qualitative analysis, counties were sorted into four primary types based on how age and race could explain their electoral shift and overall themes: normal, urban, Rust Belt, and group status theory. However, outside of demographic and societal factors, presidential candidates, such as Obama and Trump, also significantly affect partisan shifts. Multiple counties fall under various types because there is crossover in the factors which influenced their change. What follows is an explanation of each type, the counties that constitute it, and a discussion of the type.



Fig 1. Four Primary Types of Counties

4.1 Normal Type

The normal type encompasses all counties that matched the initial race and age shift hypothesis. These are the counties that aligned with predictions about their voter shift. Counties in this group include Big Stone, Pine, and Ramsey counties in MN, Lake and Saginaw counties in MI, Piscataquis County, ME, Erie County, PA, and King County, WA. Essentially, the magnitude and direction of electoral shifts could be predictable for presidential elections over the last 20 years. For instance, Piscataquis County, Maine, had a 15% total Republican shift over the previous twenty years, going from an 11% Republican margin to a 26% Republican margin in 2020. This shift aligns with the county's slight racial change, where the White population decreased by 3%, from 98% to 95%, between 2000 and 2020.

Furthermore, the Republican shift can be further indicated by their greying population, a 10% increase in the proportion of the population older than 55 years. Therefore, the low racial diversification and greying population match up with a forecasted Republican shift. Another example is Ramsey County in Minnesota, which saw sizeable racial diversification but a slight shift in age structure. Ramsey County had a 14% decrease in the White population and a negligible (less than 1%) change in its share of voters older than 55 years. This corresponds to a 25% Democratic shift in the last 20 years, matching demographic predictions.

4.2 Urban Type

The urban type describes counties with a large urban population, typically counties holding large cities. The threshold for being an urban county was possessing a population of one million or more and having a large city. These counties saw disproportionately large Democratic shifts compared to their race and age changes, typically approximately a 20% increase in Democratic voter share over the last 20 years. Unusually, the counties in the urban type also shifted Democratic in 2016, though most counties in the study and across the United States at large moved Republican (Parlapiano and Lai, 2016). This can be attributed to their sizeable nonwhite population. Urbanity seems to exacerbate Democratic shifts. The counties in this group are Bexar and Dallas counties in TX, Fulton County, GA, and King County, WA. These counties had populations close to or above one million residents, with large nonwhite populations (between 42-73% of the population is nonwhite in each county).

A notable example is Dallas County, Texas, which contains the city of Dallas, housing 1.3 million people. Dallas county has a sizable Hispanic population which has increased since 2000. In 2000, Dallas had a 30% Hispanic population and a 44% white population, leaning 8% Republican. Since then, Dallas County has seen a significant racial shift, with a 16% decrease in the White population. In 2020, Hispanics were the largest ethnic group in Dallas County, composing 41% of the county. This racial shift correlates to Dallas's electoral shift, moving 40% in favor of the Democrats since 2000. The county flipped parties in the presidential election of 2008, moving 17% in favor of Democrats. This shift can be dually attributed to the Democratic candidate Barack Obama and the financial crisis in 2008 and Hurricane Ike and Hurricane Gustav. Finally, Democratic campaign spending in 2008 dwarfed

Republican campaign spending. Obama's campaign spent 9.9 million in Texas, while McCain's only spent \$33,000. This combination of factors likely contributed to the significant shift seen between 2004 and 2008.

4.3 Rust Belt Type

The Rust Belt type demonstrates the Republican shift seen by many counties in the Rust Belt and Northeast. This type is more of a classification based on common geography than a demographic grouping. Counties in this group include Big Stone, Pine, and Nobles counties in MN, Lake and Saginaw counties in MI, Piscataquis county, ME, and Erie County, PA. These counties tend to follow a roughly similar electoral shift between elections: a Democratic win in 2000, followed by a smaller Democratic margin in 2004, a more considerable Democratic margin in 2008, followed by a Republican shift in 2012, and finally, a sizeable Republican shift (generally around 20%) in 2016, which saw all counties flip to Republican. The Republican shift of 2016 can almost certainly be attributed to Trump, with his rhetoric and campaign promise to revitalize American manufacturing. Trump is thought to have driven the significant margin shift in 2016 and through the 2020 presidential election. Between 2016 and 2020, there was little resultant change, and counties had shifts of around 2.5%.

Furthermore, the counties of the Rust Belt type tend to be very similar demographically. Racially, they had slight diversification (around 3-5%) and remained overwhelmingly White. Age-wise, these counties had older populations and saw a sizable (approximately 10%) increase in residents older than 55 years. However, the growing percentage of older people could reflect two things: the population at large is aging (this would be a standard demographic transition), or the population is declining, and young people are moving out. For most counties of this type, it seems like a general demographic transition. The birth rate is falling, and the population is growing older, as the total population between censuses remains relatively consistent.

The best example of the Rust Belt type is Lake County, Michigan. A county of about 12,000 people, Lake County, went from a 13% Democratic margin in 2000 to a 26% Republican margin in 2020. Its demographic shifts are precisely in line with the other counties of the Rust Belt type. The proportion of White people in the county increased, growing from 83% to 86% between 2000 and 2020. Age-wise, the county saw a 10% increase in residents older than 55 years, with a 13% decrease in residents between 15 and 34. The county's whitening and aging population indicate a Republican shift, proven true in the 2016 presidential election, where it moved 28% in favor of the Republicans. This enormous shift highlights Trump's impact on similar predominantly White Rust Belt counties. His campaign promises to "Make America Great Again" appealed greatly to the economically declining Rust Belt, as manufacturing had been offshored mainly due to globalization. Furthermore, Trump's conservative stance on immigration would have appealed greatly to the dominant White population of the county (Major and Major. 2018).

4.4 Group Status Type

The group status type describes a set of counties that do not follow the demographic hypothesis that racial diversification creates a Democratic shift. These are counties with large minority populations (averaging 20-30% of the county's residents) and large Republican margins. Counties in this group include Elliott County, KY, Assumption and Pointe Coupee Parishes, LA, and Nobles County, MN. A possible cause for this reverse effect includes voter policy and citizenship rates, as the minority population may not be naturalized or able to vote. Another potential reason is rooted in social and group psychology. Theories like the group status theory and the racial threat hypothesis give insights into this shift; increases in a minority population are perceived as threatening to the majority population because they threaten to disrupt the power structure and status quo (Todak and Wang, 2016; Major and Major, 2018). Many White Americans view race relations as a zero-sum game, in which gains for a minority group means a loss for the majority group(Wilkins and Kaiser, 2014). Therefore, social psychology theories about identity and intergroup relations predict that White Americans will become more conservative if they feel increasing minority populations represent a threat to White American resources and values (Tajfel and Turner, 2004; Todak and Wang, 2016). This ties in heavily with presidential candidates such as Trump, whose America first stance appealed to many White Americans. In these counties, the increased diversification has worked against the Democrats, rallying White residents to vote Republican in a conscious or unconscious attempt to maintain their power structure in the community.

One example of this kind of electoral shift is seen in Assumption Parish, Louisiana. Between 2000 and 2020, there was a 1% change in the White population, moving from 66% to 65% white. However, the county also has a large Black population, composing 30% of the county's residents. Age-wise, there was a 9% increase in residents older than 55 years. With a substantial Black population, demographic assumptions propose initial а Democratic lean in the county. However, it actually saw a Republican shift of 38%, going from an 8% margin Democratic in 2000 to a 30% Republican margin in 2020. The Republican party is overwhelmingly dominant in campaign spending, spending \$169,000 to the Democrat's \$10,000. Votes cast per election showed the total number of votes cast remained roughly the same. As opposed to new voters starting to vote, the same voters changed which party they cast their votes for. An additional dimension to consider is Louisiana's history of restrictive voter ID laws, as voter ID laws have traditionally been used to suppress the votes of Black citizens. In the past, Louisiana parishes used poll taxes, property and literacy requirements, and an understanding clause to drop the registration of Black voters. Enforcement of these laws was left to the discretion of local officials, which resulted in unequal application of the law. Therefore, though Assumption parish has a large Black population, the voting rate of that population is meager due to voter suppression. However, the large Black population could have influenced a perception of the racial threat hypothesis. This is supported by the presidential election of 2008, where Barack Obama was the first Black presidential candidate. Though Assumption Parish voted 5% Democratic in 2004, they saw a 16%

Republican shift in 2008, running contrary to the sizeable Democratic shift seen through most of the United States and possibly indicating a perceived status threat in which residents saw Obama's presidency as an irrevocable shift in status quo. In conclusion, the Republican shift in Assumption parish, despite its large minority population, could be attributed to two factors: a perceived group status threat and the irregularly low turnout of minority voters.

Though an initial assumption of demographics was that aging populations would benefit Republican candidates, this is not exclusively true. Though White Americans traditionally have become more conservative with age, this trend does not match when applied to other ethnic groups. Proof of this is seen in Jefferson County, Mississippi. Jefferson county is 85% Black and has seen a 7% increase in residents older than 55 years. However, the county has simultaneously continuously shifted Democratic over the last 20 years and had a 72% Democratic margin in 2020, one of the highest in the country. Instead, the trend with age can be revised to the creation of a "stickiness" to political identity, that as voters, in general, become older, their beliefs are more set into stone.

5. Conclusion

This paper applied a qualitative analysis to a case study set of 19 counties to evaluate how race and age change over the last 20 years could explain shifts in presidential election margins. Initial demographic hypotheses that diversification would increase Democratic votes and aging populations would increase Republican votes were applied to these counties. Demographic and electoral shift was tracked through yearly race and age data from the U.S. Census Bureau and presidential election data from MIT's election data lab. Though the electoral shift for many counties was consistent with the demographic hypotheses, limitations were present in both. There was often a "backlash" in racial diversification, where counties would see significant Republican shifts even with a growing minority population, as White residents would vote Republican against the diversification. This was

particularly true in counties with inactive or non-citizen residents because there wouldn't be many votes for Democratic candidates. This pattern indicates a fascinating social trend, as diversification has usually been seen as good for Democrats, but this backlash factor may reverse its effects in certain circumstances. A greying population being more Republican mainly was only accurate for dominantly White populations. Greying minority populations did not see the same electoral shifts as greying White populations.

The results of this study raise interesting questions about how demographic changes may indicate future electoral shifts. While there is some benefit to Democratic candidates from the increasing minority populations, the backlash of White residents seen in certain counties creates uncertainty around increasing diversification being good for Democrats, particularly given the growing group identities and polarization of today's political atmosphere. Future research should consider what circumstances and factors play into this backlash effect. The application of demographics to understanding the electorate shift continues to be an essential topic to understand, particularly as the United States continues to diversify, and it will be salient to campaigns and voter outreach efforts in future elections.

References

Bacon, P. (2018, October 18). Why Are Democrats Looking So Strong In The Midwest? FiveThirtyEight. https://fivethirtyeight.com/features/why-are-democrat s-looking-so-strong-in-the-midwest/

Brown, A., et. al. (2018, May 22). Demographic and economic trends in urban, suburban and rural communities. Pew Research Center. https://www.pewresearch.org/social-trends/2018/05/2 2/demographic-and-economic-trends-in-urban-suburb an-and-rural-communities/

Brown, M., Knopp, L., & Morrill, R. (2011). Anomalies in Red and Blue II: Towards an understanding of the roles of setting, values, and demography in the 2004 and 2008 U.S. presidential elections. Political Geography, 30 (3), 153-168. https://doi.org/10.1016/j.polgeo.2011.03.006. Budiman, A., & Igielnik, R. (2020, September 23). The Changing Racial and Ethnic Composition of the U.S. Electorate. Pew Research Center. https://www.pewresearch.org/2020/09/23/the-changin g-racial-and-ethnic-composition-of-the-u-s-electorate /

Cramer, K. J. (2016). The Politics of Resentment. The University of Chicago Press.

DelReal, J. A., & Clement, S. (2017, June 17). Rural divide. The Washington Post. https://www.washingtonpost.com/graphics/2017/natio nal/rural-america/

Frey, W. H., Griffin, R., & Teixeria, R. (2020, October 19). America's Electoral Future. Center for American Progress. https://www.americanprogress.org/issues/politics-and

-elections/reports/2020/10/19/491870/americas-electo ral-future-3/

Garnham, J. P. (2020, November 11). Democrats didn't get a blue wave, but some of the fastest-growing suburbs in Texas are still moving to the left. The Texas Tribune. https://www.texastribune.org/2020/11/11/texas-demo crats-republicans-suburb/

Hudak, J., & Stenglein, C. (2016, September 13). How demographic changes are transforming U.S. elections. Brookings.

https://www.brookings.edu/blog/fixgov/2016/09/13/h ow-demographic-changes-are-transforming-u-s-electi ons/

Hudson, W. L. (2019, June 6). Rural Resentment and 2020. Americans for Humanity. https://americansforhumanity.net/blog-1/2019/6/6/rur al-resentment-and-2020

Iyengar, S., et. al. (2019). The Origins and Consequences of Affective Polarization in the United States. Annual Review of Political Science, 22 (1), 129-146.

https://doi.org/10.1146/annurev-polisci-051117-0730 34

Iyengar, S., Sood, G., & Lekles, Y. (2012). Affect, Not Ideology: A Social Identity Perspective on Polarization. Public Opinion Quarterly, 76 (3), 405–431. doi:10.1093/poq/nfs038 Jacobson, G. C. (2016). The Obama Legacy and the Future of Partisan Conflict: Demographic Change and Generational Imprinting. The Annals of the American Academy of Political and Social Science, 667, 72–91. http://www.jstor.org/stable/24756144

Maggio, C. (2020). Demographic change and the 2016 presidential election. Social Science Research, 95, . https://doi.org/10.1016/j.ssresearch.2020.102459

Major, B., Blodorn, A., & Major Blascovich, G. (2018). The threat of increasing diversity: Why many White Americans support Trump in the 2016 presidential election. Group Processes & Intergroup Relations, 21(6), 931–940. https://doi.org/10.1177/1368430216677304

Maxwell, R. (2019, March 5). Why are urban and rural areas so politically divided? The Washington Post.

https://www.washingtonpost.com/politics/2019/03/05 /why-are-urban-rural-areas-so-politically-divided/

Milligan, S. (2019, May 14). Demographic Shifts, Political Changes. U.S. News and World Report. https://www.usnews.com/news/best-states/articles/20 19-05-14/demographic-shifts-in-cities-and-states-brin g-political-changes-too

MIT Election Data and Science Lab. (2018, October 11). County Presidential Election Returns 2000-2020. Harvard Dataverse. https://dataverse.harvard.edu/dataset.xhtml?persistent

Id=doi:10.7910/DVN/VOQCHQ

Niskanen Center. (2019, July 17). Explaining the Urban-Rural Political Divide. Niskanen Center. https://www.niskanencenter.org/explaining-the-urban -rural-political-divide/

Parlapiano, A., & Lai, R. (2016, November 9). 42 States Shifted to the Right in 2016. New York Times. https://www.nytimes.com/interactive/2016/11/09/us/e lections/states-shift.html

Pew Research Center. (2015, April 7). A Deep Dive Into Party Affiliation. Pew Research Center. https://www.pewresearch.org/politics/2015/04/07/a-d eep-dive-into-party-affiliation/

Tajfel, H., & Turner, J. C. (2004). Political Psychology. Taylor & Francis Group. https://doi.org/10.4324/9780203505984-16 The Economist. (2015, Mar 14). How to fire up america; latinos in the united states. The Economist. https://www.economist.com/leaders/2015/03/12/howto-fire-up-america The Economist. (2019, October 5). The magenta mammoth. The Economist. https://www.economist.com/united-states/2019/10/03 /texas-wont-go-blue

Todak, N., & Wang, X. (2016, July 27). Racial Threat Hypothesis. Oxford Bibliographies. https://www.oxfordbibliographies.com/view/docume nt/obo-9780195396607/obo-9780195396607-0204.x ml

U.S. Census Bureau, Population Division. (June). CC-EST2019-ALLDATA-[ST-FIPS]: Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2010 to July 1, 2019. https://www2.census.gov/programs-surveys/popest/te chnical-documentation/file-layouts/2010-2019/cc-est 2019-alldata.pdf

U.S. Census Bureau, Population Division. (October). CO-EST00INT-AGESEX-5YR: Intercensal Estimates of the Resident Population by Five-Year Age Groups and Sex for Counties: April 1, 2000 to July 1, 2010.

https://www2.census.gov/programs-surveys/popest/te chnical-documentation/file-layouts/2000-2010/interce nsal/county/co-est00int-agesex-5yr.pdf

U.S. Census Bureau, Population Division. (October). CO-EST00INT-SEXRACEHISP: Intercensal Estimates of the Resident Population by Sex, Race, and Hispanic Origin for Counties: April 1, 2000 to July 1, 2010.

https://www2.census.gov/programs-surveys/popest/te chnical-documentation/file-layouts/2000-2010/interce nsal/county/co-est00int-sexracehisp.pdf

Wilkins, C. L., & Kaiser, C. R. (2014). Racial progress as threat to the status hierarchy: implications for perceptions of anti-White bias. Psychological science, 25(2), 439–446. https://doi.org/10.1177/0956797613508412



Chemical and Biological Approach using Mixed Oxidants for the Disinfection of Drinking Water Supplies

Keilah Om^{1*}

¹ Dominican Academy NYC, New York, NY USA

Received August 15, 2021; Revised January 19, 2022; Accepted, February 21, 2022

Abstract

The use of mixed oxidants for disinfection as a public health measure reduces the spread of countless diseases and is the future of public safety. Mixed oxidant solutions, which have typically formed electrochemically through the electrolysis of brine, have emerged as an effective and potentially significant method of disinfection of drinking water supplies. Many solutions have been tested and documented regarding inactivation rates relating to mixed oxidant solutions, including chlorine, ozone, hypochlorite, and several others. Advanced technologies are interesting, however many are still in the research state, while conventional technologies are the most used and far along. However, although some of the conventional technologies are not fully verified, a vast majority of them are very significant and vital methods when it comes to disinfection. Some of the many benefits of mixed oxidant solutions include a decreased generation of potentially harmful disinfection byproducts (DBPs), inhibition of biofilm formation within distribution systems, improved residual power, safer operating environments, and lower operating costs. Disinfection byproducts (DBPs) are chemical, organic, and/or inorganic substances that are commonly found in drinking water supplies and can form during chemical reactions of a disinfectant and drinking water. The use of mixed oxidant solutions in public safety provides a safer working environment and a safer supply of drinking water, while also providing a cheaper and more effective solution to promote public health and safety.

Keywords: disinfection, drinking water treatment, DBP

1. Introduction

Mixed oxidant solutions, typically formed electrochemically through the electrolysis of brine, have emerged as an effective, and a potentially significant method of the disinfection of drinking water supplies. Numerous studies have documented the efficacy of mixed oxidant solutions in the disinfection of microorganisms including Coliphage MS2, *Escherichia coli, Bacillus subtilis, Cryptosporidium parvum* oocysts, and *Clostridium*

* Corresponding Author keilah2006@gmail.com *perfringens* spores (Son *et al.*, 2003; Venczel *et al.*, 1997; Casteel *et al.*, 2000); however, the active disinfectant species present in mixed oxidant solutions have not been fully verified, necessitating the need for continued research in this realm. Species suspected in mixed oxidant solutions generated through the electrolysis of brine include, but may not be limited to, ozone, chlorine dioxide, hypochlorite, hypochlorous acid, chlorine, hydrogen peroxide, and OH- radicals. It has been suggested that chlorine is the effective component of electrochemically mixed

Advisor: Dr. Chulsung Kim chulsungkim@gmail.com

oxidant solutions responsible for disinfection (Clevenger et al., 2007). However, studies documenting increased inactivation rates observed with mixed oxidant disinfection compared to disinfection with hypochlorite alone, as well as the ability of mixed oxidant solutions to inactivate chlorine-resistant organisms like C. parvum, indicate that chlorine species are not the only mixed oxidant constituents in effect (Hamm 2002; Son et al., 2003; Venczel et al., 1997; Casteel et al., 2000). Based on reviewed literature, the inactivation rate, as well as the extent of the increase in disinfection efficacy by electrochemically and mechanically generated mixed oxidants compared to that of chlorine- based disinfectants, it is highly dependent on the type of microorganism used in the study. Accordingly, the assessment of the major oxidant species at work is also influenced by the biological indicator used. Lastly, the literature reviewed demonstrated that synergistic effects due to pH, temperature, and the combination of different oxidant types also play an important role in the disinfection efficacy of mixed oxidant solutions (Son et al., 2003).

2. pH and Temperature Effects on Mixed Oxidant Disinfection

In the first half of the study presented by (Son et al., 2003), E. coli and B. subtilis were used to investigate the disinfection efficacy of an electrochemically generated mixed oxidant solution compared to that of free available chlorine (FAC). The study also evaluated the effects of pH and temperature on inactivation rates of both microorganisms. Son et al. found that at a pH of 8.2, electrochemically generated mixed oxidant solutions were between 20 and 50 percent more efficient in inactivating E. coli and B. subtilis spores in comparison with disinfection with free available chlorine (FAC). The synergistic effect observed with the electrochemically mixed oxidants at a pH of 8.2 was not observed for E. coli or B. subtilis spores at lower pH levels of 5.7 or 7.1. As a point of clarification, the inactivation rates of E. coli and B. subtilis spores were the highest for both electrochemically mixed oxidants and FAC at lower pH levels; it was only at pH of 8.2 that a significant

difference was recorded between the inactivation rate observed using electrochemically mixed oxidants versus the inactivation rate observed using FAC as the form of disinfection.

The effects of pH on the inactivation efficacy of different disinfectants including free chlorine, ozone and chlorine dioxide are demonstrated in the single step application experiments run in a sequential disinfection study using B. subtilis spores. Results of the study showed that free chlorine was most efficient at lower pH levels (5.6 versus 8.2), whereas faster inactivation rates were observed at higher pH levels in single step treatment with ozone (Cho et al.'s 2006). The inactivation of B. subtilis spores by chlorine dioxide was not affected by a change in pH. These findings, specifically that of a single step application of ozone resulted in increased inactivation rates at a pH of 8.2, are consistent with the greater synergistic effect observed in the inactivation of E. coli and B. subtilis by mixed oxidants at a pH of 8.2 (Son et al., 2003). They also suggest that mixed oxidant solutions, composed of several oxidant types, have the ability to be effective at varying pH levels.

Table 1. Summary of the CT values (mg/L x min) for the 2-log inactivation of *E. coli* and *B. subtilis* observed by Son *et al.* (2003).

		Electrochemi cally mixed oxidants	FAC	Electroche mically mixed oxidants	FAC
Micro organism	pН	20° C	20° C	4° C	4° C
E. coli	5.7	3.5x10 ⁻²	3.2x10 ⁻²	-	-
	7.1	7.1x10 ⁻²	7.6x10 ⁻²	-	-
	8.2	0.13	0.18	0.24	0.36
B. subtilis	5.7	46	50	204	221
	7.1	97	94	-	-
	8.2	240	280	670	800

In examining the effects of temperature (20° C vs. 4° C) at pH 8.2, Son *et al.* (2003) found that both *E. coli* and *B. subtilis* spores exhibited increased inactivation rates at the higher temperature of 20° C with both methods of disinfection; however, the percent effectiveness of inactivation by mixed oxidants versus FAC was greater for both organisms at 4° C versus 20° C. Specifically, at 4° C, the electrochemically mixed oxidants were

approximately 50% and 19% more effective than FAC in inactivation of the 2-log inactivation of E. coli and B. subtilis, respectively; whereas at 20°C, the electrochemically mixed oxidants were approximately 35% and 17% more effective than FAC in inactivation of the-2 log of E. coli and B. subtilis. These results suggest that at pH 8.2, the inactivation rates of E. coli and B. subtilis may be less affected by a drop in temperature when treated with mixed oxidants versus FAC alone. In other words, mixed oxidant solutions may be a more effective disinfectant than FAC alone across a larger spectrum of temperatures.

3. Disinfection using Electrolytically Generated Mixed Oxidants

Son et al.'s (2003; 2005)) findings are somewhat in contrast to the study conducted by Clevenger et al., in which three strains of B. subtilis and the bacteriophage MS2 were used to compare the disinfection potency of three electrolytic generation systems of mixed oxidants as well as hypochlorite (at pH 7, 22° C). While all three mixed oxidant generating systems were slightly more effective than hypochlorite in the inactivation of the macrophage, all three systems were similarly effective as hypochlorite in the inactivation of the three strains of B. subtilis spores. MS2, which was inactivated much more rapidly and effectively by all four methods of disinfection than the B. subtilis spores, displayed no apparent change in chlorine residual for the four disinfectants (since MS2 phages are very sensitive to free chlorine, very little chlorine was required for their inactivation). A more pronounced, but uniform, decrease in chlorine concentrations occurred in the inactivation of B. subtilis spores for all four disinfectants, leading Clevenger et al. to conclude that the effective disinfection component in the three mixed oxidant generating systems was chlorine alone. (Clevenger et al., 2007)

The rapid inactivation of bacteriophage MS2 by an electrochemically generated mixed oxidant solution was also reported in Casteel *et al.* (2000), in which a 2 log reduction of the macrophage occurred in 30 seconds. The study also reported a relatively high percentage of the initial dose (2.4 mg/L) of

mixed oxidant solution remained throughout the experiment, with 1.6 mg/L mixed oxidant remaining after one hour contact time. This is generally consistent with the lack of a decrease in chlorine residual observed for MS2 after 30 minutes in Clevenger et al. (2007). Casteel et al. observed similar inactivation characteristics for E. coli treated with electrochemically mixed oxidants (2 log reduction within 30 seconds with a 2.4 mg/L dose of mixed oxidant solution, with 1.6 mg/L mixed oxidant remaining after one hour contact time). In contrast, chlorine-resistant C. parvum oocysts and C. perfringens spores exhibited much slower inactivation rates at higher doses of mixed oxidants. As indentified in the table below, C. parvum oocysts treated with 2 mg/L of electrochemically mixed oxidants experienced a 0.7 log reduction with 30 minutes contact time and a 1 log reduction after 240 minutes; C. parvum oocysts treated with 4 mg/L experienced a 2.1 log reduction at 30 minutes and a 2.9 log reduction after 240 minutes. The inactivation rate of C. perfringens was similar to that of C. parvum when treated with the same two doses of electrochemically mixed oxidants.

Table 2. Inactivation of *C. parvum* and *C. perfringens* by electrochemically mixed oxidants in oxygen demand-free water at pH 8 and at 25° C (Casteel *et al.*, 2000)

Micro organism	Dose of Mixed Oxidants	Contact Time	Log(10) inactivation
Carrier	2	30 min	0.7
C. parvum	2 mg/L	240 min	1
C	4 mg/L	30 min	2.1
C. parvum		240 min	2.9
C. norfringene	2 mg/L	30 min	0.7
C. perfringens		240 min	1
C. perfringens	4 mg/L	30 min	1
		240 min	1.5

Both *C. parvum* and *C. perfringens* exhibited declining rate inactivation kinetics by mixed oxidants. This study demonstrates that the rate and extent of inactivation by mixed oxidants is dependent on the type of microorganism, the initial does of mixed oxidants, the contact time, and in the case of *C. parvum* and *C. perfringens*, the mixed oxidant residual (or oxidant demand) (Casteel *et al.*, 2000).

4. Mixed Oxidants vs. Free Chlorine Disinfection

Prior to Casteel et al.'s study, Venczel et al. (1997) demonstrated the inactivation of C. parvum perfringens oocysts and С. spores by electrochemically mixed oxidants, in contrast to disinfection by free chlorine alone. The declining rate inactivation, or "retardant die-off," kinetics also observed by Casteel et al., were also observed by Venczel et al. with disinfection of C. parvum and C. perfringens with electrochemically mixed oxidants. As identified in Table 3, 5 mg/L of electrochemically generated mixed oxidants inactivated both C. parvum oocysts and C. perfringens spores in water at pH 7 and at 25° C, with greater than 2.3 log (>99.5%) inactivation in four hours. With an equivalent dose of free chlorine and four hour contact time period, there was essentially no inactivation of C. parvum and reduced inactivation of C. perfringens spores (1.5 log inactivation, or 97%). No inactivation of C. parvum spores treated with free chlorine was recorded after 24 hours of contact time, and only a 0.2 log increase in inactivation was recorded for C. perfringens beyond four hours contact time.

Table 3. Inactivation of *C. parvum* and *C. perfringens* by electrochemically mixed oxidants and FAC in oxygen demand-free water at pH 7 and at 25° C (Venczel *et al.* 1997). The inactivation rates in the table are the averages of four replicate experiments.

Micro organism	Disinfectant	Contact Time	Log(10) inactivation
	electro-	1 h	1.3
		4 h	>3.5
C. parvum	chemically	8 h	>4.3
	mixed oxidants	12 h	>4.6
		24 h	>3.8
	FAC	1 h	
		4 h	
C. parvum		8 h	no inactivation
		12 h	
		24 h	
		1 h	2.2
C. perfringens	electro- chemically mixed oxidants	3 h	2.6
		4 h	2.7
		8 h	3.3
		12 h	3.6
		24 h	3.7

C. perfringens	FAC	1 h	1
		3 h	1.3
		4 h	1.5
		8 h	1.6
		12 h	1.7
		24 h	1.7

Based on the results observed by Venczel et al. (1997) and Casteel et al. (2000), electrochemically mixed oxidants were effective in inactivating chlorine-resistant microorganisms such as C. parvum and C. perfringens (C. parvum being more resistant to chlorine than C. perfringens (Venczel et al., 1997)). As cited in Venczel et al. (1997), Korich et al. (1990) reported that when exposed to a free chlorine dose of 80 mg/L, a 1 log inactivation of C. parvum oocysts was observed after 90 minutes; however, such a high dose of chlorine would never be used in the disinfection of drinking water. Thus, for microorganisms such as C. parvum and C. perfringens, an alternative to disinfection by chlorine alone is needed, and mixed oxidant solutions appear to be a viable approach.

5. Synergistic Effects of Mixed Oxidants on Disinfection

In addition to the optimal pH and temperature conditions, another factor influencing the increased disinfection efficacy by mixed oxidant solutions may be synergism of certain oxidant species. Synergistic effects of oxidant species can be investigated through sequential disinfection experiments. In Livanage et al. (1997), the enhanced inactivation of C. parvum oocysts was attributed to synergistic effects of combining ozone and chlorine dioxide in sequential disinfection. The study reported that the expected inactivation by single oxidants was 0.8 log units for ozone and 1.4 log units for chlorine dioxide, for a total of 2.2 log inactivation; whereas, 3.4 log inactivation was documented when C. parvum oocysts were exposed to 0.8 mg/L of ozone for 4.4 minutes followed by a 2.0 mg/L chlorine dioxide treatment for 60 minutes. Thus, the sequential treatment of C. parvum oocysts with ozone and chlorine dioxide resulted in an additional 1.2 log units of inactivation due to synergism of the two disinfectants (Livanage et al., 1997). It is quite possible that a similar synergistic effect between ozone and chlorine dioxide occurs in electrochemically mixed oxidant solutions, helping make mixed oxidant solutions effective disinfectants in the inactivation of C. parvum oocysts and C. perfringens spores. Applying the results of other sequential disinfection studies, synergism of oxidant species may also be an important factor in the increased disinfection efficacy of mixed oxidants in the inactivation of other microorganisms, including B. subtilis (Cho et al., 2006) and E. coli (Yang et al., 2012; Beber de Souza and Daniel, 2011). Synergism observed in subsequent disinfection studies has been attributed to the activity of the disinfection species reacting with specific chemical groups of the bacterial cell wall (H. Son et al., 2005).

In Son et al. (2003; 2005), synergism of oxidant species appeared to be a contributing factor to the enhanced inactivation of B. subtilis spores by mechanically mixed oxidants versus disinfection by FAC alone. Four mechanically mixed oxidant solutions were prepared by adding small amounts of ozone (1 mg/L), chlorine dioxide (1.8 mg/L and 18 mg/L), hydrogen peroxide (9.2 mg/L) and chlorite (40 mg/L) into 200 mg/L FAC stock solution at pH 2.5. The chlorite/FAC solution was prepared to take into account expected chlorite production from the reaction between FAC and ozone. The pH and concentrations of each oxidant were chosen to replicate the reported concentrations of the oxidants produced (FAC), and claimed to be produced (ozone, chlorine dioxide, hydrogen peroxide and chlorite), by the MIOX electrolytic generation system. Son et al. (2003; 2005) explained that the concentration of chlorite used in the study (40 mg/L) was much greater than that reported in MIOX literature so that the product of FAC and chlorite, chlorine dioxide, could be detected by UV absorbance.

The inactivation experiments using the four mechanically mixed oxidant solutions were carried out at pH 8.2 at 20°C. The mechanically mixed oxidant solutions of FAC and ozone (0.01 mg/L), and FAC and chlorine dioxide (0.018 mg/L and 0.18 mg/L) were 21%, 26% and 45% more effective, respectively, than FAC alone in achieving 2 log removal of *B. subtilis* spores. No notable difference

in inactivation rate was observed between the mechanically mixed oxidant solution of FAC and hydrogen peroxide (9.2 mg/L). The mechanically mixed oxidant solution of FAC and chlorite (40 mg/L) increased the inactivation rate by 52% compared to disinfection with FAC alone, but as previously mentioned, the amount of chlorite in the prepared solution was much higher than expected for chlorite generated from an electrochemical cell. Son et al. (2005) explained the enhanced disinfection efficacy of mechanically mixed oxidants containing ozone, chlorine dioxide and chlorite by the synergistic effects of the mixed oxidants, as well as by intermediates generated through the reaction of ozone, chlorine dioxide and chlorite with FAC. Throughout the measurement of UV absorbance at wavelengths specific to ozone, chlorine dioxide and chlorite, Son et al. (2003; 2005) were able to show that chlorite was produced as an intermediate from the reaction of FAC and ozone, and chlorine dioxide was generated from the reaction of FAC and chlorite.

6. Conclusion

The efficacy rate of mixed oxidant solutions compared to conventional chlorine disinfection is greatly affected by the type of microorganism used in the experiment, or found in the drinking water supply. While the bacteriophage MS2 and E. coli are inactivated rapidly by mixed oxidant solutions (Clevenger et al. 2007; Son et al. 2003; Casteel et al. 2000), B. subtilis spores, C. perfringens spores and C parvum oocysts are increasingly more resistant (Son et al. 2003; Casteel et al. 2000; Venczel et al. 1997); however, at higher doses and/or with greater contact times, mixed oxidant solutions have been found to be effective in inactivating these more resistant microorganisms (Son et al. 2003; Casteel et al. 2000; Venczel et al. 1997). This is especially significant for chlorine-resistant microorganisms like C. parvum, for which no inactivation was observed with treatment with free chlorine alone after 24 hours of contact time (Venczel et al. 1997).

Comparison of the inactivation rates of different microorganisms using the same method and dose of disinfection emphasizes the importance of selecting appropriate biological indicators. For instance, while *C. perfringens* spores have shown to be reliable indicators for *C. parvum* oocyst inactivation by mixed oxidants in water, the bacteriophage MS2 and *E. coli* would be inappropriate as indicators for *C. parvum* oocyst inactivation (Casteel *et al.* 2000).

As observed by Son et al. (2003), the species in electrochemically generated mixed oxidant solutions likely play a larger role, or have a greater capacity for disinfection, at slightly basic conditions. Mixed oxidant solutions appear to be the most effective at temperatures between 20-25° C; however, mixed oxidant solutions may be able to better maintain their disinfection potency at lower temperatures (4° C) compared to FAC alone. Synergistic effects observed in sequential disinfection experiments from the reaction of disinfectant species (Liyanage et al., 1997; Cho et al., 2006; Yang et al., 2012; Beber de Souza & Daniel, 2011) may also occur between the oxidant species in mixed oxidant solutions, contributing to the disinfection efficacy of electrochemically mixed oxidant solutions. Synergism of mechanically mixed oxidants, as well as the disinfection abilities of intermediates formed by the reaction of oxidant species, was observed in the enhanced inactivation of *B. subtilis* spores by mechanically mixed oxidants in concentrations believed to be similar to those generated by the electrolysis of brine (Son et al. 2003; 2005).

In addition to increased disinfection potency, noteworthy benefits of mixed oxidant solutions include, but are not limited to, decreased generation of potentially harmful disinfection byproducts (DBPs), inhibition of biofilm formation within distribution systems, improved residual power, safer operating environments, and lower operating costs (Venczel *et al.* 1997; Hamm 2002).

Based on the study, the use of mixed oxidants appears to be a viable approach in the disinfection of drinking water, and may prove to be very valuable with more research that investigates and confirms the constituents and mechanisms responsible for the enhanced inactivation of a number of microorganisms.

References

Beber de Souza, J. & Daniel, L.A. (2011). Synergism

effects for *Escherichia coli* inactivation applying the combined ozone and chlorine disinfection method. *Environmental Technology*. 32, 1401-1408.

Casteel, M.J., Sobsey, M.D. and Arrowood, M.D. (2000). Inactivation of *Cryptosporidium parvum* oocysts and other microbes in water and wastewater by electrochemically generated mixed oxidants. *Water Sci Technol.* 42, 127-134.

Cho, M., Kim, J.H. & Yoon, J. (2006). Investigating synergism during sequential inactivation of *Bacillus subtilis* spores with several disinfectants. *Water Res.* 40, 2911-2920.

Clevenger, T., Wu, Y., DeGruson, E., Brazos, B & Banerji, S. (2007) Comparison of the inactivation of *Bacillus subtilis* spores and MS2 bacteriophage by MIOX, ClorTec and hypochlorite. *Journal of Applied Microbiology*. 103, 2285-2290.

Hamm, B. (2002). DBP reduction using mixed oxidants generated on site. *American Water Works Association Journal*. 94, 49-53.

Korich, D.G., Mead, J.R., Madore, M.S., Sinclair, N.A. & Sterling, C.R. (1990). Effects of ozone, chlorine dioxide, chlorine, and monochloramine on *Cryptosporidium parvum* oocyst viability. *Appl. Environ. Microbiol.* 56, 1423-1428.

Liyanage, L.R.J., Finch, G.R. & Belosevic, M. (1997). Sequential disinfection of *Cryptosporidium parvum* by ozone and chlorine dioxide. *Ozone Sci. Eng.* 19, 409-423.

Son, H., Cho, M., Kim, J., Chung, H., Sohn, J. & Yoon, J. (2003). Comparison of disinfection efficiency of electrochemically or mechanically mixed oxidants with free available chlorine. *American Water Works Association. WQTC Conference.* 1-16.

Son, H., Cho, M., Kim, J., Oh, B., Chung, H. & Yoon, J. (2005). Enhanced disinfection efficiency of mechanically mixed oxidants with free chlorine. *Water Res.* 39, 721-727.

Venczel, L.V., Arrowood, M., Hurd, M. & Sobsey, M.D. (1997). Inactivation of *Cryptosporidium parvum* oocysts and *Clostridium perfringens* spores by a mixed-oxidant disinfectant and by free chlorine. *Appl. Envr. Microbiol.* 63, 1598-1601. Yang, W., Yang, D., Zhu, S., Chen, B., Huo, M. & Li, J. (2012). The synergistic effect of *Escherichia coli* inactivation by sequential disinfection with low level chlorine dioxide followed by free chlorine. *Journal of Water and Health.* 10.4, 557-564.
Using a Combination of Electroencephalographic and Acoustic Features to Accurately Predict Emotional Responses to Music

Devon Krish^{1*}

¹ Los Gatos High School, Los Gatos, CA USA

Received July 21, 2021; Revised February 11, 2022; Accepted, February 22, 2022

Abstract

Music has the ability to evoke a wide variety of emotions in human listeners. Research has shown that treatment for depression and mental health disorders is significantly more effective when it is complemented by music therapy. However, because each human experiences music-induced emotions differently, there is no systematic way to accurately predict how people will respond to different types of music at an individual level. In this experiment, a model is created to predict humans' emotional responses to music from both their electroencephalographic data (EEG) and the acoustic features of the music. By using recursive feature elimination (RFE) to extract the most relevant and performing features from the EEG and music, a regression model is fit and accurately correlates the patient's actual music-induced emotional responses and model's predicted responses. By reaching a mean correlation of r = 0.788, this model is significantly more accurate than previous works attempting to predict music from brain activity is possible. Furthermore, by testing this model on specific features extracted from any musical clip, music that is most likely to evoke a happier and pleasant emotional state in an individual can be determined. This may allow music therapy practitioners, as well as music-listeners more broadly, to select music that will improve mood and mental health.

Keywords: EEG, music therapy, acoustic features, machine learning, emotional-response predictions

1. Introduction

Music is known to be an extremely powerful tool that can make listeners feel pleasure, happiness, sadness, and even fear (Fritz et al., 2009). Music therapy is also a health-intervention that has subsequently proven to be an effective treatment for poor mental health, mood disorders, and depression (Maratos et al., 2008). For example, music therapy was determined to significantly improve mood when compared to treatment as usual (Maratos et al., 2008; Ramirez et al., 2018). Furthermore, Chen (1992) determined antidepressant drugs coupled with music therapy were more effective than antidepressant drugs alone in improving mental health. Also, Koelsch & Jäncke (2015) concluded that music can reduce pain and anxiety in patients with heart disease by lowering heart rate and blood pressure.

In music therapy, a therapist prescribes music selections to a patient to listen to based solely on the therapist's expertise, experiences, and evaluation of the patient (Tamplin & Baker 2006; Maratos et al.,

* Corresponding Author devonkrish05@gmail.com Advisor: Tyler Giallanza tylerg@princeton.edu

2008). However, to determine the optimal music to prescribe to the patient, the individual's emotional reaction to the signal must be predicted. Predicting a human's emotional response to a musical selection they have never heard is an extreme challenge because all humans experience music-induced emotions differently depending on their life experiences, moods, influences, gender (Hunter, Schellenberg & Schimmack 2010; McRae et al., 2008), and a wide variety of other factors. Despite music therapy's promising results as a healing agent for depression and mental health disorders, music therapy does not currently rely on a systematic method to predict emotional responses to music on an individual level.

Many models for the relationship between induced emotions in humans and attributes of music have been created. For example, Schubert (2004) studied the effects of musical features and music theory attributes such as dynamics and melodic contour on emotions. Gabrielsson & Lindström (2001) subsequently studied the effects of variations of tempo, articulation, dynamics, and intonation on perceived emotion by using a model in which emotional responses ranging from strong to weak were plotted across two axes: one for pleasantness and one for excitement. This model also featured plots of tempo and other musical descriptors. This model may provide information about how individual music descriptors affect emotional responses. However, because all people perceive emotions differently and therefore individual-level emotional predictions are necessary, using physiological data may be of significant value.

Other researchers have used physiological measurements as correlates for induced emotional responses. For example, Etzel et al. (2006) tested the effect of music meant to evoke different moods including happiness, sadness, and fear on cardiovascular activity. Similar research has found that certain selections of pleasant and happiness-inducing music may increase heart rate (Brouwer et al., 2013), as revealed through the analysis of electrocardiogram (ECG) signals (Kim & André 2008). More physiological measurement techniques used to correlate music-induced emotions include skin conductance response (SCR) (Khalfa,

Isabelle, Jean-Pierre, & Mannon 2002) electromyography (EMG) for facial expression (Lundqvist et al., 2008), electrodermal activity (EDA) (Craig 2005), respiration rate (Merrill et al., 2020), and pupillometry (Nakakoga et al., 2020).

On the other hand, researchers have also turned to studying brain activity to measure music-induced emotions. Such experiments have been performed using functional magnetic resonance imaging (fMRI) for example (Koelsch et al., 2006; Brattico et al., 2011). Alternatively, researchers have also used electroencephalographic (EEG) data (Daly et al., 2015; Schmidt & Trainor 2001).

Combining EEG with music, video, and other media features has proven to be effective at predicting emotional responses (Kortelainen & Seppänen 2013; Li et al., 2018) because music and their induced emotions are thoroughly believed to engage large networks of neurons and neuronal structures (Daly et al., 2019). For instance, a study by van Tricht et al. (2010) found that Parkinson's disease impaired the emotional recognition of fear and anger in music. Similarly, the deracination of the impaired anteromedial temporal lobe the music-induced recognition of terror and scariness. (Gosselin et al., 2011). Thus, music-induced emotions may relate to a variety of changes in EEG.

That being said, using EEG to predict music-induced emotions has shown to be a difficult feat because EEG is non-stationary and very noisy (Daly et al., 2015; Koelstra et al., 2012; Jiang et al., 2019). This problem is further challenged by combining EEG with variations in how humans perceive emotions (Barrett et al., 2007), differences in musical preferences (Bauer, Kreutz, & Herrmann 2015), and disparities in age and gender which may influence these preferences (Vieillard & Gilet 2013; Le Blanc et al., 1999).

Therefore, by selecting only the most important and relevant descriptors, a combination of both acoustic and EEG features may be used to train a model to predict music-induced emotions at a high accuracy on the individual level.

Since classical-style music has been shown to induce emotions at a strong level (Schaefer 2017; Kreutz et al., 2007), this study utilizes the dataset created by Daly et al. (2015), in which classical music clips are played to participants while their EEG is recorded. Descriptive features from the musical clips (from the set of stimuli from Eerola & Vuoskoski 2010) are then extracted and brain activity to train a Lasso regression model to predict each participant's emotional responses to music.

2. Materials and Methods

2. 1 Methods and Experimental Data

This analysis utilized the EEG and emotional response data gathered from Daly et. al (2015). The data consisted of thirty-one individuals, thirteen males and eighteen females, whose ages ranged from 18 to 66. Each participant's electroencephalogram (EEG) was recorded from 19 electrode channels positioned from the nasion to inion as according to the 10-20 EEG Placement System (Milnik 2006).

This study also featured classical music stimuli drawn from a dataset of 360 excerpts from film scores including titles such as *Psycho (1960)*, *Gladiator (2000)*, and *Big Fish (2003)*. The musical stimuli in this dataset were specifically chosen to induce specific emotions in human listeners (Eerola & Vuoskoski 2010).

The participants were told to stay still and each listened to 40 randomly drawn musical selections from Eerola & Vuoskoski (2010) for 15 seconds. Immediately after, they responded to a series of eight Likert-scale questions on a scale of strongly disagree to strongly agree to identify their emotional response across eight axes: happiness, sadness, pleasantness, fear, anger, tenderness, tension, and energy-level (Daly et al., 2015).

Because some of these emotion categories are likely to be highly correlated (Larsen & McGraw 2001), a Principal-Component Analysis (PCA) was used to reduce the eight axes into three principal components (PCs), which explains a large amount of the variance (75%) of the participant's music-induced emotional responses (Daly et al., 2015). These three PCs represent valence-arousal, a measurement of pleasure and happiness; energy-arousal, а quantification of liveliness; and tension-arousal, a measurement of tenseness (Ilie & Thompson 2006; Ilie & Thompson 2011), as according to the

Schimmack and Grob model (Schimmack & Grob 2000). It is these three PCs that are ultimately subjected to further analysis.

Acoustic features from the music played to each participant from the musical dataset and various anatomical features from the participants' EEG data are then extracted. Through recursive feature elimination (RFE), the most performing features are selected to accurately predict a participant's self-reported emotional response to music along each of the three axes of the PCs.

2.2 EEG Features

Because EEG signals are so noisy (Daly et al., 2012; Jiang et al., 2019), the EEG signals were first pre-processed to remove artefacts and noise. Discrete Wavelet Transform (DWT) was then used to decompose the EEG signals into alpha (α) [12-24Hz), beta (β) [24-48Hz), gamma (γ) [48-80Hz), delta (δ) (0-6Hz), and theta (θ) [6-12Hz) wavelet bands. DWT provides high-frequency resolution at the high and low EEG frequencies. DWT is an extremely convenient tool for processing non-stationary EEG and rendering these signals more suitable for feature extraction than just standard EEG alone (Qazi et al., 2016). From the α , β , γ , δ , and θ wavelet bands, a total of 285 features were extracted from the EEG signals using Minimum Norm Estimates (MNE). RFE was further utilized to determine that the features below were beneficial to training the model.

The energy of each wavelet band, or strength of the signal at any time interval as it represents the area under the curve (AUC), was computationally calculated.

The mean and standard deviation of each wavelet band was also extracted and used to train the model. The standard deviation supplies information about how close the features are to the signal's mean and is mathematically represented as shown below:

$$STD = \sqrt{\sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N - 1}}$$
 (1)

Where x_i represents the random variable with a mean signal of μ



Figure 1. A visual depiction of the energy of the EEG signals for each of the 19 independent electrodes for one participant while listening to one of the songs. The intensity of the signal in millivolts is represented by the y-axis and the x-axis represents time (each musical excerpt is 16 seconds long).

The entropy of an EEG signal represents the outcome uncertainty measurement and is calculated through the following formula:

$$H(x) = -\sum_{i=1}^{n} P(x_i) log[P(x_i)]$$
(2)

Where x represents the random variable with possible outcomes $x_{(i=1)}, ..., x_n$, which occur with a probability $P(x_i)$

The Hjorth parameter of mobility is a normalized slope descriptor that represents the square root of the variance of the first derivative of the signal divided by the variance of the signal (Nascimben et al., 2019). It is otherwise stated as the proportion of standard deviation to a signal's power spectrum and has been noted to be advantageous to emotion recognition (Li et al., 2018). Results from the addition of the Hjorth parameter to the developed model significantly improved the model's accuracy, as hypothesized. This mobility parameter is mathematically represented as shown below:

$$H_m(t) = \sqrt{\frac{var(\frac{dy(t)}{d(t)})}{var(y(t))}}$$
(3)

 $H_m(t)$ equates to the square root of the variance of the first derivative of signal y(t) divided by the total variance of signal y(t)

Power frequency bands extract features by using

frequency bands to compute the power spectrum of an EEG signal (Al-Fahoum & Al-Fraihat 2013). The five frequency wavelet bands extracted the power of each of the 19 electrodes' signals, resulting in 95 features.

2.3 Acoustic Features

From each of the 360 musical clips used as stimuli, a range of specific acoustic features was extracted. These extracted features included both spectral and temporal types, ranging from describing each piece's key and tonal qualities to the speed. A total of 352 musical features were selected. RFE was also used to make sure the acoustic features were beneficial to the model. Furthermore, for each of the eleven acoustic feature types (except key and bpm), the kurtosis, maximum, mean, median, minimum, skewness, and standard deviation were also extracted. The acoustic feature types are described below.

Chroma and chroma grams can be described as the transformation of a musical signal's pitches (twelve possible pitches in total) through time. From chroma, both the intensity and certain pitch are extracted - a total of 84 features.



Figure 2. A chromagram in which the x-axis is represented by time in seconds of a signal while the y-axis represents the pitches. The redder a pitch is at a certain time, the more intense that pitch is.

BPM or beats per minute provides the number of beats in the signal per minute. A low tempo of 40 BPM, for example, is very likely to indicate a slower, solemn musical stimulus while a tempo of 160 BPM is likely to indicate an upbeat, faster-paced musical piece. BPM proved to be an extremely important feature in determining the emotional responses to a musical selection.

Zero-crossing rate (ZCR) can be described as the rate at which the signal of a music stimulus may change from positive to negative or negative to positive, thus crossing zero (zero-crossings) in a fixed amount of time; ZCR amounted to a total of seven features.



Figure 3. A plotted depiction of ZCR versus samples of one musical clip, in which every time the signal crosses y = 0 (illustrated by the red line), the number of zero-crossings (n) increases by one.

Major and minor keys turned out to be an extremely significant feature in determining the emotional responses to an audio signal. A major key versus a minor key is determined by analyzing the notes or pitches present in a signal. It is also widely acknowledged that songs in major keys tend to sound bright and cheerful while songs in minor keys are more melancholy.

MFCC or Mel-frequency cepstral coefficients are computed by taking the logarithm of the Fourier coefficients of an audio signal that has been converted to the Mel-scale. MFCCs ultimately represent the timbre of the audio signal (Stevens 1937; Garima & Barkha 2013). These 140 features are also commonly used in speech recognition systems as detailed in (Anggraeni et al., 2018).

Tonnetz is a feature that determines the tonal centroids, or harmonic components of the signal when extracted. The spectral centroid of a musical signal defines which frequency the energy of a spectrum is centered upon or where the center of mass of the spectrum is located. The spectral roll-off is the frequency below which a percentage (normally 0.85) of the spectral energy of the signal lies (Jang et al., 2008). The spectral contrast is the difference in level between the crest and troughs of the spectrum.

The spectral flatness can determine how noisy a signal is in decibels (with 1.0 indicating the spectrum is white noise). Finally, the spectral bandwidth, or the extent of the power transfer around the center frequency of the audio signal, is extracted (Theimer et al., 2008). Altogether, the spectral features contribute 77 features to the model.



Figure 4. An MFCC spectrum plot in which the x-axis represents the time in seconds of a signal and the y-axis represents the increasing MFCC coefficients of a signal. Similar to the chromagram in Figure 1, the redder the pitch is at a certain time, the more intense that coefficient is.

2.4 Training the Model

A total of 637 extracted features (285 EEG features and 352 acoustic features) were used to train a Lasso-regularized linear regression model (with an alpha parameter set to 10) to predict the participants' emotional responses to music in terms of the PCs. To best fit the data and generate accurate predictions, a multi-task, cross-validated, Lasso regression model with five folds was used. In each of the five folds, the features were split into a training and testing set. Combinations of the training set features were then related to the PCs to create a linear regression model that would fit the emotional responses of the participants as noted by their recorded PCs.

After being trained on the said training set, the model attempted to predict the participant response PCs from the previously unseen features in the testing set. By identifying how close the model's predicted response PCs are to the participants' actual, recorded PCs for each data point in the testing set of each cross-fold, the model's performance and statistical can be determined.

2.5 Recursive Feature Elimination (RFE)

As first utilized for gene selection, RFE is an effective method that determines the importance of features. The RFE algorithm was trained on the whole set of features initially (637 features: 285 from the EEG and 352 from the music). The algorithm could then determine the importance of each feature to the correlation model by assigning weights and eliminating the lowest-ranking features (Li et al., 2018). This process occurred recursively for several rounds until all the features had been selected. Using an RFECV (cross-validation) selector, the 17 feature types (comprising the 637 features) were ranked to determine which features were most beneficial to obtain the regression model's correlation. RFE was also used to determine that the features would only improve the model's prediction accuracy.

3. Results

3.1 Correlation Analysis

The model's prediction performance is first evaluated on the EEG and acoustic features separately and then combined. These feature subsets are then used to evaluate the response PCs individually. For each PC, the mean correlation (r), between the actual recorded PC and the predicted PC is calculated. The results of these correlations are displayed in Table 1. Given the noise and non-stationarity of the EEG data (Hassani & Karami 2015) and the 75% variance of the participant's emotional responses (Daly et al., 2015), training the model on each feature subset resulted in a predicted response PC of high correlation (p<0.001). As displayed in Table 1, the model's correlation was highest when both the EEG and acoustic features are combined.

The results of this correlation are unparalleled to previous works dealing with predicting music-induced emotions and EEG. For example, our r-values are almost 370% better than a comparable study by Daly et al., (2015).

3.2 RFE Analysis

The RFE selector's ranking of the 637 features determined which of the music and EEG feature types were the most important to the regression model. For the acoustic features, the key and tempo of the song performed best for the model. Similarly, the Hjorth parameter of mobility was the most vital EEG feature for training the model. Additional rankings are summarized below in Table 2.

Table 1. The mean correlation performance of the regression model at predicting the response PC from both acoustic and EEG features, acoustic features alone, and EEG features alone. All results are highly statistically significant (p<0.001), but the model performs notably well when trained on both acoustic and EEG features.

Pagpanga PCa	Mean correlation (r)		
Response PCS	Both	Acoustic	EEG
PC1 (Valence-arousal) PC2 (Tension-arousal) PC3 (Energy-arousal)	0.774 0.791 0.798	0.474 0.362 0.435	0.355 0.381 0.494

Table 2. A ranking of the acoustic and EEG features. The acoustic features were ranked in the selector from 1 to 352. The mean feature score of each of the 11 acoustic feature types is notated by μ , in which the lower the μ value, the more important the feature type was to the regression model. Similarly, the EEG features were ranked in the selector from 1 to 285. For the EEG, μ ' denotes the mean feature score of each of the 6 EEG feature types.

ACOUSTIC FEATURES	EEG FEATURES
Major/Minor Key (µ=1.00)	Hjorth Mobility (μ '=10.00)
BPM/Tempo (µ=2.00)	Energy (µ'=76.50)
Chroma (µ=43.48)	Standard Deviation (μ '=143.00)
Spectral Centroid (µ=89.00)	Mean (μ '=162.00)
Spectral Rolloff (μ =124.71)	Entropy (µ'=181.00)
MFCC (µ=171.50)	Power Spectrum (μ '=238.00)
Spectral Flatness (μ =174.29)	
Spectral Bandwidth (µ=250.00)	
Spectral Contrast (µ=284.18)	
ZCR (µ=307.00)	
Tonnetz (µ=331.50)	

3.3 Demographic Analysis

Because the subjects of this study had a variety of age and gender differences and age and gender have been noted to convey differences in affective responses to emotions conveyed by music (Vieillard & Gilet 2013; Hunter et al., 2011), we tested if participant gender and age affected our model's results. Therefore, predictions were calculated on an individual, per-participant level. Correlations between participant gender and ages and the predictions were then noted.

T-tests proved that the model's ability to predict emotional responses was not influenced by gender or age (p=0.413 and p=0.278 respectively).

The strength of the subjects' emotional responses to the music as taken from the Likert-scale was also tracked. On a scale from 0 to 4, with 4 being an extremely strong emotional response and 0 being a very weak emotional response, analyses found that female participants had an average emotional response strength of 2.48 while males' strengths were 1.99. This difference indicates that the female participants were 24.2% more emotionally expressive towards the music than the males (p<0.001). However, in comparing the emotional response strengths for ages 18-66, no statistically significant differences were found (p=0.250).

3.4 Pop Songs (Acoustic Test) For Valence

The need for predicting music-induced emotions at an individual level is vital to music therapy and prescribing people music to improve mood and mental health. Because this model achieved such a relatively high correlation between the predicted emotions and actual responses, we tested if the model could benefit music therapy by predicting which songs would elicit the most positive emotional response in the participants. Although the combination of EEG and acoustic features performed the best, the acoustic features alone yielded a relatively high correlation in comparison to previous research (Daly et al., 2015; Song & Dixon 2015). Thus, the acoustic features were tested to see if they could be used to predict responses to pieces of music not previously listened to by the participants.

As stated in 2.1, the first PC, valence, describes a musical stimuli's positiveness (McConnell & Shore 2010). For example, music with higher valence induces more happy emotions than clips with low valence.

Given that most people listen to pop music over classical music and because music therapy most likely requires music longer than 15 seconds, the model was tested on a random set of longer, three-minute pop songs, as opposed to the original, short classical music excerpts. The model was trained on the acoustic features extracted from the classical music in 2.3 and tested on the held-out, acoustic features extracted from the pop songs. The model was then isolated for valence and ordered the songs from largest to smallest valence, thus indicating which songs would elicit the most cheerful emotional responses in the 31 participants of this study.

The model's performance in predicting participant's valence responses to pop songs they've never heard, based on their emotional responses to short classical music clips is summed up in Table 3.

Table 3. The model was trained on acoustic features extracted from short classical music clips and ordered a random selection of 10 pop songs from highest to lowest. As valence rating increases, the song's level of induced-happiness increases.

Song Name	Artist	Valence Rating	Valence Order
Dancing Queen	ABBA	1.000	High Valence
Jump	Van Halen	0.985	†
Stayin' Alive	Bee Gees	0.973	
Good Life	OneRepublic	0.965	
Нарру	Pharrell Williams	0.745	
Uptown Girl	Billy Joel	0.606	
Yesterday	Beatles	0.580	
Say Something	A Great Big World	0.054	
Unchained Melody	Righteous Brothers	0.003	
Tears in Heaven	Eric Calpton	0.000	Lower Valence

4. Discussion

In this section, observed results and data should be interpreted concisely. In addition, the authors need to explain the significance of the research findings and revisit the hypothesis described in the introduction.

Although music therapy is an effective treatment for poor mental health disorders and depression (Maratos et al., 2008), it does not currently employ a systematic method to predict emotional responses to music on an individual level. Predicting human emotional responses to music at high accuracy is a challenging problem because factors such as a person's age, gender, mood, and memories may affect how people emotionally respond to music. This challenge is heightened by combining the acoustic features of the music with the non-stationariness and noisiness of EEG.

Results from this study show that by combining EEG-derived features with acoustic features, emotional responses to music can be predicted at a significantly higher accuracy. This suggests that emotional responses to music are not just based on the musical properties of the music, but also the listener's idiosyncrasies and internal processes.

By training a regression model with the most performing EEG and acoustic features. music-induced emotions can be predicted at a higher accuracy than previously reported. For example, this experiments' outcomes were compared to Daly et al. (2015), which used the same dataset. By processing the EEG and audio differently and extracting unique and performing features only, our model achieved much better correlations. For example, for valence-arousal, they achieved a correlation of 24.3% \pm 0.5%, while we achieved a correlation of 77.4%. They achieved a mean correlation of $15.8\% \pm 0.6\%$ for energy-arousal, while our model reached a correlation of 79.1%. Finally, for tension-arousal, while they only achieved an accuracy of $10.2\% \pm$ 0.5%, we met a mean correlation of 79.8%. This represents a 370% increase in accuracy over Daly et al. (2015).

Other researchers have analyzed emotions elicited via the DEAP dataset, using EEG to predict music video-induced emotions (Nascimben et al., 2019; Kumar et al., 2016). However, these experiments extracted features solely from the EEG data and used music videos as opposed to just music as stimuli. Ultimately, Nascimben et al., (2019) achieved a cross-validation accuracy of 65.4%. Additionally, (Kumar et al., 2016) used a single cross-validation run to achieve accuracies of 57.6% for valence-arousal and 62.0% for arousal., Our model is 32% more accurate. Age and gender differences have been noted to affect emotional responses to music (for example, Vieillard & Gilet (2013) and Hunter et al., (2011)). While there were no differences in emotional responses across age, emotional responses across gender varied heavily, thus supporting Hunter et al., (2011). The strength of female responses to music was 24.2% more intense than those of males. This indicates that females are more emotionally expressive to music than men. That being said, our model's prediction accuracy for music-induced emotions was still consistent across gender and age.

Results from testing our model on a variety of pop songs show that this model can successfully be used in a clinical setting to improve music therapy techniques. For example, by giving people small selections of songs and recording their emotional responses to the clips on eight axes, our model can prescribe them a list of songs targeting specific emotions, e.g. happiness. Ideally, these treatments would also use patients' EEG as a parameter, but given that EEG may be expensive and infeasible, it is not necessary (as demonstrated in 3.4).

Ultimately, the model created in this experiment can predict emotional responses to music on an individual level at significantly higher correlations than previously recorded. Our model has the potential to work as a music system to prescribe songs to patients to induce happier and more pleasant emotional states and advance music therapy as a treatment for mental health disorders and the emerging field of brain-computer music interfaces. Future work will strive to use these findings to create machine-generated music with acoustic properties that induce certain emotional states.

Acknowledgments

The author would like to thank Tyler Giallanza for his support and contributions.

References

Al-Fahoum, A. S., & Al-Fraihat, A. A. (2014). Methods of eeg signal features extraction using linear analysis in frequency and time-frequency domains. *ISRN Neuroscience*, 2014, 1-7. doi:10.1155/2014/730218 Anggraeni, D., et al. (2018). The implementation of speech recognition USING Mel-Frequency Cepstrum COEFFICIENTS (mfcc) and support vector Machine (SVM) method based on Python to CONTROL robot arm. *IOP Conference Series: Materials Science and Engineering*, 288, 012042. doi:10.1088/1757-899x/288/1/012042

Barrett, L. F., et al. (2007). The experience of emotion. *Annual Review of Psychology*, *58*(1), 373-403. doi:10.1146/annurev.psych.58.110405.085709

Bauer, A. R., Kreutz, G., & Herrmann, C. S. (2014). Individual musical Tempo preference correlates with EEG beta rhythm. *Psychophysiology*, *52*(4), 600-604. doi:10.1111/psyp.12375

Brattico, E., et al. (2011). A functional mri study of happy and sad emotions in music with and without lyrics. Frontiers in Psychology, 2. doi:10.3389/fpsyg.2011.00308

Brouwer, A., et al. (2013). Perceiving blocks of emotional pictures and sounds: Effects on physiological variables. *Frontiers in Human Neuroscience*, 7. doi:10.3389/fnhum.2013.00295

Chen, X. (1992). Active music therapy for senile depression. *Zhonghua shen jing jing shen ke za zhi* = *Chinese Journal of Neurology and Psychiatry*, 25(4), 208–210. 252–3. https://pubmed.ncbi.nlm.nih.gov/1478135/

Craig, D. G. (2005). An exploratory study of physiological changes during "chills" induced by music. *Musicae Scientiae*, *9*(2), 273-287. doi:10.1177/102986490500900207

Daly, I., et al. (2012). What does Clean EEG look like? 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. doi:10.1109/embc.2012.6346834

Daly, I., et al. Music-induced emotions can be predicted from a combination of brain activity and acoustic features. *Brain Cogn. 2015 Dec*; 101:1-11. doi: 10.1016/j.bandc.2015.08.003.

Daly, I., et al. (2019). Electroencephalography reflects the activity of subcortical brain regions during APPROACH-WITHDRAWAL behaviour while listening to music. *Scientific Reports*, 9(1).

doi:10.1038/s41598-019-45105-2

Fritz, T., et al. (2009). Universal recognition of three basic emotions in music. *Current Biology*, *19*(7), 573-576. doi:10.1016/j.cub.2009.02.058

Eerola, T., & Vuoskoski, J. K. (2010). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39* (1), 18-49. doi:10.1177/0305735610362821

Etzel, J. A., et al. (2006). Cardiovascular and respiratory responses during musical mood induction. *International Journal of Psychophysiology*, *61*(1), 57-69. doi:10.1016/j.ijpsycho.2005.10.025.

Gabrielsson, A., & Lindström, E. (2001). *The influence of musical structure on emotional expression*. In P. N. Juslin & J. A. Sloboda (Eds.), *Series in affective science. Music and emotion: Theory and research* (p. 223–248). Oxford University Press.

Gosselin, N., Peretz, I., Hasboun, D., Baulac, M., & Samson, S. (2011). Impaired recognition of musical emotions and facial expressions following anteromedial temporal lobe excision. *Cortex*, 47(9), 1116-1125. doi:10.1016/j.cortex.2011.05.012

Hassani, M., & Karami, M. (2015). Noise estimation in electroencephalogram signal by using volterra series coefficients. *Journal of Medical Signals & Sensors*, 5(3), 192. doi:10.4103/2228-7477.161495

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity, and the Arts, 4(1), 47–56.* https://doi.org/10.1037/a0016873

Hunter, P. G., Glenn Schellenberg, E., & Stalinski, S. M. (2011). Liking and identifying emotionally expressive music: Age and gender differences. *Journal of Experimental Child Psychology, 110*(1), 80-93. doi:10.1016/j.jecp.2011.04.001

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception, 23*(4), 319-330. doi:10.1525/mp.2006.23.4.319

Ilie, G., & Thompson, W. F. (2011). Experiential and

cognitive changes following seven minutes exposure to music and speech. *Music Perception*, 28(3), 247-264. doi:10.1525/mp.2011.28.3.247

Jang, D., Jin, M., & Yoo, C. D. (2008). Music genre classification using novel features and a weighted voting method. *2008 IEEE International Conference on Multimedia and Expo*. doi:10.1109/icme.2008.4607700

Jiang, X., Bian, G., & Tian, Z. (2019). Removal of artifacts from eeg signals: A review. *Sensors*, *19*(5), 987. doi:10.3390/s19050987

Khalfa, S., et al. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters*, *328*(2), 145-149. doi:10.1016/s0304-3940(02)00462-7

Kim, J., & Andre, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(12), 2067-2083. doi:10.1109/tpami.2008.26

Koelsch, S., & Jäncke, L. (2015). Music and the heart. *European Heart Journal*, *36*(44), 3043-3049. doi:10.1093/eurheartj/ehv430

Koelstra, S., et al. (2012). DEAP: A database for emotion Analysis using physiological signals. *IEEE Transactions on Affective Computing*, *3*(1), 18-31. doi:10.1109/t-affc.2011.15

Kortelainen, J., & Seppanen, T. (2013). EEG-based recognition of video-induced emotions: Selecting Subject-independent feature set. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). doi:10.1109/embc.2013.6610493

Kreutz, G., et al. (2007). Using music to induce emotions: Influences of musical preference and absorption. *Psychology of Music*, *36*(1), 101-126. doi:10.1177/0305735607082623

N. Kumar, K. Khaund, & S. M. Hazarika, "Bispectral Analysis of EEG for Emotion Recognition," Procedia Computer Science, vol. 84, pp. 31–35, 2016.

Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology*,

81(4), 684-696. doi:10.1037/0022-3514.81.4.684

LeBlanc, A., et al. (1999). Effect of Age, Country, and Gender on Music Listening Preferences. *No. 141, The 17th International Society for Music Education: ISME Research Seminar,* 72-76.

Li, X., et al. (2018). Exploring EEG features In Cross-Subject emotion recognition. *Frontiers in Neuroscience, 12*. doi:10.3389/fnins.2018.00162

Lundqvist, L., et al. (2008). Emotional responses to music: Experience, expression, and physiology. *Psychology of Music*, *37*(1), 61-90. doi:10.1177/0305735607086048

Maratos, A., S., et al. (2008). Music therapy for depression. The Cochrane Database of Systematic Reviews (1), CD004517. doi: 10.1002/14651858.CD004517.pub2

McConnell, M. M., & Shore, D. I. (2010). Upbeat and happy: Arousal as an important factor in studying attention. *Cognition and Emotion*, *25*(7), 1184-1195. doi:10.1080/02699931.2010.524396

McRae, K., et al. (2008). Gender differences in emotion regulation: An fmri study of cognitive reappraisal., *Group Processes & Intergroup Relations*, *11*(2), 143-162. doi:10.1177/1368430207088035

Merrill, J., Omigie, D., & Wald-Fuhrmann, M. (2020). Locus of emotion influences psychophysiological reactions to music. *PLOS ONE*, *15*(8). doi:10.1371/journal.,pone.0237641

Milnik, V. (2006). Instruction of electrode placement to the international 10-20-system. *Neurophysiologie-Labor*. 28. 113-143.

Nakakoga, S., et al. (2020). Asymmetrical characteristics of emotional responses to pictures and sounds: Evidence from pupillometry. *PLOS ONE, 15*(4). doi:10.1371/journal.,pone.0230775

Nascimben, M., Zoega Ramsoy, T., & Bruni, L. E. (2019). User-Independent classification of emotions in a Mixed AROUSAL-VALENCE Model. 2019 *IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*. doi:10.1109/bibe.2019.00086 Qazi, E., et al. (2016). An intelligent system to classify epileptic and NON-EPILEPTIC Eeg signals. 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). doi:10.1109/sitis.2016

Ramirez, R., et al. (2018). EEG-Based analysis of the emotional effect of music therapy on palliative Care cancer patients. *Frontiers in Psychology*, *9*. doi:10.3389/fpsyg.2018.00254

Schaefer, H. (2017). Music-evoked emotions—current studies. *Frontiers in Neuroscience, 11.* doi:10.3389/fnins.2017.00600

Schimmack, U. & Grob, A. (2000), Dimensional models of core affect: a quantitative comparison by means of structural equation modeling. *Eur. J. Pers.*, *14: 325-345.* https://doi.org/10.1002/1099-0984(200007/08)14:4

Schmidt, L. A., & Trainor, L. J. (2001). Frontal brain electrical activity (EEG) DISTINGUISHES Valence and intensity of musical emotions. *Cognition & Emotion*, *15*(4), 487-500. doi:10.1080/02699930126048

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, *21*(4), 561-585. doi:10.1525/mp.2004.21.4.561

Song, Y., & Dixon, S. (2015). How Well Can A Music Emotion Recognition System Predict the Emotional Responses of Participants? *Queen Mary University of London EECS*. doi:https://www.eecs.qmul.ac.uk/~simond/pub/2015/ SongDixon-SMC2015-EmotionPrediction.pdf

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8(3), 185-190. doi:10.1121/1.1915893

Tamplin, J., & Baker, F. (2006). *Music Therapy Methods in Neurorehabilitation: A Clinician's Manual.*, https://njmt.w.uib.no/2007/08/20/music-therapy-meth ods-in-neurorehabilitation-a-clinicians-manual/

Theimer, W., Vatolkin, I., & Eronen, A. (2008). Definitions of Audio Features for Music Content Description. *Faculty of Computer Science Algorithm* Engineering (LS 11), 18.

van Tricht, M. J., et al. (2010). Impaired emotion recognition in music in Parkinson's disease. *Brain and Cognition*, 74(1), 58-65. doi:10.1016/j.bandc.2010.06.005

Vieillard, S., & Gilet, A. (2013). Age-related differences in affective responses to and memory for emotions conveyed by music: A cross-sectional study. *Frontiers in Psychology, 4*. doi:10.3389/fpsyg.2013.00711

Vyas, G., & Kumari, B. (2013). SPEAKER RECOGNITION SYSTEM BASED ON MFCC AND DCT. International Journal of engineering and advanced technology. 2. 167-169.



The Environmental and Economic Impacts of an Invasive Species in South Florida

Yardena Rubin^{1*}

¹ NSU University School, Davie, FL USA

Received August 30, 2021; Revised February 22, 2022; Accepted, March 7, 2022

Abstract

For decades, invasive species have continued to enter our environment, causing multiple issues. In order to effectively prevent these effects from getting worse, a closer look needs to be taken at both the economic and environmental issues caused by invasive species. This paper aims to bring awareness to the public and to policymakers of the harmful impacts that invasive species have on our planet. With specific research on the chosen species of iguanas (*Iguana iguana*) in South Florida, this project highlights the damage iguanas are doing not only to the environment and economy, but also to the infrastructure. Recent research and experimentation have brought to light the issues that iguanas cause by eating away at the surrounding areas of foundations of different buildings, roads, and bridges, which causes them to become less structurally safe. There have been some advances in the right direction with newly proposed regarding infrastructure. However it is still a long way from directly targeting the problem. This review paper calls attention to these issues currently faced as the threat of these invasive species is expected to continue.

Keywords: Invasive species, Florida iguanas, infrastructure, Iguana iguana, climate change, regulations, invasion mitigation

1. Introduction

An invasive species is an organism that is not native to a specific area or ecosystem (National Wildlife Federation, "Invasive Species", n.d.). In the United States alone, there are approximately 4,300 invasive species (U.S. Fish and Wildlife Service, 2012). Invasive species are usually taken from their natural habitat, where their abundance is relatively controlled, and brought to an ecosystem that is vulnerable to their life history strategies and usually does not have their natural predators. This off-sets the predator-prey relationships in the food chain (National Geographic, 2012). Vertebrata invasive species can cause severe negative impacts on their new environment because they can prey on native species and compete for food or other resources, which can lead to food and space competition between different organisms. These invasive species can even carry disease to these environments, which pose a threat to the native ecosystems (Ricciardi, 2013). Furthermore, invasive species are usually considered r-selected species, meaning that they have high reproduction rates and can adapt easily to new environments. These distinctive characteristics make them extremely dangerous as they can influence and

^{*} Corresponding Author yrubin2022@gmail.com

harm almost all environments they are introduced into (Ricciardi, 2013).

An invasive species is usually introduced into a new ecosystem due to accidental releases, people housing them as pets, and on shipping containers from around the world. This paper specifically focuses on the iguana (Iguana iguana) in South Florida (National Geographic, 2012). Iguanas are invasive in South Florida and pose a threat to the environment (Fig.1). They were originally brought over to the US as pets, yet were released due to the difficulty of caring for them. Iguanas are known to dig up areas around the foundations and supporting structures of seawalls, bridges and buildings in order to lay their eggs. This further affects Florida's infrastructure and ecosystems as it could reduce the effectiveness of systems that prevent flooding such as embankments and seawalls (Table 1). The purpose of this review paper is to provide insight into the struggles of mitigating the iguana invasion in South Florida.



Figure 1. Photos of the Florida green iguana (*Iguana iguana*) taken in June and July of 2021. A) A single iguana in a garden in Hollywood, FL. B) A different individual iguana climbing on a fence in Hollywood, FL. C) Multiple iguanas on a waterfront deck in El Portal, FL.

Table 1. Data of the damages caused by iguanas in Southern Florida. Damages were categorized by type of location, such as dams and power line shortages.

Type of Location	Number of Reported Cases	Cost	Year(s)	Location(s)	Reference
Floodgate/dam	1	\$1,800,000	2020	West Palm Beach, FL	(Childs, 2020)
Power Lines	1	\$40,000	2018	Lauderdale Lakes, FL	(Childs, 2020)

1.1 Florida Iguana Invasion

The iguana is native to Central America, South and the Caribbean islands. America, This cold-blooded species prefers to live in tropical climates and is herbivorous, consuming leaves, flowers, and fruit (National Geographic, 2012). Iguanas have no natural predators in South Florida, can grow up to 1.8m in length, and can weigh over 6.8kg (Florida Fish and Wildlife Conservation Commission, FWC, 2021). Their preferred habitat is near the water as they are strong swimmers. Iguanas were brought to the United States as pets, yet have mainly been released due to the difficulty of caring for them during hurricanes and other natural disasters (FWC, 2021).

Iguanas have species-specific needs in order to survive. They can live anywhere from 12-20 years and prefer to live in extremely warm and humid climates (National Geographic, 2012). When temperatures reach below 10°C, iguanas go into a hibernation state in which they become limp and immobile. This can be dangerous as one may believe that the iguana is dead, yet they are just relatively frozen so once the weather warms up, they defrost and continue with their lives (Townsend, et al., 2003). Iguanas also show their agility and speed as they can run up to 33.8km/h. Additionally, iguanas lay their eggs from 0.6-1.8m underground and can lay up to 40 eggs at a time (Townsend, et al., 2003). Iguana eggs are not exposed to many predators because they are underground, however occasionally some eggs that do not hatch end up being consumed by raccoons and snakes.

Iguanas are invasive in South Florida and are increasingly becoming more dangerous to the environment. Primarily, iguanas pose a threat to the native wildlife because they destroy native crops and plants, which can cause a bottom-up effect in the food chain (Townsend, et al., 2003). This increases the competition for food and resources between iguanas and other species in the environment, which can ultimately lead to a lack of overall food. Likewise, the iguana has recently been observed feeding on flowering plants that are essential to butterfly survival, specifically the Miami Blue Butterfly thomasi (Cyclargus bethunbakeri; Bakkalapulo, 2018). These butterflies have faced critical endangerment due to iguanas in 2018 and needed human intervention in order to be reintroduced in their native habitat (Bakkalapulo, 2018). It is still likely that these butterflies could become critically endangered again, as the increase in the iguana population poses a great threat.

Due to the iguana's preference to burrow near water and underground, iguanas pose huge threats to the infrastructure of South Florida. They are known to dig in areas around the foundations and supporting structures of docks, bridges, and buildings to lay their eggs. This could reduce the effectiveness of systems installed that prevent and control flooding, such as levees and dams (Table 1, Smith, et al., 2008). As seen in Figure 2, the burrowing of the iguanas near the over-water patio could cause significant structural issues and become unsafe to use. Some of the damage that these iguanas are causing includes increasing the number of people displaced from their homes as well as increasing the threat of flooding in an already flood-prone region (Chakraborty, et al., 2014). This invasive species is harming the environment for the local wildlife as well as humans, and regulations have been implemented to help combat the effects of their damage.

2. Methods

For this review paper, articles from the past 15 years were identified after using Google Scholar as a search engine. The following keywords were used either separately or in combinations to determine relevant articles for this review paper: infrastructure, invasive species, iguana, florida iguana invasion, regulation, iguana removal, infrastructure bill, *Iguana iguana*, climate change, invasion mitigation. Using these keywords, the non-peer-reviewed articles were

also found by using the Google search engine. This project focused on invasive animal species in South Florida, 26.3014° N, 80.6327° W (Fig. 3). Additionally, the pictures in this review paper of iguanas and the damage they have caused were taken by the author in El Portal, FL, at a lake-side house and in Hollywood, FL in a home-garden.



Figure 2. Photograph of burrowing damage, shown in blue circles, done by invasive green iguanas (*Iguana iguana*) to lay their eggs near a waterfront home in El Portal, FL, 2021.



Figure 3. A map sourced from the Early Detection & Distribution Mapping System (2021) of South Florida and the recent green iguana (*Iguana iguana*) populations that have been recorded thus far during 2021.

3. Results

3.1 Regulation

While iguanas continue to harm our ecosystems and infrastructure, there have been some developments in regards to the removal of this species. For example, in South Florida, iguanas are protected solely by animal anti-cruelty laws. These laws are in contradiction with the recommendations of the FWC who urge landowners to remove iguanas from their private property.

Removing iguanas while remaining compliant with exhaustive anti-cruelty regulation can prove to be difficult. For instance, to "humanely" kill using a pellet gun, the iguana must be killed using a single shot to the head (Bryan, 2018). However, many cities do not allow these guns to be fired on private property (Bryan, 2018). Another problematic removal method is decapitation. This involves hunting down the iguana and swiftly using a shovel to behead the reptile, which is known to have impressive agility and speed (National Geographic, 2012). If more than one swing is taken at an iguana, it is considered animal cruelty and there could be a fine of up to \$5,000 (Bryan, 2018). Lastly, captured iguanas must be handed over to pest and critter control for eradication and control purposes, yet a permit is required to possess live-captured iguanas, which are difficult to obtain (Bryan, 2018). These strict regulations on controlling the iguanas continuously prove to be burdensome as these animals are posing threats to not only the environment but also to the economy.

Environmental Limitations

In addition to the strict regulations for killing iguanas, there are also lengthy limitations that differ among cities in Florida on the proper removal of the deceased animal. For instance, once an iguana is humanely killed, it is prohibited to dispose of the iguana's body in a road or waterway (Geggis, 2018). Depending on different cities, it may also be illegal to put the animal remains in the trash. In the city of Hallandale Beach, in order to remove a deceased iguana, homeowners must contact a private pick-up service (Bryan, 2018). For example, in the city of Davie, dead animals including iguanas can be placed in the trash, however, in the city of Hollywood, the city considers iguanas to be "domestic animals" and does not let homeowners put them into the trash (Bryan, 2018). This leaves three additional options in all cities, the burial of the animal at least 1m underground, cremation at a pet cemetery (for a fee), or hiring a local critter removal company to discard the carcass (for a fee; Bryan, 2018).

Economic Limitations

Despite the issues faced by city regulations regarding iguana removal, another pressing concern involves the United States economy. With the U.S. facing hardships in the maintenance of regulations and the funding for infrastructure (Nallathiga, 2013), combined with these rapidly burrowing iguanas (FWC, 2021), the expediting of many more infrastructural issues will be encountered in the years to come. As current infrastructure is mainly outdated and maintenance costs are on the rise (Nallathiga, 2013), civil engineers are advocating for change due to the rise in safety concerns. These concerns include structural corrosion and failure that affects drinking water provided to the public (Council on Foreign Relations, CFR 2021), which poses a risk to the public's health (CFR, 2021). This can cause major issues for not only the health of the public but also can cause structural failure to important and frequently used public highways, bridges, and buildings. Both permanent or temporary closures of these resources for maintenance or conservation would create complications with transportation and traffic control (CFR, 2021). Attempts are currently being made to help combat the need for federal aid to help maintain infrastructure, such as President Biden's American Jobs Plan (FACT SHEET: The American Jobs Plan, 2021), however it lacks the focus on targeting invasive species and their effect on the environment.

Regardless, President Biden has proposed \$600 billion in funding to improve the infrastructure of airports, bridges, highways, roads, and even water systems (FACT SHEET: The American Jobs Plan, 2021). The American Jobs Plan mainly targets the working class as the major idea is to create more jobs in America. This would greatly help the issue

currently faced with the deterioration of infrastructure due to invasive species. Yet the inconsistencies in this plan begin when the majority of this allocated money will go towards training workers, manufacturing facilities, and research (FACT SHEET: The American Jobs Plan, 2021). This plan predominantly illustrates how funding will be used to create jobs which will hopefully help fix the problems with our infrastructure. Although this shows the great potential of the plan as a step in the right direction, it is not enough to target the specific issues of eliminating the threat of invasive species. As of August 10, 2021, part of this infrastructure bill has passed through the senate, yet there is no guarantee that it will help target these specific issues regarding invasive species.

4. Discussion

With the extensive regulations on iguana removal, there are a few solutions that have yet to be proposed. For example, using a fast-acting poison that only affects iguanas in areas that are being targeted could easily help decrease population and growth. An example of this can be found in Antarctica, where invasive rodents were eradicated. To accomplish this, scientists created a rodenticide bait containing brodifacoum to help remove the species (Springer, 2016). An unintended consequence of this solution, if used in Florida, could be the poison entering the water supply, which might pose a risk to human A commonly proposed solution is health. human-introduced species. Human-introduced species can specifically target a category of animals, such as the iguana, yet once the iguana population becomes diminished, this newly introduced animal or plant could become harmful for other species as well. For example, one of the only animals that are predators to iguanas are eastern racer snakes (Coluber constrictor), which are endemic to South Florida (Oxley, 2018). This poses an issue not only because they are extremely dangerous animals, but iguanas and racer snakes live in two completely different environments (Oxley, 2018). This would make it hard for the snakes to target all iguanas, as the iguanas live not only in parks and rural areas, but in residential areas as well, where the snakes cannot

thrive. This technique has been used on several islands where there were feral cats which were responsible for playing a big part in the disappearance of many native and global species. By human-introduction of rats into the area, the cats became more likely to target rat species rather than native plants or animals (Nogales, et al., 2004). These rats were also poisoned, which killed the cats after they consumed the rats (Nogales, et al., 2004).

Other than proposed environmental solutions, there are also economic propositions that could fight the negative effects of iguanas. As discussed earlier in this paper, attempts are being made to help increase funding for infrastructure maintenance and projects (Nallathiga, 2013). However, in order to directly tackle the issues that infrastructure and the environment face, policymakers need to change their focus to include adding new job categories that focus more on invasive species and their effects on the environment. This would greatly help when proposing new budgets and legislation for the maintenance and protection of infrastructure as it would include a section dealing with invasive species.

Future studies should include focusing on how climate change could affect the range of the iguanas. Currently they have a lower temperature tolerance of 7°C, restricting them to South Florida (National Geographic, 2012). An increase in local and global temperatures (Zaval et al., 2014) could increase their range across Florida and nearby states. This change in distribution could cause more widespread environmental and economical issues in new areas.

5. Conclusion

The rapid damage of iguanas in South Florida continues to deteriorate and affect infrastructure security and wildlife conservation. Mitigation of these invasive iguanas will bring massive improvements in the environment. Full removal of the species will make sure that burrowing, which affects structural components of infrastructure, is reduced, yet it will not be able to fix the current damage. The downside to not changing anything about the problems created by invasive species is that they will continue to harm the ecosystems in which they live. This could also potentially force already problematic structural issues to become even more dangerous and hazardous to humans as well as wildlife. Differences in regulations across three different South Florida cities make killing and disposal of iguanas a lengthy and tiresome process. These added barriers make it complicated for residents to adhere to the calls of the FWC to eliminate the iguana population. With current proposed bills for infrastructure development, it seems as if there have been advances but it is not directly targeting problems that invasive species generate. To reduce this problem, new job categories should be introduced to specifically study the issues that invasive species cause. With this newly introduced job category and its research, when bills are drafted, policymakers will have more background on the source of many infrastructure problems. As recent invasive iguana populations are on the rise, threats to the environment and the economy will grow. An increase in resources for invasive species damage mitigation will result in a reduced inverse effect on the environment all while increasing public awareness.

Acknowledgements

I would like to thank my mentor Malia Smith for advising and guiding me through the process of researching, writing, revising, and publishing my paper. I would also like to thank Polygence and my family for providing me with the opportunity to accomplish this project.

References

Bryan, S. (2018, December 12). *That pesky iguana may be dead, but your troubles aren't over yet.* sun. https://www.sun-sentinel.com/local/fl-ne-iguanas-dea d-legal-removal-20181011-story.html.

Chakraborty, J., Collins, T. W., Montgomery, M. C., & Grineski, S. E. (2014). Social and Spatial inequities in exposure to flood risk in Miami, Florida. *Natural Hazards Review*, *15*(3), 04014006. https://doi.org/10.1061/(asce)nh.1527-6996.0000140 Childs, J. W. (2020, January). *Burrowing iguanas helped cause \$1.8 million in damage in one Florida town: The weather channel - articles from The* *Weather Channel*. The Weather Channel https://weather.com/science/nature/news/2020-01-23florida-burrowing-iguanas-18-million-damage-west-p alm-beach

Council on Foreign Relations. (n.d.). *The state of u.s. infrastructure*. Council on Foreign Relations. https://www.cfr.org/backgrounder/state-us-infrastruct ure.

Do iguanas lay eggs? Critter Control - Since 1983. (n.d.).

https://www.crittercontrol.com/wildlife/iguana/do-igu anas-lay-eggs.

Eddmaps. (2021). Early Detection & Distribution Mapping System. The University of Georgia - Center for Invasive Species and Ecosystem Health. Available online at https://www.eddmaps.org/florida/distribution/viewma p.cfm?sub=12119&lat=27.230210&lng=-80.551758 &zoom=9

Electrocuted iguana knocks out power at Florida Nursing Home- Miami Herald. Miami Herald. (2018). Retrieved February 14, 2022, from https://www.miamiherald.com/news/local/community /broward/article219241815.html

Geggis, A. (2018, December 15). *How to kill an iguana (legally)*.Sun. http://www.sun-sentinel.com/news/florida/fl-sb-iguan a-handling-issues-20180131-story.html.

Green iguana. Florida Fish And Wildlife Conservation Commission. (n.d.). https://myfwc.com/wildlifehabitats/profiles/reptiles/g reen-iguana/.

Green iguana: National geographic. Animals. (n.d.). https://www.nationalgeographic.com/animals/reptiles /facts/green-iguana.

Guardian News and Media. (2019, July 4). *Florida residents urged to kill iguanas 'whenever possible'*. The Guardian.

https://www.theguardian.com/us-news/2019/jul/04/fl orida-residents-urged-to-kill-iguanas-whenever-possi ble.

Iguana care Tips: Reptile Vet: Long island bird & exotics. libirdexoticsvet. (n.d.). https://www.birdexoticsvet.com/iguana-care-tips#:~:t ext=The%20lifespan%20of%20an%20iguana,live%2 0more%20than%2020%20years.

Invading iguanas tear through Florida's ecosystem.(www.dw.com), D. W. (n.d.). DW.COM. https://www.dw.com/en/invading-iguanas-tear-throug h-floridas-ecosystem/a-43148878.

Invasive iguanas chewing their way through South Florida infrastructure landscape. The Weather Channel. (n.d.). https://weather.com/science/nature/news/2018-06-26south-florida-green-iguanas-wreaking-havoc-damage.

Invasive species. National Wildlife Federation. (n.d.). https://www.nwf.org/Educational-Resources/Wildlife -Guide/Threats-to-Wildlife/Invasive-Species.

Nallathiga, Ramakrishna. (2013). Infrastructure Development: Issues and the Way Forward.

National Geographic Society. (2012, October 9). *Invasive species*. National Geographic Society. https://www.nationalgeographic.org/encyclopedia/inv asive-species/.

Oxley, C. (2018, August 8). *Invasive species in Galapagos*. Galapagos Conservation Trust. https://galapagosconservation.org.uk/invasive-species /.

Review of feral cat eradications on islands. (n.d.). http://www.issg.org/pdf/publications/Island_Invasive s/pdfHQprint/1Campbell.pdf.

Smith, H. T., Sementelli, A. J., Engeman, R., & Meshaka, W. E. (2008). *Just green Iguanas?: The associated costs and policy implications of Exotic invasive wildlife in South Florida*. SAGE Journals. https://journals.sagepub.com/doi/abs/10.1177/108772 4x08316157.

Springer, K. (2016). Methodology and challenges of a Complex multi-species eradication in The sub-antarctic and immediate effects of invasive species removal. *New Zealand Journal of Ecology*, *40*(2), 273–278. https://doi.org/10.20417/nzjecol.40.30

Townsend, Josiah & Krysko, Kenneth & Enge, Kevin. (2003). Introduced Iguanas in Southern Florida: A History of More Than 35 Years. U.S. Fish and Wildlife Service, J. E. (n.d.). *U.S. Fish* & *Wildlife Service*. FAQs - Invasive Species - U.S. Fish and Wildlife Service.

https://www.fws.gov/invasives/faq.html#:~:text=How ever%2C%20of%20that%2050%2C000%20species,c onsidered%20invasive%20species%20.

The United States Government. (2021, May 4). *fact sheet: The American jobs plan*. The White House. https://www.whitehouse.gov/briefing-room/statement s-releases/2021/03/31/fact-sheet-the-american-jobs-pl an/.

Zaval, L., Keenan, E. A., Johnson, E. J., & Weber, E. U. (2014). How warm DAYS increase belief in global warming. *Nature Climate Change*, *4*(2), 143–147. https://doi.org/10.1038/nclimate2093

The Neglected Equine Agency in "Stopping by Woods on a Snowy Evening"

Yuheng Wang^{1*}

¹ Beijing No. 101 High School, Beijing, People's Republic of China

Received January 27, 2022; Revised March 2, 2022; Accepted, March 14, 2022

Abstract

This article explores Robert Frost's "Stopping by Woods on a Snowy Evening" from the posthumanist perspective of equine agency and argues that the narrator's actions represent man's belief in anthropocentric superiority. The horse, through his exquisite kinetic senses, perceives the impending danger that awaits in the freezing woods and advises the narrator to reverse course. The narrator, however, obtusely misinterprets the message, believing the horse to be incapable of comprehending aesthetics and promises, reaffirming human superiority over the equine species. Under closer scrutiny, however, this anthropocentric vanity crumbles. First, human beings, while privileged to observe the world through a keen sense of vision, are led astray in their professed ability to recognize beauty. For example, the narrator remarks on the beauty of the "dark, deep" woods, but in doing so his visual penetration is blocked and confined within a superficial level. Then the speaker's lofty promises to travel miles before stopping to rest acquire an ironic touch. Instead of exploring an uncharted territory, the speaker embarks upon the same trips within a familiar region. Given his frequent travels on well-traveled paths, if the narrator still has a long way to go after evening, he has poorly planned his trip. Rather than acknowledging his blunder, however, the narrator boastfully promises, "And miles to go before I sleep," and the repetition of the last line reveals his underlying anthropocentric vanity.

Keywords: Robert Frost, the Horse, Anthropocentric Superiority, Aesthetics, Promises

Stopping by Woods on a Snowy Evening	He gives his harness bells a shake
By Robert Frost	To ask if there is some mistake.
	The only other sound's the sweep
Whose woods these are I think I know.	Of easy wind and downy flake.
His house is in the village though;	
He will not see me stopping here	The woods are lovely, dark and deep,
To watch his woods fill up with snow.	But I have promises to keep,
	And miles to go before I sleep,
My little horse must think it queer	And miles to go before I sleep.
To stop without a farmhouse near	
Between the woods and frozen lake	1. Introduction
The darkest evening of the year.	
	"Stopping by Woods on a Snowy Evening," the
Corresponding Author	Advisor: George Lazo

yuhengwang2020@126.com

Advisor: George Lazo george.lazo@beijing101.com.cn most memorable poem of Robert Frost, has a never-waning appeal. Even "my little horse" has received extensive scholarship, albeit overlapping in nature. The horse shakes its harness bells of social "obligations" to keep the dreamy poet awake (Henry, 1978, p.38). Similarly, the equine embodies "social condemnation" for the guilty poet who takes a break in a workaholic society (Frank, 192, p. 44). Richard Moore, in comparison between Frost's horse and Swift's Houyhnhnms, maintains that the animal is "rational, predictable, mechanical" (Moore, 2000, p. 96). Few critics have paid attention to the horse as a horse per se. How does the horse perceive the woods? different understandings, From their what relationship between animals and human beings is revealed? This article explores the poem from the perspective of posthumanism. Posthuman theory, according to Rosi Braidotti, mainly consists of "becoming-animal, becoming-earth and becoming-machine." These branches have different accentuations but they all share the same view, that is, to contest against "arrogance of anthropocentrism and the 'exceptionalism' of the Human as a transcendental category" (Braidotti, 2013, p.66). As far as animals are concerned, Cary Wolfe proposes an enlightening approach to gain apprehension of nonhuman species. "To 'understand' the animal," Wolfe writes, we have to "stand 'under,' not above her-by surrendering the dream of mastery troped as vision" (Wolfe, 2003, p.5). Stepping back from an anthropocentric view enables men to have a more accurate understanding of human positioning in the world and a better opportunity to learn about agency of nonhuman species. This leads to a paradox: acknowledging nonhuman agency broadens human knowledge but it simultaneously challenges human vanity and desire for superiority. "Stopping by Woods on a Snowy Evening" reveals the narrator's unconscious process of struggling to resolve this tension. According to this essay, the poem demonstrates the speaker's frustrated efforts to establish anthropocentric superiority over the horse through aesthetic appreciation and social promises.

2. The Neglected Horse in the Poem

The little horse plays an essential role in the poem

but it is often neglected by readers. If we examine the world from the equine perspective, we will be enlightened by the poem's nuances. "My little horse" has a distinctive perception of the scene from the speaker "I." The structure of "Stopping by Woods" is crafted to present the first and last stanzas from human perspective, but to narrate the second and third quatrains from animal viewpoint, beginning with "My little horse must think it queer" (Frost, 1995, p.207). Horses are known for their exquisite kinesthetic sensitivity. "Every muscle twitch of the rider will be like a loud symphony to the horse" (Hearne, 1987, p.108). The horse's extremely fine sense of touch finds its full expression in the poem. "Feathery flake," the usual English expression, accentuates the size of snow, while "downy flake" emphasizes the tactile quality of softness of snow. Flake, a piece of something especially "that has broken off from something larger" (Wehmeier, 1997, p. 768), suggests that a single flake of snow is connected with a mass of snows.

Likewise, "easy wind," the reference to gentle breeze, not only underscores the tactual feeling of balmy air but also conveys an equine evaluation, but what is easy? The combination of "easy wind and downy flake" is "the sweep," which according to Oxford Dictionary, means "move quickly with force" (Wehmeier, 1997, p. 2041). "[T]he sweep of easy wind and downy flake," neither quick nor forceful, is contradictory to human logic, but not to the horse's delicate sense of touching. The imperceptible accumulation of falling snow under continual wind during a long period is likely to result in potential danger of massive snowballs. The word "up" in "his woods fill up with snow" has already indicated an enormous amount of snows that has almost filled up the woods. The horse intuitively senses the hidden peril and becomes aware of human negligence: it is "easy" for men to relax their caution under the impression of "easy wind and downy flakes." Therefore "He gives his harness bells a shake" to remind the master of the hazard. Before that, the horse had already given a caveat to the narrator. Resonating with the aggrandizement of fragmentary flakes, the "frozen lake" has unified water into a smoothing wholeness. "Frozen" also indicates the rapid decrease of temperature after winter evening

and the subsequent freezing coldness. Thus the horse expresses its disapproval of the speaker's "queer" activity. The message is taken up, but not taken in, because the poet misinterprets its motivation: "My little horse must think it queer / to stop without a farmhouse near." In the mind of the speaker, animals could not raise themselves above basic life necessities: a farmhouse to access food and shelter. Therefore the complicated speculations of the animal are monotonously reduced to the "lower, practical, instrumental reason" of simple utilitarian needs (Coetzee, 1999, p. 29).

3. The Speaker's Vanity of Establishing Anthropocentric Superiority

Furthermore, the narrator endeavors to establish human superiority through aesthetic appreciation. The purpose of devaluing animals is to establish anthropic superiority by aesthetics. "He will not see me stopping here / To watch his woods fill up with snow" (Frost, 1995, p.207). Artistic admiration, a utility-transcending activity, is regarded as a privilege of human mind. The ascendancy of visual stimuli from olfactory stimuli, a line of demarcation to distinguish human beings from animals, leads to "the threshold of human culture" (Freud, 1961, pp. 51-2). Man, in his "upright gait," begins to be distinguished by aesthetic sensitivity and distant contemplation. Likewise, Una Chaudhuri also holds a similar idea. One prominent feature of Human beings is "the privileging of sight in the human evolution from quadruped to biped," which is reinforced by "the derogation of the 'lower senses' (touch and smell)" (Chaudhuri, 2007, p.11). The importance of human sight, highlighted by such verbs as "see" "watch," culminates in visual pleasure: "the woods are lovely." However, the woods are also "dark and deep." "Dark" suggests the impenetrability of vision, and "deep" allows a superficial sight but blocks further insight. The conjunction of "dark and deep" rouses the feeling of uneasiness, even anxiety, canceling the momentary aesthetic pleasure. Therefore, human superiority is traumatically thwarted.

To reestablish human superiority, the narrator has ushered a new vantage: promises. "But I have promises to keep." Promises are a unique and exclusive human phenomenon because a man's honor depends on his word which in turn defines his responsibility. Promises include personal ones and "those that my ancestors made for me, known as the social contract" (Frost, 1995, p. 823). The one in the poem is "And miles to go before I sleep, / And miles to go before I sleep" (Frost, 1995, p.207). The literary repetition of poetic "miles to go" produces an impression of a far-away odyssey while the placement of "miles to go" in the context of metaphorical "before I sleep" suggests an exploration of risky and unknown territories. Richard Gray also holds that the poet feels compulsive to "take risks and explore the truth" in his "spiritual quest" (Gray, 2012, p.352).

However, under scrutiny this lofty mission crumbles. First, the poet does not explore a foreign land, and he just makes usual trips in a familiar region. He knows "whose woods these are" and is even acquantant with the owner whose "house is in the village" (Frost, 1995, p.207). Given his frequent travels in familiar surroundings, if he still has a long way to go after evening, he has ill-planed his trip; he is too unrealistic to fulfill his pompous promise. If it is not so long, the repetition of "miles to go" is his bombast of trivial walks back and forth between the village and market; he is prone to exaggerations of Either way, his promise or petty swears. responsibility, is severely undermined. In contrast, the equine, as "my little horse," demonstrates its due responsibility towards the master. Intuitively convinced of the potential hazard, the horse cautions a "queer" expression to the poet who does not understand it. Then the animal hardens its attitude to show that "there is some mistake." Moreover, the horse adjusts its way of communication with human beings and skillfully employs a man-made instrument to convey its message: "He gives his harness bells a shake." The sound reminds the poet of his "promises" and drives him to continue the journey. Twice embarrassed by the intelligent horse, the poet hastes to mechanically repeat a forced closure to stabilize his human identity: "And miles to go before I sleep, / And miles to go before I sleep."

In short, this article explores the kinetic world from the perspective of the horse and challenges the assumed superiority of the human narrator in "Stopping by Woods on a Snowy Evening." The decentering of the anthropocentric gaze, along with due attention to nonhuman agency, initializes an indispensable step to appreciate the full complexity of the natural world that frequently foregrounds Frost's poetry.

Acknowledgement

We would like to express our thanks for the anonymous reviewer and the editor of *Journal of Research High School* whose constructive comments have greatly improved the quality of this paper.

References

Braidotti, R. (2013). The Posthuman. Polity Press.

Chaudhuri, U. (2007). (De)Facing the animals: Zooësis and performance. *TDR: The Drama Review*, 51(1), 8-20.

Coetzee J.M. (1999). *The Lives of Animals* (A. Gutmann, Ed.). Princeton UP.

Frank, B. (1982). Frost's "Stopping by Woods on a Snowy Evening." *The Explicator*, 40 (4), 43-45.

Freud, S. (1961). *Civilization and its discontents* (J. Strachey, Trans and Ed.). Norton.

Frost, R. (1995). *Collected poems, prose, and plays*. Library of America.

Gray, R. (2012). *A history* of *American literature* (2nd edition). Blackwell Publishers Ltd.

Hearne, V. (1987). Adam's task. Random House.

Henry, N. (1978). Frost's 'Stopping by Woods on a Snowy Evening." *The Explicator*, 37(1), 37-38.

Moore, R. (2000). Frost's 'Stopping by Woods on a Snowy Evening' and out, out--, and Swift's *Gulliver's travels. The Explicator*, 58 (2), 95-97.

Wehmeier, S, editor-in-chief. (1997). Oxford advanced learner's English-Chinese dictionary (7th edition). The Commercial Press.

Wolfe, C. (2003). Animal rites: American culture, the

discourse of species, and posthumanist theory. University of Chicago Press.



Breakthroughs in Honey Bee Health: Continuous-Release Mist Diffusion of Thymol-Based Essential Oils

Kaitlyn N. Culbert^{1*}

¹ Toms River High School North, Toms River, NJ USA

Received September 1, 2021; Revised March 5, 2022; Accepted, March 17, 2022

Abstract

Honey bee (*Apis mellifera*) pollination is responsible for approximately 80% of all cultivated crops. Unfortunately, reports suggest losses of 30-50% of all bee colonies in the US. The greatest single contributor to the decline of bee health is the Varroa mite. Synthetic chemicals are currently used to control Varroa, but the mites are developing methods of resistance. Essential oils (EOs) may be a viable alternative. EOs are cheaper, environmentally-friendly, and pose fewer health risks to bees and consumers. Furthermore, Varroa have not yet developed resistance to EOs. EOs' shortcoming is the limitation of exposure. Humidity and temperature affect the rate of evaporation and the mites' exposure to the EOs. Thymol-based essential oils dispersed via a battery-operated mist diffuser would provide effective miticide efficacy without causing harm to honey bees due to the natural miticide properties of the EOs and the ability of the mist diffuser to maintain a constant temperature and humidity. Utilizing a gel capsule system to rear larvae/pupae/Varroa, thymol-based EOs were found to be effective against Varroa without harming bees. Miticide activity was as follows: thyme>oregano>rosemary>spearmint> control (vegetable glycerin). Currently, all commercially available thymol-centered systems are gel-based and work by direct contact with the mite. The use of mist diffusers effectively eliminates fluctuations in temperature and humidity and was more cost-effective than commercially available thymol-based systems (US\$3.20 versus US\$15-\$18 per application). Continuous-release mist diffusion permits the disbursement of EOs throughout the entire hive, perhaps including the brood cell where natural miticides do not reach and where *Varroa* reproduction occurs.

Keywords: Varroa destructor; Apis mellifera; mites; honey bees; essential oils; mist diffusion

1. Introduction

Honey bees, *Apis mellifera*, are crucial for agriculture, responsible for over 80% of all cultivated crops in the United States (Randall, 2020). In fact, bee pollination accounts for approximately US\$15 billion in added crop value (Bryant, 2006). Additionally, bees also produce honey, pollen, royal jelly, beeswax, propolis, and venom for nutritional

* Corresponding Author culbertkaitlyn@gmail.com and medicinal uses. In the US, there are approximately 2.5 million commercially farmed honey bee hives and around 500,000 colonies kept by hobbyists and semi-professional bee keepers (Frazier, et al., 2013). Unfortunately, these numbers are declining at a rapid rate. Various reports suggest a loss of 30% to 50% of bee colonies in the US, (Colony Collapse Disorder, 2021), its lowest point in the past 50 years. The greatest single contributor to

Advisor: Christine Girtain cgirtain@trschools.com

the decline of bee health is the *Varroa destructor* mite (Rosenkranz, et al., 2010). 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* weakens and ultimately kills colonies by out-reproducing their host.

Varroa is an ectoparasitic mite that exploits the honey bee's life cycle. It feeds on the fat bodies of developing honey bee larvae and adult bees and aggressively reproduces within an infected bee colony. Recent research by Ramsey, Ochoa, Bauchan (2019) brings to light the *Varroa* mite's focus on the fat body tissue, and not the hemolymph as previously believed, which ultimately render the honey bee susceptible to harm from pesticides. *Varroa* mites also facilitate the transmission of multiple viruses and other diseases which have been linked to honey bee larvae develop into pupae inside the individual brood cell of the wax comb and emerge as adults.

Female Varroa mites attach onto an adult or immature honey bee and enter the bee hive's brood cells to reproduce. The female Varroa punctures the underside of the abdomen of the pupae and feeds on the fat body cells before laying the first egg. Like honey bees, the male Varroa develops from an unfertilized egg. The female continues to lay about one egg per day. These eggs develop into daughters that mature and mate with the male before the bee emerges from the brood cells. Immature and male Varroa can only be found in the capped brood cell. Only the mature Varroa female will survive after leaving the brood cells. All immature mites and Varroa males die soon after the adult bee emerges. The female mite then repeats the cycle by entering cells of other developing larvae (Hunt, & Given, 2021).

Bahreini, et al., (2020) clarify, in order to control the spread of *Varroa*, bee keepers initially used acaricides, pyrethroids, and organophosphates pesticides. Unfortunately, these chemicals have been found to increase the early replacement of the queen, heighten mortality in adults and brood, reduce body weight in queens, decrease the amount of lipids, carbohydrates and proteins in workers. Furthermore, these synthetic compounds have also been found to accumulate in the beeswax which impacts the development of bee larvae (Bahreini et al., 2020). The frequent use of these synthetic miticides to control *Varroa* infestations has resulted in the development of resistance to many of the chemical components of these miticides (Bahreini et al., 2020; Traynor, et al., 2016). Likewise, the excessive or improper use of synthetic compounds has also resulted in the contamination of hive products which pose a health threat to both hive bees and human consumers (Traynor et al., 2016). For these reasons, the need for alternative *Varroa* control measures is critical.

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). Essential oils are concentrated hydrophobic liquids extracted from aromatic plants and capture the "essence" or characteristic odor of the plants. Only plants containing more than 0.1% oil can be called essential oils (Imdorf, et al., 1999). The essential oil composition of each plant is unique. Terpenes (mainly monoterpenes) are the main components of essential oils. A wide variety of essential oils (over 150) have been tested as potential miticides; unfortunately, very few of them have proven successful when tested in field trials (Sabahi, et al., 2017).

A key problem evidenced in trials conducted to evaluate essential oils as miticides in hives is the consistency and significant variability that exists between seasons and localities. Sabahi et al. (2017) concluded that one factor affecting this variability is the pattern of climatic conditions which are dependent upon the ambient temperature and relative humidity which can affect the properties of essential oils such as the rate of evaporation of the oils and, consequently, the mites' exposure to the essential oils.

Currently, numerous essential oil compounds have been evaluated for miticidal activity. One of the proven successful essential oils is thymol. Thymol is a phenolic monoterpene that is present in many plants such as thyme, basil, rosemary, mint, and sage. It works by disorienting the mite and blocking its pores (Tennessee's Honey Bees, 2021). The efficacy of thymol has previously studied and has been recommended to be effective against Varroa. Thymol is the only compound of essential oils widely used in beekeeping with 70%-90% efficacy against Varroa (Garrido, 2018). Thymol also has the added advantage of being active against fungus (chalkbrood) and some efficacy with tracheal mites (Davis, & Ward, 2003). The most widely used and popular products with thymol as a main ingredient are Apiguard®, ApiLifeVar® and Thymovar®. These systems only kill mites on the adult bees, but do not reach the reproducing Varroa mites in the brood cells (Garrido, 2018).

In light of the current crisis surrounding honey bees and the *Varroa* mite, new solutions are desperately needed. The following essential oils: thyme, oregano, spearmint, *and* rosemary were evaluated for miticide efficacy and safety. The use of a battery-powered mist diffuser to deliver essential oils may allow for the continuous-release of thymol regardless of environmental conditions, perhaps reaching the brood cell where *Varroa* reproduction takes place.

2. Materials and Methods

2.1 Safety Precautions

Appropriate attire and proper safety and handling precautions were taken during the collection and handling of *Varroa* mites, honey bee larvae and pupae. Appropriate attire consisted of a complete beekeeping suit which includes: hat with veil, full body suit, and foot gear. Mentors from Stockton and Rutgers provided guidance and oversight during the handling and collection process.

2.2 Varroa Mite Collection

A. mellifera colonies with bee hives naturally infested by *Varroa destructor* mites were utilized for this study. Live *Varroa* were collected from Stockton University from hives in Galloway, NJ and from Rutgers University from hives in Eastampton Township, NJ. All *Varroa* collection was conducted on October 10, 2020 to coincide with increased *Varroa* populations in the fall/early winter and to minimize interference with honey production during the spring.

For this experiment, live *Varroa* was collected from the powdered sugar roll methodology as described by the Honey bee Health Coalition (2016). However, the roll/shake was performed over a white paper plate and sprayed with a spray bottle containing water to dissolve the powdered sugar. A paintbrush was utilized to collect all live *Varroa*. All collected live *Varroa* were placed into a clean, damp paper towel. The towel was folded up gently with the mites inside and placed into a petri dish for transport.

2.3 Larvae/Pupae Collection

Fourth-instar worker bee larvae and capped brood larvae were collected from established A. mellifera colonies at Stockton University from hives in Galloway, NJ on October 10, 2020. Following larvae collection methods as described by Jack, et al., (2020), honeycomb brood frames were placed into an incubator at 94 degrees Fahrenheit (34.4 degrees Celsius) at 65% relative humidity (RH). After 3 hours, larvae crawled out of the cells and paper was placed underneath the combs to catch the larvae. Each larvae was carefully examined for the presence of attached mites. For the capped brood, cells were uncapped and brood was cautiously removed with a pair of fine tipped forceps as described by Honey Bee Health Coalition. Pupae were carefully checked for the presence of mites. All larvae and pupae were examined with a probe, all reacting larvae and pupae were scored as live (Honey Bee Health Coalition, 2016).

2.4 Gelatin Capsule System

As described by Jack et al. (2020), all collected live larvae and pupae were placed into size "0" gelatin capsules (7 mm, Now Foods, Bloomingdale, IL, USA). Using a number 2 insect pin (BioQuip, Rancho Dominguez, CA, USA), small ventilation holes were made in the capsules to simulate the porous capped brood cells. *Varroa* females collected from the powdered sugar roll system were individually inserted into a gelatin capsule containing a larva or a pupa. Each capsule contained a larva/pupa and one or two *Varroa* depending on availability. The capsules were then placed vertically into an empty micro-pipette tip container and maintained in an incubator (FBA Magicfly-MF014, Hong Kong, China) at 94 degrees Fahrenheit (34.4 degrees Celsius) at 75% RH. According to Jack et al. (2020), the gelatin capsule system is reported to have a mite survival rate of >95%. The temperature and humidity were monitored with the Broodminder mobile app (Broodminder, Stoughton, WI, USA).

2.5 Essential Oils and Continuous-Release Battery-Operated Mist Diffuser

For this experiment, four USDA, certified organic, premium food-grade essential oils, thyme, oregano, spearmint, and rosemary, were evaluated along with a control consisting of organic, premium food-grade vegetable glycerin (Plant Guru, Plainfield, NJ, USA).

Continuous-Release Treatments 1,2,3,4 consisted of essential oils (1-thyme, 2-oregano, 3-spearmint, 4-rosemary) for continuous-release via а battery-powered mist diffuser (AirWick® Essential Mist Essential Oil Diffuser, Reckitt Benckiser, Slough, England). Each individual incubator (FBA Magicfly-MF014, Hong Kong, China) contained one of the tested essential oils in a continuous-release battery-powered mist diffuser (or vegetable glycerin only in the control) along with the gelatin capsule system (Jack et al., 2020). This methodology is designed to simulate the brood cell environment within the hive. Sabahi et al. (2017) evaluated oxalic acid, oregano/clove, and oregano alone utilizing a continuous-release electric diffuser. However, in light of the potential fire safety concerns and the lack of accessible electrical outlets in the field, the mist diffuser for this experiment was battery-powered.

To provide continuous-release, two alternating mist diffusers were used in each incubator to provide 24/7 release as each mist diffuser automatically shuts down after eight hours of continuous use. The AirWick Mobile App was utilized to ensure that all diffusers were operating properly. To calculate the rate of evaporation of the essential oil, the mist diffuser container was weighed while empty and then

weighed again when filled with the essential oil at the beginning (the net weight was obtained by subtracting the tare weight from the gross weight) and at the end of the trial. Control Treatment 5 contained no essential oil, only vegetable glycerin. Cumulative mite and bee survival was recorded at 24, 48, 72, 96, 120, 144, 168, 192, and 216 hours.

In total, 88 live larvae/pupae and 91 live mites were collected from the apiaries at Stockton University and Rutgers University. After 12 hours, all collected mites and larvae/pupae were assessed for movement as a measurement of lividity using fine tipped forceps. In total, 56 live larvae/pupae and 63 live mites were recovered to commence the experiment. Each live larvae/pupae was separated into its own gel capsule and 1 (or 2 mites) was allocated to each larvae/pupae.

3. Results and Discussion

The recorded mite and bee mortality from the diffusion of tested essential oils in 24-hour increments are provided in Table 1.

E.O. Tested	Time (hours)	# of Fallen	# of Fallen
		Varroa	Larvae/Pupae
	24	5	0
	48	13	0
	72	13	0
	96	13	0
Thyme	120	13	0
ingine	144	13	5
	168	13	9
	192	13	11
	216	13	11
	240	13	11
	24	3	0
	48	13	0
	72	13	0
Oregano	96	13	0
	120	13	0
	144	13	4
	168	13	8
	192	13	11
	216	13	11
	240	13	11

Table 1: Number of fallen *Varroa* and fallen larvae/pupae by essential oil tested

	24	0	0
Casaratint	48	0	0
	72	1	0
	96	5	0
	120	8	0
Spearmin	144	9	3
	168	12	5
	192	12	8
	216	12	11
	240	12	11
	24	0	0
	48	0	0
	72	3	0
	96	9	0
Posomany	120	13	0
Rosemary	144	13	2
	168	13	7
	192	13	10
	216	13	10
	240	13	11
	24	0	0
	48	0	0
	72	0	0
	96	1	0
Control	120	2	0
Control	144	3	2
	168	6	4
	192	10	7
	216	13	11
	240	13	12
	Time to last fallen Varroa		
	Time to last fallen larvae/pupae		

Across all essential oils, an increase in the hours of exposure led to an increase in *Varroa* mortality as depicted in Figure 1. After 48 hours, all *Varroa* mites in both the thyme and oregano had completely fallen. After 120 hours, all mites exposed to the rosemary had succumbed. After 144 hours, all mites exposed to spearmint had fallen. For the control, containing only vegetable glycerin, after 240 hours, all mites had succumbed. The overall miticide activity from the first recorded mite fall to last recorded mite fall was as follows in terms of essential oil efficacy: thyme > oregano > rosemary > spearmint > control. Statistical significance was determined using regression analysis. All p-values were statistically significant (p-value ≤ 0.05) across all mite falls (thyme (p=.00000651); oregano (p=.0000451); spearmint (p=.00391); rosemary (p=.000294)).



Figure 1. Percentage of *Varroa* mites fallen by essential oil per 24-hour periods

Across all essential oils, increasing the hours of exposure did not lead to an increase in honey bee mortality as depicted in Figure 2. Honey bee mortality was first noted at 144 hours for thyme, oregano, rosemary and spearmint. These values are comparable to the placebo which also noted honey bee larvae/pupae mortality at 144 hours. Overall bee mortality from first fall to last fall was as follows: thyme = oregano > rosemary > spearmint = placebo. All p-values were not statistically significant (p-value ≤ 0.05) across all bee larvae/pupae fall (thyme (p=.176); oregano (p=.594); spearmint (p=.255); rosemary (p=.452)).



Figure 2. Percentage of bee larvae/pupae fallen by essential oil per 24-hour periods

4. Discussion

This investigation of various essential oils demonstrates the effectiveness of thymol-based

essential oils and a battery-operated continuous-release mist diffusion delivery system as an effective miticide. All tested essential oils were found to be effective. The mist diffuser system was successful in its dispersal of the essential oils independent of temperature and humidity variables.

A key advantage of using essential oils in hives is that it is generally recognized as safe and environmentally-friendly. In fact, all the essential oils utilized for this experiment were premium food-grade. Due to the sustained lividity of the larvae/pupae from this investigation, the essential oil concentration tested is concluded to be safe for honey bees.

Furthermore. a brief analysis of the cost-effectiveness of the battery-operated continuous-release mist diffuser technology in concert with essential oils was found to be more economical than the annual cost of commercially available thymol products. The average cost per application of the continuous-release diffuser technology and essential oil is approximately US\$3.20 per hive per application (assuming the use of 2 diffusers) in comparison to currently available thymol-based commercial products at US\$15-US\$18 per hive per application.

5. Conclusion

The goal of this in vitro study was to demonstrate that continuous-release mist diffusion delivery of thymol-based essential oils can serve as an effective miticide while also being safe for bees. The results of this in vitro investigation warrant further examination of the continuous-release mist diffusion delivery system and the essential oils of thyme, oregano, and rosemary in a field investigation. Because the effectiveness of essential oils is dependent on multiple factors, including temperature, humidity, and time of year (Sabahi, et al., 2017), the use of a battery-operated continuous-release mist diffuser system minimized these variables and, perhaps, may allow for miticide exposure inside the capped brood cells. Currently, available thymol-centered systems are gel-based and work only by contact with the mite (Honey Bee Health Coalition, 2016). As technology evolves, other continuous-release devices (solar

powered diffusers) may also be investigated. Currently, no natural miticides penetrate into brood cells where *Varroa* reproduction takes place (Garrido, 2018). The use of thymol-based essential oils with a continuous-release mist diffuser delivery system may one day be incorporated with other control measures to effectively manage *Varroa* mites and other pests in order to enhance the chance of colony survival and residue-free hive products.

Acknowledgement

I would like to acknowledge Dr. Ron Hutchison for providing the *Varroa* mites and the honey bee larvae and pupae and Professor Michael Haberland for also providing the *Varroa* mites for my research. I would also like to thank Dr. Cameron Jack for his advice on *Varroa* collection and transport as well as setting-up the gel capsule system. Lastly, I also acknowledge the support of beekeepers Randy Oliver, Clifford Moore, and Geff Vitale.

References

Bahreini, R., Nasr, M., & Docherty, C. (2020). Evaluation of potential miticide toxicity to *Varroa destructor* and honey bees, *Apis mellifera*, under laboratory conditions. *Sci Rep* 10, 21529.

Bryant, D. (2006, February 16). USDA Bee Lab Teams Pursue *Varroa* Mite Controls. FarmProgress. January 22, 2021 from https://www.farmprogress.com/usda-bee-lab-teams-p ursue-varroa-mite-controls

Colony Collapse Disorder, US Environmental Protection Agency. Retrieved January, 2021, from https://www.epa.gov/pollinator-protection/colony-col lapse-disorder

Damiani, N., et al. (2009). Acaricidal and insecticidal activity of essential oils on *Varroa destructor* (Acari: Varroidae) and Apis mellifera (Hymenoptera: Apidae). *Parasitol Res.* Dec;106(1):145-52.

Davis, C., & Ward, W. (2003, December). Control of Chalkbrood Disease with Natural Products. Australian Government Rural Industries Research and Development Corporation. https://www.agrifutures.com.au/wp-content/uploads/p ublications/03-107.pdf Frazier, M. et al. (2013, June 13). *Beekeeping -Honey Bees*. Penn State Extension. https://extension.psu.edu/beekeeping-honey-bees Garrido, C. (2018, February 27). Thymol *Varroa* Control. Bee Culture. 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.* Oct;55(2):147-54.

Honey Bee Health Coalition. (2016, February 16). Tools for *Varroa* Management. Retrieved January 23, 2021, from

Hunt, G. J., & Given, K. Beekeeping What Beekeepers Should Know About Bee Mites. Purdue University Extension.

https://extension.entm.purdue.edu/publications/E-201 /E-201.html

Imdorf, A., et al. (1999). Use of Essential Oils for the Control of *Varroa jacobsoni Oud*. in Honey Bee Colonies. *Apidologie, Springer Verlag*, 30 (2-3), pp.209-228.

Jack, C. J., et al. (2020). Comparing four methods of rearing *Varroa destructor in vitro*. *Experimental and Applied Acarology*, 80(4): 463-476.

Moore, P. A., Wilson, M. E., & Skinner, J. A. (2019 August 20). Honey Bee Viruses, the Deadly *Varroa* Mite Associates. Bee Health, January 23, 2021 from https://bee-health.extension.org/honey-bee-viruses-th e-deadly-varroa-mite-associates/

Ramsey, S. D., 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. https://www.farmers.gov/blog/conservation/value-bir ds-and-bees

Rosenkranz, P., Aumeier, P., & Ziegelmann, B. (2010) Biology and control of *Varroa destructor*. J Invertebr Pathol 103(Suppl 1):S96–S119.

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.

Tennessee's Honey Bees. (2021). Three Primary Essential Oils are Useful in Beekeeping. https://www.tennesseeshoneybees.com/index.php?ma in_page=page&id=27&chapter=0

Traynor, K., Pettis, J., & Tarpy, D.(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.

Tools for *Varroa* Management. (2016, February 16). Honey Bee Health Coalition. January 23, 2021 from https://honeybeehealthcoalition.org/wp-content/uploa ds/2016/03/HBHC-Guide_Varroa_Interactive_BW_2 3FEB2016.pdf



Investigation of Vulnerability of Gwinnett County to Floods

Haelin Lee^{1,*}

¹ Gwinnett School of Mathematics, Science, and Technology, Lawrenceville, GA USA

Received February 7, 2022; Revised March 24, 2022; Accepted, April 1, 2022

Abstract

This research paper studies flooding dangers in Gwinnett County to discover the most susceptible areas to significant flood damage. Three different flood year recurrence interval flood scenarios (i.e., 100 year, 200 year, 500 year) in Gwinnett County were analyzed using HAZUS, outlining the floodplains and calculated flood depths of each scenario. These three different year recurrence intervals resulted in different flood depths for each scenario. This study also investigates the vulnerability of essential facilities, such as electric power facilities and water treatment plants. The vulnerable essential facilities in Gwinnett County were found with geological information provided by HAZUS. This investigation can be used as an initiative to develop a more resilient and safer Gwinnett County by developing greater preparedness and creating more organized responses to flood events and natural disasters.

Keywords: Gwinnett County, flood, essential facilities, vulnerability, HAZUS

1. Introduction

From September 16 to September 22, 2009, riverine flooding occurred in Georgia, caused by heavy rain and rainfall runoff, which is when rain continues to fall over an area for a long period of time. This case of riverine flooding resulted in nearly 500 million dollars' worth of damages, declaring 23 of 159 counties in Georgia to be disaster areas under the Presidential Disaster Declaration, FEMA-1858-DR (Federal Emergency Management Agency, 2018). Gwinnett County was one of these 23 counties, although Gwinnett County is located inland and is either north of or at the upper threshold of three of the major watersheds; upper Chattahoochee, Oconee, and Ocmulgee. Compared to the counties around it, Gwinnett County received rather lower damages and impacts from flooding, and along with

hhaelinlee@gmail.com

Cobb County, had 6 stream gages that measured floods which breached the 0.2% annual exceedance probability (i.e., 500-year recurrence interval), meaning that the chances of a flood of this large magnitude or greater happening again in this area is 0.2% (USGS, 2018). In the month September of 2009, an inordinate amount of rain fell on Northern Georgia. Over a hundred homes were decimated in just Gwinnett County during this period, also causing over a million dollars of damages to roadways and drainage systems. This is due to the fact that Gwinnett County had only been prepared for a 100-year flood (Flood Risk Management, 2019).

From 2000, there has been an increased number of flood-related natural disasters along the Atlantic Coast of North America. Devastations left by such floods include thousands of lives taken and billions of dollars in damages. The impact on electric power

Advisor: Dr. Jay Om jo729@nyu.edu

^{*} Corresponding Author

facilities during such flood-related natural disasters can result in devastating power outages, while flooded wastewater treatment plants can lead to the system failure and release of untreated water to the effluent channel, which contaminates the main water stream and highly elevates the risk for harmful disease and illness to spread.

The objectives of this study are to investigate the effects of riverine flooding in Gwinnett County, the assessment of vulnerable essential facilities, and to aid in preparation, such as best management practice (BMP) and a stormwater pollution prevention plan (SWPPP) to manage stormwater runoff and prevent flooding on such essential facilities to minimize potential damages.

2. Materials and Methods

HAZUS (Hazards United States) 4.2 was used for this study, and this program works with ArcGIS (Aeronautical Reconnaissance Coverage Geographic Information System) 10.5. To analyze flood hazards in Gwinnett County, the study region was defined and the watersheds were developed, also developing a stream network based on terrain's surface data. This study created three different riverine flooding scenarios by giving different values of return periods, such as 100 year floods (1% annual exceedance probability), 200 year floods (0.5% annual exceedance probability) and 500 year floods (0.2% annual exceedance probability). After creating different riverine flooding scenarios, the flood depths and floodplains for three different return periods on the selected reaches in the study area were calculated. This study also assesses the vulnerability of the nine wastewater treatment plants and three electric power facilities located in Gwinnett County, using the information of different locations.

2.1. Riverine Flooding in Gwinnett County

This study developed a stream network with segments of 1 square mile, which were used for the hydrologic analysis. Figure 1 shows the terrain surface and locations of three electric power facilities and nine wastewater treatment plants in Gwinnett County. A total of three different riverine flooding scenarios (100 years, 200 years and 500 years flood) is used for the simulation using HAZUS. Simulation using HAZUS produced flood depths and floodplains for three different return period floods on the selected reaches in Gwinnett County. This study incorporated the information of locations for wastewater treatment plants and electric power facilities which should be protected from flooding with the floodplains delineated to investigate their vulnerabilities.



Figure 1. Terrain surface, and locations of electric power facilities and wastewater treatment plants in Gwinnett County

The initial water elevation is required to run the Hurricane Model in HAZUS, and the HAZUS manual suggests to estimate the initial water level by using NOAA tide forecasts, plus a pre-storm tidal anomaly (NOAA, 2013). From the simulation, the averaged water levels were calculated using the predicted data by NOAA during the Hurricanes Irma and Michael. The difference between observed and predicted tidal elevations (two days before the hurricane landfall) was also calculated from the simulation. The predicted and observed water

elevation data at NOAA Fort Pulaski station was used because this station was available to store tidal elevation data during Hurricane Irma and Michael near the affected area. Figures 2 and 3 represent the comparison of predicted and observed tidal elevations at NOAA Fort Pulaski station during Hurricane Irma and Michael, respectively.

Tidal water surface elevations are often predicted by summing a series of cosine tidal constituents using Eq. (1).

$$\boldsymbol{\zeta} = H0 + A_1 \cos(a_1 t + E_1 - k_1) + A_2 \cos(a_2 t + E_2 - k_2) + \dots + A_n \cos(a_n t + E_n - k_n).$$
(1)

where ζ is the water surface elevation at any time, H_0 is the height of mean water level above a selected datum, Ai is the amplitude of constituent *i*, as is the frequency (speed) of constituent *i*, Ei is the equilibrium argument of constituent *i* at *t*=0, and ki is the phase lag to the maximum amplitude of constituent *i*. Speed is the rate of change in the phase of a constituent, expressed in degrees per hour; it is equal to 360 (degree) divided by the constituent period expressed in hours (Ahn and Ronan, 2019).

Figure 2 represents the comparison of predicted and observed tidal elevations during Hurricane Irma (from 9/11/2017 to 9/12/2017) at the Fort Pulaski NOAA station. The averaged predicted tidal elevation during Hurricane Irma was found to be 3.63 feet, and the difference between the forecast and the observed water level (two days before the hurricane landfall) was found to be 0.40 foot. The calculated initial water level, 1.27 feet, was used to run the hurricane model of HAZUS.



Figure 2. Comparison of predicted and observed tidal elevations during Hurricane Irma

Figure 3 shows the comparison of predicted and observed tidal elevations during Hurricane Michael (from 10/10/2018 to 10/12/2018) at the Fort Pulaski NOAA station. The averaged predicted tidal elevation during Hurricane Michael was found to be 0.190 foot, and the difference between the forecast and the observed water level (two days before the

hurricane landfall) was found to be 0.733 foot. The calculated initial water level, 0.923 foot, was used to run the hurricane model of HAZUS. From the figure 2 and figure3, calculated tidal elevation shows differences compare to observed data, since tidal height is calculated based on tidal constituents as seen on equation (1). However, water elevations are affected by wind and wave effects.



Figure 3. Comparison of predicted and observed tidal elevations during Hurricane Michael

3. Results

In the 100-year flooding in Gwinnett County, some flooding can be observed at the rivers in Gwinnett County, especially in the western region of the county. Based on the simulation, it was identified that the 200-year and 500-year floods do not have noticeably greater impacts than that of the 100-year flood. The results of the simulation are clearly shown in Figures 4, 5, and 6. To assess the vulnerable critical facility, this study did spatial data analysis which investigates the susceptible location. In this study, flood depth is the independent variables and inundated area is dependent variables. The highest flood depth due to the 100-year flood is found to be

63.852 feet, even though the 200-year flood is found to be 64.478 feet and the 500-year flood is found to be 58.852 feet. Applying finer resolution for employing Digital Elevation Model (DEM) and developing stream networks will improve the simulation.



Figure 4. 100-year Floodplain in Gwinnett County



Figure 5. 200-year Floodplain in Gwinnett County





In September 2009, NOAA reported that two wastewater pumping stations were shut down and wastewater treatment plants were also flooded during that time (NOAA, 2013). HAZUS results (Figures 4, 5, and 6) clearly show that for the Gwinnett County floods, eight of nine wastewater plants are located near a major stream or rivers and can be flooded if the water level was raised by the storm which is shown at the figures. The Gwinnett County Yellow River Water Pollution Treatment Plant is the only facility that is considered not at risk of flooding since this facility is situated at a far enough distance from the major stream. All three electric power plants are not affected by riverine flooding since those facilities are located at a far enough distance from any major rivers.

4. Conclusions

This study considered historical flooding in order to investigate the inundation area in Gwinnett County in order to aid in future preparation for flood related natural disasters (i.e., hurricanes, natural disasters). Evacuation plans can also be built by using inundation maps from these scenarios. Due to eight of nine wastewater treatment plants in Gwinnett County being flooded in each of the three scenarios; it has come to the conclusion that all the essential facilities, especially wastewater treatment plants must develop plans to protect themselves from flooding and further damages.

For future studies, fine resolution can be used to develop stream networks and employ DEM to create improved simulation results. Also, before the construction of wastewater treatment facilities and power plants, the evaluation and preparation of types of best management practice (BMP) and a stormwater pollution prevention plan (SWPPP) should be conducted to manage stormwater runoff and prevent flooding on such essential facilities to minimize potential damages.

References

Ahn, J. E., Ronan, A.D., (2019). Impact of Discrepancies between Global Ocean Tide Models on Tidal Simulations in the Shinnecock Bay Area, *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 145 (2)

NOAA (2013), Case Study, Georgia: Upper Apalachicola-Chattahoochee-Flint River Basin, Water Resources Strategies and Information Needs in Response to Extreme Weather/Climate Events (https://cpo.noaa.gov/sites/cpo/Projects/SARP/CaseSt udies/2013/ACF%20River%20Basin%20GA_Case% 20Study %20Factsheet_Extreme%20Weather%20Events_201 3-1-30v2.pdf)

FEMA (2018), HAZUS Flood Model User Guidance NOAA tidal stations.

https://pubs.usgs.gov/of/2010/1230/pdf/ofr2010-1230 .pdf.

USGS (2018) "The 100-Year Flood.", www.usgs.gov/special-topic/water-science-school/sci ence/100-year-flood?qtscience_center_objects=0#qt-science_center_objects.

"Flood Risk Management | Gwinnett County." Gwinnett, Gwinnett County Board of Commissioners, 2019, www.gwinnettcounty.com/web/gwinnett/departments/ water/whatwedo/stormwater/floodriskmanagement



Versatile PDMS Microfluidic Biosensing Platform for Rapid Detection of Viruses

Hersh Nanda^{1,2*}

¹ BASIS Chandler, Chandler, Arizona USA, ²SCENE Program, Arizona State University, Tempe, Arizona USA

Received August 14, 2021; Revised March 21, 2022; Accepted, April 1, 2022

Abstract

This project entailed the fabrication of a versatile PDMS (polydimethylsiloxane) microfluidic device to detect virus secreted glycoproteins and small molecules. Specifically, this device implements surface detection to detect virus secreted glycoproteins and small molecules using a gold biosensor which is integrated into a PDMS mold. The fabrication procedure for this device is innovative, as it significantly simplifies the traditional process of PDMS mold making, and it is versatile in that it can be adapted to fabricate different microfluidic devices that enable several other detection methods. The novel and simplified approach to PDMS mold fabrication developed as a part of this research resulted in 55% reduction in lead time and eliminated the need for a plasma cleaner, decreasing the cost of production by at least \$6000. This device can theoretically be used to detect any virus provided that the virus secretes a glycoprotein, and there is a commercially available compound that can be applied to bond that glycoprotein to the gold biosensor. This lab-on-a-chip (LOC) device offers several benefits - versatility in detection of diverse analytes, small sample size requirement, increased efficiency (less time to fabricate the device and quickly scale production to large volumes), lower production cost, device portability in the field, and ease of use for healthcare professionals.

Keywords: Varroa destructor; Apis mellifera; mites; honey bees; essential oils; mist diffusion

1. Introduction

Microfluidics involves the design and study of devices that are used to analyze the properties of fluids, using channels measuring from tens to hundreds of micrometers (Choi, et al., 2013). Microfluidics is not only revolutionizing biological approaches for DNA and enzymatic analysis, but also clinical pathology. Clinical pathology is concerned with the diagnosis of diseases, as well as the detection of toxins or pathogens in the human body.

This project has major applications in clinical

pathology, as it introduces a simpler, time-efficient, and cost-efficient procedure to fabricate microfluidic biosensing devices for the rapid detection of viruses and small molecules.

2. Research Questions

A microfluidic chip is a device that enables a tiny amount of liquid to be processed or visualized (Abdelgawad, et al., 2010). Polydimethylsiloxane (PDMS) is commonly used in these devices due to its transparency, unique rheological properties, and

Advisor: Dr. Chao Wang wangch@asu.edu

^{*} Corresponding Author hersh.nanda@gmail.com
excellent elasticity that are suitable for constructing microchannels and reaction chambers.

This project aimed to answer three questions:

- 1. How can a PDMS microfluidic device be fabricated to serve as a versatile biosensing platform to detect viruses and small molecules?
- 2. How can the PDMS to substrate bonding process be simplified?
 - a. Bonding is an essential step during microfluidic device fabrication. After a PDMS mold is made, it must be bonded to a glass slide usually via oxygen plasma bonding. Oxygen plasma bonding requires the use of a plasma cleaner, and the bonding process is error prone. One of the objectives of this project was to simplify this bonding process.
- 3. How can the process of making PDMS microwells be simplified?
 - a. Usually, microwells are made by punching holes into the PDMS mold, but this method requires additional materials and the microwells are harder to replicate.

3. Materials and Methods

This experiment involved a three-part procedure based on prior research and proprietary to Arizona State University (Chan-Park, et al., 2004; Chikkaveeraiah, et al., 2009; Friend and Yeo, 2010; Hardy, et al., 2008; Hongbin, et al., 2009; Natarajan, et al., 2008; Oh, 2008; SadAbadi, et al., 2013; San-Miguel and Lu, 2013; Subramani and Selvaganapathy, 2009; Zaytseya, et al., 2005):

- 1. Fabrication of gold biosensor device
- 2. Fabrication of PDMS device (PDMS mold making)
- 3. Device validation (sensing)

First, simulation and design optimization of the gold biosensor was done using FDTD simulation in LUMERICAL software (Figure 1). The biosensor was then fabricated and involved repetitive sample cleaning, layer depositions, EBL (electron-beam lithography for drawing microstructures), and inspection. The biosensor detects glycoproteins which are found on the surfaces of viruses. Gold (Au) nanoparticles help increase the signal strength during

the detection process due to their high surface-to-volume ratio and conductivity. Compatible substances such as streptavidin and biotin are necessary for bonding glycoproteins or small molecules to the biosensor for downstream signaling. Streptavidin is a biotin-binding protein, making streptavidin-biotin interactions useful for bonding molecules in surface detection. Other biotin-binding substances, such as avidin, could also potentially be used as an alternative.



Figure 1: Schematic of the gold NP (nanoparticle)based biosensor

During the PDMS mold making step, an empty paper bowl was first weighed and PDMS was poured into the bowl. The bowl was weighed again to determine the mass of the PDMS. Using a 1000 µL pipette, a curing agent was added to the bowl (1:10 ratio of curing agent to PDMS) and the mixture was stirred vigorously for 5 minutes. The bowl was placed into a desiccator for ~35 minutes to remove air bubbles from the mixture. A large glass slide was cleaned with acetone, isopropyl alcohol, and DI water and it was placed in an empty petri dish lid. Four 20 mL scintillation vials (lids removed) were obtained (alternatively, glass O-rings may also be used). A new glass slide (76 x 26 x 1mm) was cleaned with acetone, isopropyl alcohol, and DI water. Using a plastic dropper, 2-3 drops of the PDMS was added onto the glass slide. The rim of each vial was lubricated with PDMS and attached to the large glass slide within the petri dish. The vials were bonded to the glass slide by curing the PDMS for 30 minutes at ~90°C and the degassed PDMS was poured into the petri dish. The PDMS was cured again for 90 minutes at ~90°C. After the final curing step, the vials were carefully removed to ensure there was no PDMS leakage (Figure 2).



Figure 2: Final curing step for the device

4. Results

The final step was to validate the device capability and sensitivity through a reflection simulation using an Olympus light microscope and spectrometer to determine whether the microstructures within the biosensors were properly functioning (a picture of the final product could not be taken as the laboratory was closed immediately after completion of the experiment due to the ongoing COVID-19 pandemic). Figure 3 shows a schematic of the reflection measurement setup.

The reflection simulation provided quantitative data on each biosensor that was tested to show their respective sensitivities. Reflection measurements for each biosensor were taken in water and thiol to test their detection capability. Water was used as the control substance, while thiolated biomolecules were used as the compatible ligand for testing surface detection. A thiol is classified as any compound containing a sulfhydryl functional group, and it was used in this experiment as it is commonly found in small molecules and proteins. The reflection simulation revealed that the biosensors functioned properly and were able to detect thiolated biomolecules according to the wavelength data. Table 1 displays the quantitative test results.



Figure 3. Schematic showing the setup of the experimental spectrum measurement

_____ reflection measurement using the white light source. Measurements were taken using this technique.

alternative reflection measurement technique using laser source and 10x objective lens.

_____ alternative reflection measurement technique using laser source and parabolic mirror.

<u>data</u> from reflection simulation (for all measurement techniques) is retrieved using a spectrometer.

Table 1: Table of reflection wavelengths in nanometers for each biosensor in water (control) and thiol (experimental). These results were received from the reflection measurement analysis.

	Array 1	Array 2	Array 3	Array 4	Array 5	Array 6	Array 7	Array 8		
Water	695.524	722.058	732.123	776.308	693.369	739.514	727.093	778.426		
Thiol	696.627	735.118	737.996	783.265	697.675	742.056	732.45	783.265	Avg Peak Shift	Error
Peak Shift	1.103	13.06	5.873	6.957	4.306	2.542	5.357	4.839	5.504625	1.26337

Ideally, when measurements are taken in water and thiol, the peak reflection wavelengths for thiol should be higher when compared to water. Furthermore, the wavelength peak shift from water to thiol should be positive for each sample array, indicating a relatively higher reflection wavelength for thiol. As shown in table 1, results from the analysis correctly displayed a positive peak shift for all arrays with an average peak shift of ~5.505 nanometers, thus indicating that the biosensors functioned properly. On a graph, the curve representing the wavelengths when measurements are taken in thiol should be shifted slightly to the right of the curve representing the measurements taken in water. This shift is also indicated numerically in Table 1. Figure 4 provides a visualization of the data shown above.



Figure 4: Graphs show the comparison of average peak shifts in wavelength for each biosensor when tested in water (control) and thiol (experimental)

Figure 4 once again shows that all eight of the nanoparticle-based biosensors gold exhibited appropriate reflectance wavelengths for measurements taken in thiol and water. The experimental conditions were identical for all sample arrays, as the sole intention of the reflectance measurement analysis was to validate the biosensors' reliability in detecting compatible biomolecules. As shown, the average peak wavelength for measurements in thiol (represented by the yellow curve) is greater than the average peak wavelength for the measurements in water (represented by the blue curve). These results are consistent for all biosensors that were tested, thus validating their detection capability.

Regarding fabrication process and outcome, the device exhibited no PDMS leakages, and using the newly optimized procedure, two functional PDMS molds were produced within the span of ~5 hours by one experimenter. Traditionally, this process would have taken ~11 hours, indicating that this modified procedure resulted in a 55% increase in efficiency. Based on prior empirical data, this device can detect

the Ebola virus, as well as the small molecule cannabidiol (CBD, a natural compound in marijuana).

5. Discussion

This device falls in the category of lab-on-a-chip (LOC) devices that integrate one or several laboratory functions on a single highly miniaturized device. It offers many benefits, such as a simpler and faster fabrication process, higher efficiency in detection, lower cost for production, and small sample size requirement for detecting viruses or small molecules. This device can potentially detect any virus if the virus produces glycoproteins, and there exists a molecule that can indirectly bond the glycoprotein to the gold biosensor. Therefore, this device is theoretically capable of detecting the influenza virus, hepatitis C virus, SARS-CoV, and even the novel COVID-19.

The improvement and production of these microfluidic devices also deliver several marketing and economic benefits. For example, the wide-scale production of these devices can take place with lower cost requirements: by eliminating the use of a plasma cleaner in the PDMS mold-making process, manufacturers are able to avoid an additional expense of at least \$6000. Additionally, the increased versatility associated with this biosensing platform allows it to be easily modified, which could be a potential asset when used in response to outbreaks caused by newly discovered viruses.

Acknowledgment

This project was conducted as part of the SCENE (Science and Engineering Experience program) at the Nanoscience and Biotechnology labs at Arizona State University under the supervision of Dr. Chao Wang, Assistant professor at the School of Electrical, Computer and Energy Engineering at Arizona State University (ASU). This research was conducted under the guidance of MD Ashif Ikbal, who is a Ph.D. student at the Biodesign institute at ASU.

References

Abdelgawad, M., et al. Wu, C., Chien, W.-Y., & Sun, Y. (2010). A fast and simple method to fabricate circular microchannels in Polydimethylsiloxane (PDMS). 2010 IEEE 23rd International Conference on Micro Electro Mechanical Systems (MEMS). https://doi.org/10.1109/memsys.2010.5442395

Chan-Park, M. B., et al. (2004). Fabrication of large su-8 mold with high aspect ratio microchannels by UV exposure dose reduction. Sensors and Actuators B: Chemical, 101(1-2), 175–182. https://doi.org/10.1016/j.snb.2004.02.049

Chikkaveeraiah, B. V., et al. (2009). A microfluidic electrochemical device for high sensitivity biosensing: Detection of nanomolar hydrogen peroxide. Electrochemistry Communications, 11(4), 819–822.

https://doi.org/10.1016/j.elecom.2009.02.002

Choi, J. S., Piao, Y., & Seo, T. S. (2013). Fabrication of a circular PDMS microchannel for constructing a three-dimensional endothelial cell layer. Bioprocess and Biosystems Engineering, 36(12), 1871–1878. https://doi.org/10.1007/s00449-013-0961-z

Friend, J., & Yeo, L. (2010). Fabrication of microfluidic devices using polydimethylsiloxane.

Biomicrofluidics, 4(2), 026502. https://doi.org/10.1063/1.3259624

Hardy, B. S., Uechi, K., Zhen, J., & Pirouz Kavehpour, H. (2009). The deformation of flexible PDMS microchannels under a pressure driven flow. Lab Chip, 9(7), 935–938. https://doi.org/10.1039/b813061b

Hongbin, Y., Guangya, Z., Siong, C. F., Shouhua, W., & Feiwen, L. (2009). Novel polydimethylsiloxane (PDMS) based microchannel fabrication method for lab-on-a-chip application. Sensors and Actuators B: Chemical, 137(2), 754–761. https://doi.org/10.1016/j.snb.2008.11.035

Natarajan, S., Chang-Yen, D. A., & Gale, B. K. (2008). Large-area, high-aspect-ratio SU-8 molds for the fabrication of PDMS microfluidic devices. Journal of Micromechanics and Microengineering, 18(4), 045021.

https://doi.org/10.1088/0960-1317/18/4/045021

Oh, S. R. (2008). Thick single-layer positive photoresist mold and poly(dimethylsiloxane) (PDMS) dry etching for the fabrication of a glass–PDMS–glass microfluidic device. Journal of Micromechanics and Microengineering, 18(11), 115025.

https://doi.org/10.1088/0960-1317/18/11/115025

SadAbadi, H., Badilescu, S., Packirisamy, M., & Wüthrich, R. (2013). Integration of gold nanoparticles in PDMS microfluidics for lab-on-a-chip plasmonic biosensing of Growth Hormones. Biosensors and Bioelectronics, 44, 77–84. https://doi.org/10.1016/j.bios.2013.01.016

San-Miguel, A., & Lu, H. (2013). Microfluidics as a tool for C. Elegans Research. WormBook, 1–19. https://doi.org/10.1895/wormbook.1.162.1

Subramani, B. G., & Selvaganapathy, P. R. (2008). Surface micromachined PDMS microfluidic devices fabricated using a sacrificial photoresist. Journal of Micromechanics and Microengineering, 19(1), 015013. https://doi.org/10.1088/0960-1317/19/1/015013

Zaytseva, N. V., Goral, V. N., Montagna, R. A., & Baeumner, A. J. (2005). Development of a microfluidic biosensor module for pathogen detection. Lab on a Chip, 5(8), 805. https://doi.org/10.1039/b5038

The Meme Stock Phenomenon: Exploring the Correlations and Growth Patterns of Seven Meme Stocks in 2021

Jordan Lee^{1*}

¹ Chadwick School, Palos Verdes Peninsula, CA USA

Received December 2, 2021; Revised February 20, 2022; Accepted, March 1, 2022

Abstract

The meme stock phenomenon, which began in January of 2021, saw several different stocks of companies like GameStop, AMC, and BlackBerry surge in growth exponentially. Many of the questions that arose from this fascinating event have already been addressed, but there is one that has yet to be answered: To what extent did the growth of meme stocks correlate? Data from historical quotes of seven different meme stocks were used in order to perform a regression analysis, thereby providing the adequate resources to find the extent to which the growth of the seven meme stocks correlated. SPSS was used in order to perform the regression analysis, and significant figures that resulted from said analysis were noted. With such a new area of study, it is difficult to encapsulate all elements of the phenomenon, but using the available data, these findings show that the growth of meme stocks is indeed correlated, but there are some exceptions that are indeed notable.

Keywords: Math, Statistics, Stocks, Regression

1. Introduction

The United States stock market has existed since the late 18th-century and has witnessed significant upturns and downturns throughout its 220-year history (Terrell, n.d.). The Great Stock Market Crash of 1929, for example, saw stocks in the New York Stock Exchange plummet, causing the biggest economic recession since the market's conception. Then, the Golden Age of Capitalism, fueled greatly by the wartime economic boom during World War II, saw stock prices gradually rise, thereby signaling economic upturn. During the 2008 housing bubble crisis, stock prices fell yet again due to subprime mortgages and defaulting on loans. Most recently, the coronavirus pandemic in early 2020 caused stock prices to fall tremendously, leading to the worst economic recession since the Great Depression. However, as the economy began to slowly recover, young adults became far more interested in the stock market and cryptocurrency than ever before, thanks to new, accessible pathways like Robinhood and NFTs, or non-fungible tokens. With this drastic increase in accessibility to the stock market to the younger generation, a new kind of stock rose to prominence: meme stocks (Tepper, 2021).

Meme stocks originated from the Reddit forum known as "WallStreetBets," in which members discuss stock and option trading (Tepper, 2021). These meme stocks have high short interest and are highly volatile stocks that these WallStreetBets traders selected as potential candidates for significant

Advisor: Sophie Kim skimdow@gmail.com

^{*} Corresponding Author jordanlee2022official@gmail.com

growth (Ponczek, 2021). Examples of these meme stocks include the likes of GameStop, BlackBerry, Bed Bath and Beyond, AMC Entertainment, Nokia, Koss, and Express (Tepper 2021). These seven companies are the meme stocks that will be examined in this research project. Now, a question may be: Why are these companies considered to be "meme stocks?" There are various reasons, but the most salient reason is that these companies are popular in something commonly known as "meme culture" and amongst younger generations. For example, GameStop, AMC, BlackBerry, and Nokia used to be extremely popular during the 2000s and early 2010s. However, GameStop now faces heavy competition from online video game retailers, streaming services like Netflix and Hulu caused AMC's revenue to gradually decline. In addition, lockdown orders caused by the COVID-19 pandemic also exacerbated AMC's decline whilst accelerating the growth of the aforementioned streaming services. Tech giants like Apple and Samsung nearly drove BlackBerry and Nokia out of the cell phone industry. In fact, according to S&P Global, Netflix experienced its first ever cash-flow-positive year, raking in approximately \$1.9 billion in 2020 (Akins, 2021). Meanwhile, in 2020, AMC warned its investors that it could exhaust its cash holdings as a result of the coronavirus pandemic and fierce competition from streaming services like Netflix, clearly showing how much of an impact new competition has had on these meme stocks (Akins, 2021). Therefore, participants in the aforementioned forum "r/WallStreetBets" chose these companies because of the fact that they used to be so popular amongst their generation and are now facing bankruptcy (Backman, 2021).

Ultimately, the driving question of this research project is: Do meme stocks move together? More specifically, are AMC and GME's stock growth correlated? To uncover a clear answer to this query, regression analysis is used to find the extent to which the meme stocks are correlated from historical quotes from each of the seven meme stocks: GameStop, BlackBerry, AMC, Bed Bath and Beyond, Nokia, Express, and Koss.

2. Materials and Methods

Regression analysis is one of the fundamental and most important concepts within the realm of statistics. It can be utilized in many ways, ranging from an analysis of a high-school science experiment to asset pricing in financial economics. It allows researchers and statisticians to understand which certain independent variables, or inputs, have an impact on the dependent variable, or output. Using regression analysis, they can know which independent variables should be noted, which can be ignored, how these independent variables interact with one another, and the degree of certainty of each independent variable's impact on the dependent variable.

In this specific research project, utilize historical quotes from the aforementioned meme stocks dating back from January 2021 to July 2021, thereby encompassing the time period in which these meme stocks saw a surge in growth. These historical quotes were downloaded from MarketWatch and Nasdaq in the form of csv files, which were then imported into Google Sheets for easier usability (MarketWatch and Nasdaq, 2021). Use the open, close, high, and low prices as well as the total trading volume for every day between said time period, and this data will be averaged to create specific independent variables for our research project. Therefore, the independent variables are quantitative. The dependent variable is also linear because it essentially provides the extent to which the meme stocks are correlated by producing numerical evidence. This kind of regression is known as a linear regression, and because we will be utilizing multiple independent variables and only one dependent variable, it is also a multiple regression. Therefore, the type of regression we will be using is the multiple linear regression. In our experiment, the dependent variable is the close price of GameStop whilst the independent variables are the open, close, high, and low prices as well as the volume of each meme stock, which amounts to one dependent variable and 34 independent variables.

In order to perform said multiple linear regression, SPSS will be used, also known as Statistical Package for the Social Sciences. SPSS, a statistical software platform created by IBM, allows users to better understand and find specific patterns in complex datasets, and with historical quotes datasets with thousands of individual data points, SPSS is extremely useful (Laerd, 2018). After transferring the data collected from MarketWatch into SPSS, check specifically for collinearity by conducting collinearity diagnostics. Essentially, run a test to observe how closely related the independent variables within our experiment are. Ultimately, there should be as little collinearity as possible to ensure accurate results.

Then, assumptions testing needs to be performed (Laerd, 2018). Firstly, in order to have independence of observations, check the Durbin-Watson statistic, which should be included in the model summary of the regression. If the Durbin-Watson statistic is greater than 1.5 and less than 2.5, which means that there is little to no autocorrelation in the sample (Laerd, 2018) (Hayes, 2019).

Another aspect of the regression to check is whether or not there is a linear relationship between the dependent variable and each independent variable (Laerd, 2018). In order to find this out, use the P-P plot created with the regression in which the data points should follow the line of best fit closely. The P-P plot essentially measures the closeness of two data sets, and it reveals information about the skewness of the data set (ScienceDirect) (NIST, 2019).

This leads conveniently into the next assumption test, which regards the presence of significant outliers in the data (Laerd, 2018). There should be no significant outliers in the data, and skewness helps determine whether or not they exist (Laerd, 2018).

In addition. the data must not show heteroscedasticity and, in turn, should show homoscedasticity (Laerd, 2018) (LibreTexts, 2017). In order to determine this, use the scatterplot generated by SPSS Statistics. A scatterplot that shows little to no heteroscedasticity is one in which the data points are distributed in some sort of pattern or shape, most likely along a line, whilst a scatterplot that shows heteroscedasticity is one in which the data points do not have a discernible pattern or shape (LibreTexts, 2017).

Another assumption test regards the presence of multicollinearity, which was discussed earlier.⁶ Lastly, check if the residuals are distributed normally (Laerd, 2018). One way to determine this is by analyzing the histogram generated by SPSS Statistics.

The histogram should follow normal distribution, which is also known as the bell curve. Kurtosis can measure to find out the extent to which the data follows normal distribution (NIST, 2019).

In addition to this assumption testing, use the f-statistic, which helps determine if the means of two datasets are significantly different, which can also determine the p-value. (ScienceDirect). The p-value is useful because it measures the probability of getting a result at least as extreme as the actual result (Tanha and Janani, 2017).

Table 1: List of Variables Involved

Table 1. Lis	
Variable	Description
GME_	The dependent variable and the close
Close/Last	price of GameStop.
GME_	An independent variable and the open
Open	price of GameStop.
GME_	An independent variable and the high
High	price of GameStop.
GME_	An independent variable and the low
Low	price of GameStop.
GME_	An independent variable and the number
Volume	of shares traded of GameStop.
AMC_	An independent variable and the close
Close/Last	price of AMC.
AMC_	An independent variable and the open
Open	price of AMC.
AMC_	An independent variable and the high
High	price of AMC.
AMC_	An independent variable and the low
Low	price of AMC.
AMC_	An independent variable and the number
Volume	of shares traded of AMC.
BB_	An independent variable and the close
Close/Last	price of BlackBerry.
BB_	An independent variable and the open
Open	price of BlackBerry.
BB_	An independent variable and the high
High	price of BlackBerry.
BB	An independent variable and the low
Low	price of BlackBerry.
BB	An independent variable and the number
Volume	of shares traded of BlackBerry.
BBBY	An independent variable and the close
Close/Last	price of Bed Bath and Beyond.
BBBY	An independent variable and the open
Open [–]	price of Bed Bath and Beyond.
BBBY_	An independent variable and the high
— High	price of Bed Bath and Beyond.

BBBY_	An independent variable and the low
Low	price of Bed Bath and Beyond.
BBBY_	An independent variable and the number
Volume	of shares traded of Bed Bath and Beyond.
KOSS_	An independent variable and the close
Close/Last	price of Koss.
KOSS_	An independent variable and the open
Open	price of Koss.
KOSS_	An independent variable and the high
High	price of Koss.
KOSS_	An independent variable and the low
Low	price of Koss.
KOSS_	An independent variable and the number
Volume	of shares traded of Koss.
NOK_	An independent variable and the close
Close/Last	price of Nokia.
NOK_	An independent variable and the open
Open	price of Nokia.
NOK_	An independent variable and the high
High	price of Nokia.
NOK_	An independent variable and the low
Low	price of Nokia.
NOK_	An independent variable and the number
Volume	of shares traded of Nokia.
EXPR_	An independent variable and the close
Close/Last	price of Express.
EXPR_	An independent variable and the open
Open	price of Express.
EXPR_	An independent variable and the high
High	price of Express.
EXPR_	An independent variable and the low
Low	price of Express.
EXPR_	An independent variable and the number
Volume	of shares traded of Express.

3. Results

Before discussing the results, one of the most important tests is the collinearity diagnostics test, which essentially runs a test across the independent variables that finds the extent to which variables have a relationship. In the end, the collinearity should be low, fulfilling the meaning of "independent variable". Remember, the dependent variable is the close price of GameStop, and the independent variables are the close, open, high, and low prices as well as the volume of the other seven meme stocks.



Figure 1. Histogram of multiple linear regression for AllClose, a compilation of all "close" prices of the meme stocks

Histograms represent the distortion or asymmetry of normal distribution, also known as kurtosis (NIST, 2019). Although kurtosis and skewness may be similar, the main distinction is that kurtosis measures variation within the same tail, whilst skewness measures variation between opposite tails (NIST, 2019). The kurtosis value is determined by the extent to which the chart is peaked or flattened in comparison to the normal distribution, which is the typical bell curve outlined in the chart (NIST, 2019). The data above (Figure 1) follows the bell curve relatively well, but there is a significant outlier just before the peak of the bell curve. The standard deviation, 0.980, which is relatively low, shows that the above data is not very spread out.



Figure 2. Normal P-P plot of multiple linear regression for AllClose, a compilation of all "close" prices of the meme stocks

P-P plots tell us about the skewness of data (ScienceDirect) Here, in the above graph (Figure 2), the P-P plot is not very varied, and there are no significant outliers. This means that the skewness is near 0 (ScienceDirect).



Figure 3. Scatterplot of GME close variable in multiple linear regression predicted vs. residual values for AllClose, a compilation of all "close" prices of the meme stocks.

Scatterplots tell us about the heteroscedasticity of data, which is when the standard deviation of a specific variable differs overtime, which could be over different periods of time or different values of independent variables (LibreTexts, 2017). Essentially, heteroscedasticity is in regards to data that has unequal variability, or scatter. The plot typically follows a cone shape, with the base facing in the positive x-direction. In the above graph (Figure 3), the data is rather varied, showing little correlation or patter



Figure 4. Histogram of multiple linear regression for GMEElse, a compilation of data specifically comparing GameStop and AMC

In this histogram (Figure 4), the kurtosis is greater than 0 due to the fact that the chart is peaked far above the peak of the bell curve, or the normal distribution. This means that the data is leptokurtic. The standard deviation is relatively low, 0.986, meaning that the data is not very spread out.



Figure 5. Normal P-P plot of multiple linear regression for GMEElse, a compilation of data specifically comparing GameStop and AMC.

This P-P plot (Figure 5) is more varied than the AllClose P-P plot, which means that the skewness is higher. The distribution between the two samples differs, as the points deviate from the 45 degree line (x=y).



Figure 6. Scatterplot of GME close variable in multiple linear regression predicted vs. residual values for GMEElse, a compilation of data specifically comparing GameStop and AMC.

This scatterplot (Figure 6) shows heteroscedasticity due to the fact that the data is relatively scattered, and the plot follows a cone shape.

Within this research project, there are multiple independent variables and one dependent variable that we will be observing during the regression analysis. The quantitative independent variables are as follows: the open, close, high, and low prices of GameStop and AMC averaged over a specific time period.

Before discussing the details about each regression, it is imperative that the variables involved in each regression are known. The first and largest

regression, the AllElse regression, includes the dependent variable, GME Close/Last, and all 34 independent variables mentioned in Table 1. The second regression, GMEElse, is more of an internal comparison between the variables that relate to GameStop, which includes the dependent variable, GME Close/Last, and four independent variables: GME High, GME Open, GME Low, and GME Volume. Lastly, the third regression, AllClose, includes all variables related to the close price of the meme stocks: GME Close/Last as the dependent variable and six other close price variables for each meme stock as the independent variables.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change
1	.998	.996	.994	\$5.34795	.996
Model	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	763 029	34	115	< 001	2.066

Table 2. Model Summary for AllElse

Table 3	. Model	Summary	for	GMEElse
---------	---------	---------	-----	---------

		-			
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change
1	.990	.980	.979	\$10.21659	.980
Model	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	1748.772	4	145	<.001	2.305

Table 4. Model Summary for AllClose

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change
1	.890	.791	.783	\$32.96648	.791
Model	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	90 460	6	143	< 001	500

Above are the model summaries for each regression. They list all the important values, including the likes of the R-squared value, which measures the extent to which the dependent variable varies based on the independent variable(s); p-value, which is the probability of having test results at least as extreme as the actual results, assuming that the null hypothesis is correct; and the degrees of freedom, which are the numbers of independent

entities of information that went into calculating the estimate (Laerd, 2018).

The r-squared value, which is also the square of the Pearson correlation coefficient, represents the closeness of the data plotted against the line of best fit (Laerd, 2018). A statistically significant r-squared value is usually 0.8 or above, which shows significant correlation between the data and the line of best fit (Laerd, 2018). However, any r-squared value greater than 0.7 shows that the data is strongly correlated, whilst an r-squared value less than 0.3 shows that the data is not strongly correlated (Laerd, 2018). In both the AllElse and Else regressions (Table 2, Table 3), the r-squared values are statistically significant, as they are 0.996 and 0.980, respectively. However, the r-squared value for the AllClose regression (Table 4) is not statistically significant due to the fact that the r-squared value is less than 0.8.The statistical implication of a statistically significant r-squared value is that the dependent variable is closely related to the independent variables. So, why is the r-squared value so high for these two regressions?

4. Discussion

The first regression, also known as the AllElse regression, exhibits a very high r-squared value, 0.996. This is significant because it shows that the correlation between the dependent variable, the closing price of the GameStop stock, is very closely correlated to the 36 other independent variables, which include the high, low, open, and close prices as well as the volume of the other meme stocks, as an r-squared value of 1 means that the dependent and independent variables are perfectly correlated (Laerd, 2018). From the correlations table, there are some independent variables that are more closely correlated than others. Examples of this phenomenon include very close correlations between independent variables of the same meme stock, like the correlation between the high and low prices of BlackBerry. In addition to these correlations, however, there are others that are not of the same meme stock: an example of this is how the GameStop open price is closely correlated to the Koss low price and the Express close, open, high, and low prices, with Pearson correlation values (r values) of 0.851, 0.813, 0.831, 0.905, and 0.793, respectively. There are also statistically significant variables that can be determined from the correlations table. As a result, these findings show that our assumptions were true: the seven meme stocks do move together and behave similarly. However, the real question regards what actually causes these seven meme stocks to move together. This question will be answered in the second and third regressions.

The second regression, also known as the GME Else regression, includes several important values: R-Squared = 0.980, Adjusted R-Squared = 0.979, df1 = 4, df2 = 145, Sig. F-Change < 0.001, and the GME High Unstandardized b = 0.406. The adjusted R-Squared adjusts for the bias of R-Squared, which is usually an overestimate due to the large number of predictor, or independent, variables, and the small sample size. The df values, also known as the degrees of freedom, represent the number of pieces of information that contributed to the calculation of the estimate. The F-Change value is the R-Squared change tested by an F-Test, which helps determine the model of best fit for a certain population from the original dataset (ScienceDirect). A significant F-Change, which is less than 0.005, means that the prediction was greatly improved by the predictor variables added. Finally, the unstandardized B is the slope of the line of regression (UCLA). Ultimately, the data is going down because the low price, which is the lowest price of a certain trading day, is closely correlated with the closing price, so it is possible to identify the low price to the close price, which makes sense because after the January-February meme stock bubble, stock prices for these meme stocks have decreased overtime.

Finally, the third regression, also known as the CloseElse regression, compares all the closing values across the seven meme stocks. There are a few important values to highlight, especially the unstandardized B values of several of the meme stocks: R-Squared = 0.791, Adjusted R-Squared = 0.783, df1 = 6, df2 = 143, Sig. F-Change < 0.001, AMC Close/Last Unstandardized B = 0.977, KOSS Close/Last Unstandardized B = 4.462, BB Close/Last Unstandardized B = -8.598, EXPR Close/Last Unstandardized B = 0.046, and BBBY Close/Last Unstandardized B = 3.882. The R-squared value is lower in comparison to the other two regressions, but it is still considered to be strongly correlated, as the value is greater than 0.7. as seen by the mostly positive Clearly. unstandardized B values, there are unknown forces that are driving certain stock prices, but there is most definitely a common driving force. This common driving force can be the continued activity of investors and members of the Reddit forum r/wallstreetbets, which is the catalyst of the meme stock phenomenon (Bhansali). In addition, Robinhood, an investment platform catered towards young adults in the digital age, allowed this demographic to engage in the stock market to an unprecedented extent, thereby having much influence on the meme stock phenomenon. This "herd mentality" allows for such a high correlation of these meme stocks and their open, close, high, and low prices. These external factors are what mostly likely caused these meme stocks to be so closely correlated. An interesting aspect of this regression, however, is the uniqueness of the BlackBerry unstandardized B value, which is actually negative. Why this is negative may have to do with the fact that BlackBerry has characteristics that distinguish itself from the other six meme stocks. BlackBerry, though outperformed in the smartphone industry, has a strong grasp in the software industry, as their connections to various governments, especially the US government, are extremely robust. Therefore, instead of being depicted as a meme stock, BlackBerry could very much be a legitimate investment that could produce profits in the long-run.

Though multicollinearity may have posed an issue in this experiment, it is important to note that with stock analysis, sometimes this phenomenon occurs. With the largest variable set, which is included in the AllElse regression, multicollinearity is to be expected. The reason why we ran a stats test on the largest variable set was to provide a general picture of what can be studied and what the dependent variable looks like in relation to the independent variables. The second and third variable sets, which are included in GMEElse and AllClose, respectively, are much smaller, especially GMEElse's variable set, which only includes five variables. These regressions were performed in order to provide more specific results, which were found and mentioned above.

5. Conclusion

We applied regression analysis to a recent event like the meme stock phenomenon by identifying significant values in the model summaries as well as the charts generated by SPSS Statistics. Through extensive background research, data collection, and analysis of significant values, we discovered that yes, meme stocks do indeed move together. The fact that they do is not necessarily surprising, but it is fascinating nonetheless. r/wallstreetbets, Robinhood, and other platforms were possible causes for the correlated behavior between meme stocks. Through the use of regression analysis, we discovered an extraordinary aspect of the meme stock phenomenon, but there is most definitely ample room for further research that taps into meme stocks' relationships with the rise of cryptocurrency, what factors caused meme stocks to grow so quickly, and many others.

Acknowledgements

I would like to acknowledge and thank Miss Sophie Kim, who was my research mentor for this independent research project. Her valuable insights, unwavering support, and vast knowledge made it possible for me to pursue my passions to such an extent.

References

1.3.5.11. Measures of Skewness and Kurtosis. (2019). Nist.gov.

https://www.itl.nist.gov/div898/handbook/eda/section 3/eda35b.htm

4.5: Homoscedasticity and Heteroscedasticity. (2017, June 27). Statistics LibreTexts.

https://stats.libretexts.org/Bookshelves/Applied_Stati stics/Book%3A_Biological_Statistics_(McDonald)/0 4%3A_Tests_for_One_Measurement_Variable/4.05% 3A Homoscedasticity and Heteroscedasticity

Backman, M. (2021, February 22). *Poll: How Well Did Meme Stock Investors Understand What Was Happening*? The Motley Fool. https://www.fool.com/research/poll-how-well-did-me me-stock-investors-understand/

Bhansali, V. (n.d.). *Why The GameStop Phenomenon Is Not Surprising*. Forbes. Retrieved December 12, 2021, from

https://www.forbes.com/sites/vineerbhansali/2021/01/ 27/why-the-gamestop-phenomenon-is-not-surprising/ ?sh=439254772635

F Test - an overview | *ScienceDirect Topics*. (n.d.). Www.sciencedirect.com.

https://www.sciencedirect.com/topics/mathematics/f-t est

Hayes, A. (2019). *Understanding the Durbin Watson Statistic*. Investopedia. https://www.investopedia.com/terms/d/durbin-watson

-statistic.asp

Laerd Statistics. (2018). *How to perform a Multiple Regression Analysis in SPSS Statistics* | *Laerd Statistics*. Laerd.com. https://statistics.laerd.com/spss-tutorials/multiple-regr

ession-using-spss-statistics.php

Laerd Statistics. (2018). Pearson Product-Moment Correlation - When you should run this test, the range of values the coefficient can take and how to measure strength of association. Laerd.com. https://statistics.laerd.com/statistical-guides/pearsoncorrelation-coefficient-statistical-guide.php

MarketWatch. (2019, March 15). *MarketWatch: Stock market news - financial news*. MarketWatch. https://www.marketwatch.com/

NASDAQ. (2019). NASDAQ's Homepage for Retail Investors. NASDAQ.com. https://www.nasdaq.com/ Netflix surges on earnings; Investors cheer AMC's debt deal. (n.d.). www.spglobal.com. Retrieved December 12, 2021, from

https://www.spglobal.com/marketintelligence/en/new s-insights/latest-news-headlines/netflix-surges-on-ear nings-investors-cheer-amc-s-debt-deal-62225648 *P-P Plot - an overview* | *ScienceDirect Topics*. (n.d.). Www.sciencedirect.com.

https://www.sciencedirect.com/topics/mathematics/p-p-plot

Ponczek, S. (2021, February 2). *Meme Stocks Lose* \$167 Billion as Reddit Crowd Preaches Defiance. Bloomberg.com.

https://www.bloomberg.com/news/articles/2021-02-0 2/moonshot-stocks-lose-167-billion-as-crowd-preach es-defiance

Regression with SPSS Chapter 1 – Simple and Multiple Regression. (n.d.). Stats.idre.ucla.edu. https://stats.idre.ucla.edu/spss/webbooks/reg/chapter1 /regressionwith-spsschapter-1-simple-and-multiple-re gression/

Tanha, K., Mohammadi, N., & Janani, L. (2017). P-value: What is and what is not. *Medical Journal of the Islamic Republic of Iran*, *31*(1), 377–378. https://doi.org/10.14196/mjiri.31.65

Tepper, T. (2021, February 4). *The Rise And Fall Of The GameStop Meme Stocks*. Forbes Advisor. https://www.forbes.com/advisor/investing/gamestop-meme-stocks-bb-amc-nok/

Terrell, E. (n.d.). *Research Guides: Wall Street and the Stock Exchanges: Historical Resources: Stock Exchanges*. Guides.loc.gov. https://guides.loc.gov/wall-street-history/exchanges

The Effect of Retinal Scan Image Resolution on the Performance and Accuracy of Deep Learning Models

Saaketh Madabhushi^{1*} and Anish R. Roy²

¹Mission San Jose High School, Fremont, CA USA ²Stanford University, Stanford, CA, USA

Received November 18, 2021; Revised April 4, 2022; Accepted, April 13, 2022

Abstract

Deep learning approaches have increasingly been used in the diagnosis of disease and other image recognition problems. Training a neural network requires a high amount of computational power. An assessment of the performance of neural networks on less than optimal image sets and hardware is desirable for these less developed nations. This research seeks to test the performance of neural networks under non-ideal conditions. Examples include fewer resources and training data for the neural network to utilize or outdated hardware. A dataset of images of diabetic retinopathy, with five different levels of severity, was procured. These images were split into four different test resolutions, none of which were the highest possible resolution images of the retinal scans available in the dataset, before being trained across three different neural network architectures: LeNet, AlexNet, and Encoder-Decoder. Images of diabetic retinopathy were utilized specifically as these are more complex images and would allow for the discovery of the limitations of deep learning across different resolutions. Furthermore, after conducting research, it was observed that AlexNet performed the best overall and achieved the highest accuracy of 57.2%. Encoder-Decoder was able to achieve 50.2% and LeNet was able to achieve 51.3%. In addition, the time until convergence for all three neural networks varied depending on the amount of layers and depth of each neural network, as well as the image resolution inputted. In general, neural networks trained with higher resolution images had higher accuracy, but still did not reach optimal performance.

Keywords: Deep Learning, Convolutional Neural Networks, Diabetic Retinopathy

1. Introduction

With the turn of the twenty-first century, the emergence of neural networks (NNs) are increasingly important in day-to-day life. Put simply, a neural network is a computer algorithm that uses multiple data points to try and find a specific pattern amongst the data in order to complete a certain task. The main goal of NNs is to develop algorithms so that machines can complete tasks normally attributed to humans For example, NNs can identify and classify

* Corresponding Author saakethraj@gmail.com

different objects in images. As machines are given access to more data they are able to make more accurate predictions using complex algorithms. The main benefit that an NN provides is increased efficiency and ability to solve complex problems. Examples include autonomous driving and medical image analysis. NNs can be used in driving for geographic mapping and vehicle detection, as well as any other landmarks, potential hazards, or road signals (Huang & Chen, 2020). In terms of medical image analysis, most medical diagnostic and image

Advisor: Anish R. Roy arroy@stanford.edu interpretation have been performed by human experts. With the rate of progress of computational medical image analysis, NNs have the potential to be extremely effective in this field, especially in terms of time saved for medical professionals (Shen, et al., 2017).

An array of real world problems can be solved using one specific type of NN: a Convolutional Neural Network (CNN), a subclass of Artificial Neural Networks (ANN). To better understand the layers of a CNN, we must first understand how ANNs are structured. ANNs, which are many neurons (basic units of NNs) connected to each other. A neuron takes the weighted sum of the input and adds a bias to produce one output (Zhou, 2019). All artificial neural networks have an activation function, which turns a bounded input into a predictably formed output. In addition, activation functions modify the input allowing for non-linear problems to be solved. Inputs from neurons are passed down to other neutrons, and this process continues until the desirable output is obtained. Many artificial neural networks have hidden layers, which are any layer between the input and the output (LeCun, et al., 2015). In order to train a neural network, a dataset must be provided for the neural network to "practice" on so that it can fine tune the appropriate weights of the kernels and nodes in the neural network. In order to conduct this training, an adequate hardware setup must be available. However, not everyone has access to such hardware.

When large, detailed images are inputted into the neural network, it is very resource intensive. The main advantage of CNNs is their ability to reduce the complexity of a given model, which often leads to faster training and is generally more efficient, specifically with regards to images. CNNs can be trained to better understand complex images as more images are fed through the system. By reducing any given image into a form which is easier to process without losing any of the main features of that image in order to get an accurate prediction, a CNN can maintain its efficiency and apply what it learns to a wider range of images. Computer vision is vital when tackling complex issues such as image classification. Typically, convolutional neural networks are used to solve these problems efficiently. However, CNNs

involve large and complex data sets for training. In this paper, we wanted to determine the effect of image resolution of the training data set on the performance of CNNs for a specific problem: classifying the disease state of retinopathy.

CNNs generally consist of three main types of layers: convolutional layers, pooling layers and fully-connected (dense) layers (O'Shea & Nash). The first structure in a CNN is the convolutional layer. The first part of this layer is the kernel/filter, which helps to truncate a given input into a form which makes it easier to process a large input of info and to output a more compact, easier-to-understand image/result. The CNN takes the input and convoles it with the kernel and may add a bias. The elements in the kernel are the trainable weights (O'Shea & Nash). The pooling layer is responsible for reducing the size of any given input image. This allows the program to process the image using less computer resources, as it has been scaled down. The pooling layer also makes sure to extract all useful and dominant information from any input, while still allowing the neural network to learn the overall value of any given input (O'Shea & Nash). The most common type of pooling is max pooling. Max pooling takes a given output from the convolutional layer, which includes the key features of a given sector of an image, and records the most dominant part of that sector. Therefore, it cuts out any unnecessary information or "noise" contained in that sector, only returning the most dominant and important feature.

Depending on how complex a CNN is or the inputs are, the number of convolutional and pooling layers' can vary in any given CNN. The more detail that is needed to be captured from an input, the more convolutional and pooling layers may be added - but at the expense of more computational power being used. A dense layer, or fully-connected layer, is an efficient way to obtain non-image outputs from image datasets. However, fully-connected layers are also very computationally expensive, and therefore, cannot be overused in a CNN model.

Some examples of basic neural network architectures include: LeNet-5 (Figure 1A), AlexNet (also Figure 1A), and Encoder-Decoder (Figure 1B). LeNet-5 is the first major CNN architecture that was developed in 1998. LeNet-5 was a simple architecture that was compact and efficient and was mainly used for letter and number recognition. As it did not have many layers, it was considered a fairly shallow neural network compared to more modern architectures. AlexNet is a more advanced version of LeNet-5. AlexNet consists of eight layers, five of which are convolutional layers and three of which are fully connected (dense) layers. The model won the 2012 ImageNet competition by more than an eleven percent error difference (Gupta, 2020). This was a significant improvement and milestone in the advancement of neural networks. An Encoder-Decoder structure first down-samples images in a dataset to a low resolution, all while using convolutional layers to create a feature map. Then, it up-samples the data back to its original size while running more convolutional layers to add to the creative feature map.



Figure 1. (A) This figure depicts the general schematic of the LeNet and AlexNet CNN architectures. The image is processed through convolutional layers, max pooling layers, and sent to the dense layers before obtaining an output. (B) This figure depicts the general schematic of the Encoder-Decoder CNN architecture. This architecture down-samples and up-samples the image before its information is sent to the dense layer(s) to give an output. The lighter shade of blue highlights the downsampling of the image, while the darker shade of blue highlights the upsampling of the image.

Occasionally, the procurement of good datasets is not feasible. Internationally, datasets are not always easily available, and many datasets have images which are not the best resolution or optimized for machine learning. In order to avoid bias, a dataset should represent the classes that are meant to be predicted (Mestre, 2018). When collecting data of different categories, such as race, gender, or color, it is often difficult to achieve consistency between images in each category, as well as in the dataset overall. Usually, datasets must be stitched together using various online sources and sets (Hooker, 2021). Moreover, adding/finding labels to these images often takes a tremendous amount of time and resources and leads to inaccurate and inefficient datasets. This problem can occur in a dataset of any type when it comes to image classification.Furthermore, not all individuals have access to high performance graphical processing units (GPUs), hardware that drastically decreases time of training, as well as dedicated computers from which to run the neural network. This project simulates the effectiveness of having a less-than-optimal setup for the performance of CNNs.

As the resolution of an image set increases, the computational power required to process the image set also increases. The resolution of the image set is critical as the higher the resolution, the more detail and information the neural network can extract from each given image. To test the effect of non-ideal resolution and hardware on neural networks, we assessed the ability CNNs to predict the severity of diabetic retinopathy when given a retinal scan. Diabetic retinopathy is a diabetes complication that affects the eyes. It is mainly caused by damage to certain blood vessels in the retina. Usually, diabetic retinopathy starts with little to no symptoms which increasingly worsen over time. It is a chronic condition and can last for years or be lifelong. This disease typically affects around 200,000 people in the U.S. per year. Some problems related to diabetic retinopathy include difficulty with vision and perception of different colors (Mayo Clinic, 2021). As stated above, CNNs can be very useful in diagnosis of diseases and other image recognition type situations that may arise. Since an image set of varying eye disease cases is more complex when compared to simple problems such as classifying numbers, it will expose the limitations of neural networks when the resolutions of the datasets are reduced for medically relevant data.

А

в

2. Materials and Methods

The data inputted to each neural network was normalized and shuffled. Tensorflow was used in the creation of the three neural networks. The parameters among the three CNNs are similar. All of the nets use an increasing number of convolutions starting with 16 up to 128 in powers of 2. All max pooling layers are of size (2,2) and a stride of 2. Each network had 1024 dense layers with a dropout layer parameterized to 0.2. All CNNs used the Relu activation function and the SoftMax classification function was used in the last layer. We used the Adam optimizer to train. The batch size used was 32. A callback function was used to reduce the learning rate if the validation loss plateaued, with a patience of 2 epochs. Each reduction reduced the learning rate by a factor of 0.2. The initial learning rate used was 0.01. LeNet consists of three convolutional layers and two dense layers. AlexNet consists of five convolutional layers, with two dense layers. Encoder-Decoder consists of three downsampling layers and three upsampling layers, with two dense layers.

To achieve the four different resolutions of our imageset, we used the *resize* function from the CV2 library of Python. This function changes the dimensions of an image and preserves the aspect ratio. As we decrease the resolutions of the images, the pixel size increases. This will help to keep the image set relatively uniform in terms of size while allowing simulation of different levels in dataset image quality. We made sure to label each category of severity in increasing order, with zero being no eye disease and four being severe eye disease.

Our computer is homebuilt and only equipped with 16 gigabytes of random access memory (RAM), as well as 6 gigabytes of GPU memory. The computer setup overall is not specialized for computationally demanding applications. So, our computer was not able to handle image resolution sizes up to 164, as it was not able to allocate enough memory during training for larger image sizes.

3. Results

We chose these three convolutional neural network architectures (LeNet-5, AlexNet, and

Encoder-Decoder). The performance of each architecture can be tested when given the same dataset with varying resolutions. (See materials and methods for further details about the nets).



Figure 2. (A) This figure depicts some sample images from the dataset that was used. Each of the images have a 512x512 resolution and depict the different categories of eye disease present in the dataset. (B) This figure depicts some sample images from the dataset that was used. This figure depicts the different resolutions tested in this project using the same sample eye image which is from the same scan of an eye that does not have disease.

We procured our dataset of various eye scans with variations in severity of diabetic retinopathy to train the nets from Kaggle, a popular website where many datasets from a range of genres can be found (Tandon, 2021). This dataset has roughly 35,000 images, as well as individual training sets, testing sets, and validation sets within the superset. All images are 512x512 base resolution (Figure 2A). All images are formatted in gray-scale and are organized into five different categories based on the severity of diabetic retinopathy present in the given eye. These categories are no diabetic retinopathy, mild, moderate, proliferate, and severe. Each of the categories in the training set has 7,000 images. Each of the categories in the validation set has 2,000 images that were not used in training. We worked with four different resolutions: 32x32, 64x64, 128x128, and 164x164 (Figure 2B).

In terms of LeNet, as the image resolution increased, the validation accuracy tended to increase. However, the validation accuracy plateaued around roughly fifty-one percent. It should be noted that LeNet architecture is the oldest of the three architectures on which we chose to base our models. Given the shallow nature of LeNet and the fact that it was created for simpler datasets with less complexity in each image, it was expected that LeNet would not perform up to the same standards as the other two models. Even if the image size was increased up to 512x512, it is not clear whether this model would perform with a similar accuracy to AlexNet and Encoder-Decoder. As shown in Figure 3A and Table 1, the accuracy only gets marginally better (0.5%)when given image inputs 164x164 resolution. It is also noted in the table that as the resolution of the image increases, the time until convergence of each test run of the neural network also increases.

In terms of AlexNet, the validation accuracy increased as the image resolution was increased. As AlexNet is more complex than LeNet in terms of its layers, it is possible that the accuracy could have kept on increasing at even higher resolutions tested in this paper. AlexNet gave the best results for this experiment. As AlexNet is a deeper network with more convolutional and pooling layers compared to LeNet, the higher accuracy is to be expected with training with the base image resolution of 512x512. As mentioned above and shown in Figure 3A and Table 2, the time until convergence for AlexNet increases as the image resolution increases. Since AlexNet is a more complex neural network compared to LeNet, we can see that there is a much longer time until convergence for the two upper resolutions of the images.



Figure 3A. This figure depicts a line graph of Accuracy v.s. Epochs for each CNN at 128x128 image resolution.



Figure 3B. This figure depicts three segmented bar graphs that each pertain to one CNN architecture. The graphs show how each network performed for each given image resolution. Orange represents the percentage of correctly classified images, and blue represents the percentage of incorrectly classified images. For each network, accuracy tended to increase as the image resolution increased. Accuracy is defined as (number of images classified correctly) / (total number of images in the dataset).

In terms of Encoder-Decoder, the validation accuracy increased up to the 128×128 image and then strikingly slightly regressed when given an input image of 164×164 . Likely, this was caused by such a

drastic reduction in the image size, as well as only inputting images that are less than half of the original image size. Therefore, the amount of detail that is captured by the network is lowered due to the extremely low resolution of the image. Since the 164x164 image resolution is not a significant difference in image resolution compared to 128x128, the amount of information that the network was able to extract from 164x164 was likely very similar to the 128x128 resolution image. Accordingly, as both image resolutions were fairly similar as inputs for the neural network, the fact that 128x128 performed better is likely due to chance. This likely explains the cause of some of the inefficiencies that occured in this network and why it did not improve as much as AlexNet when provided with a marginally higher resolution image. As mentioned above and shown in Figure 3A and Table 3, the time until convergence for Encoder-Decoder increases as the image resolution increases. We can see that Encoder-Decoder takes the longest time until convergence at all given image resolutions, as it is the most complex of the three neural networks.

Table 1: This datatable gives a context for the time taken by the LeNet CNN architecture to converge with each given image resolution. This table helps to contextualize the line graph measuring the validation accuracy of LeNet over 15 epochs by informing how long in seconds LeNet took to train with each resolution.

Resolution	32x32	64x64	128x128	164x164
Time until				
convergence	75	75	120	195
(seconds)				

Table 2: This datatable gives a context for the time taken by the AlexNet CNN architecture to converge with each given image resolution. This table helps to contextualize the line graph measuring the validation accuracy of AlexNet over 15 epochs by informing how long in seconds AlexNet took to train with each resolution.

Resolution	32x32	64x64	128x128	164x164
Time until				
convergence	75	105	255	360
(seconds)				

Table 3: This datatable gives a context for the time taken by the Encoder-Decoder CNN architecture to converge with each given image resolution. This table helps to contextualize the line graph measuring the validation accuracy of Encoder-Decoder over 15 epochs by informing how long in seconds Encoder-Decoder took to train with each resolution.

Resolution	32x32	64x64	128x128	164x164
Time until convergence (seconds)	90	135	375	540

4. Discussion

These results all show that when inputting a complex image in a significantly downscaled form, the neural network overall performs worse. In our specific case, the neural networks performed around 10% worse with the smallest resolution compared to the largest resolution. As shown by Figure 3A and 3B, AlexNet performed the best overall. AlexNet was able to achieve the highest accuracy of 57.2%, compared to 50.2% for Encoder-Decoder and 51.3% for LeNet. One problem that we encountered when performing this research is that our computer, with our Nvidia GTX 1660 GPU, was not able to handle image sizes larger than 164x164 resolution. Given that the original image dataset consisted of images of resolution 512x512, even our highest resolution image at 164x164 was significantly downscaled. Therefore, some of our results may be different if this experiment had been performed with a stronger computer that was able to handle image resolutions up to 512x512. As image size increases up to 512x512, it is very likely that the accuracy of these convolutional neural networks also increases.

Additionally, we have considered that, for countries of lower than average socioeconomic status, such as developing countries, there are serious developmental challenges in terms of technology due to a range of factors. These include low levels of investment and low rates of education and skilled labor. This leads to lower levels of diffusion of technology and innovation (Utoikamanu, 2019). This, coupled with low income, results in setbacks similar to those that were exhibited during our research. Based on our data and the findings above, it is seen that a certain level of computer specifications, as well as dataset image quality, must be met in order to gain accurate, meaningful results for neural networks. When using a midrange specification computer with detailed images of resolutions ranging from 32x32 up to 164x164, it is not possible to accurately train a neural network to modern standards (greater than 90% accuracy). As shown by our research, one would need to procure a much more powerful computer with more RAM, a state of the art CPU, and a GPU with more memory and more parallel processing cores (Christensson, 2006; Dettmers, 2020). In addition to high end computer specifications, a large dataset with detailed, colored, high resolution images is also required to achieve high end accuracy.

If, for any reason, both of these necessities are not adequately met, the results may be very subpar. When a dataset is used for a medical or some other professional, important purpose, image resolution and image quality should be as high as possible to ensure the greatest accuracy. Downscaling the image to anything below 50% of the image size will result in accuracy that, at most, can reach up to 60%.

5. Conclusion

Neural networks are useful in a variety of circumstances, from medical disease detection to autonomous driving. In order to tackle more complex issues, computer vision is needed, which requires the use of convolutional neural networks, which use image sets as the input values. There are different architectures of neural networks that were created for different purposes, including the three which we used LeNet, in our project: AlexNet, and Encoder-Decoder. The main purpose of our project is to simulate the effectiveness of having a less than optimal setup when trying to train a convolutional neural network, as we wanted to show the limitations of our hardware and the difficulties of finding adequate datasets. We chose a more complex topic, which is the detection of diabetic retinopathy from a given retinal scan. We specifically chose this topic as classifying disease states from images is highly relevant for application in the medical field. Benchmarking the performance of CNNs on medical images would be very beneficial to doctors and

patients. We measured the accuracy of the three neural networks' detection capabilities when given a range of lowered resolutions of images: 32x32, 64x64, 128x128, and 164x164. Based on our findings, as the image resolution increased, so did the accuracy of the neural networks. However, all the neural networks struggled, with the maximum accuracy being only 57.2%. Some issues that we faced included our computer specifications not being powerful enough to handle larger image resolutions, as well as the use of simple architectures, which likely led to underperformance. These results imply that, if both high specification of computer and quality image datasets are not utilized, the neural network will not be properly trained, causing large inaccuracies when compared to modern standards. For countries which do not have the required infrastructure or funding to provide these resources, it will be difficult to achieve good results when using convolutional neural networks. Further experiments need to be conducted to research the lowest possible resolution and least complex hardware required to sufficiently train an accurate network.

Acknowledgements

We would like to acknowledge Pramod Madabhushi and Madhuri Madabhushi for their help editing and looking over this paper through its multiple revisions. They also aided us in the organization and logistics of the project, and the funding for this project.

References

Christensson, P. (2006). *VRAM Definition*. TechTerms. https://techterms.com

Dettmers, T. (2020, September 7). *Which GPU(s) to Get for Deep Learning: My Experience and Advice for Using GPUs in Deep Learning.* TimDettmers.com. https://timdettmers.com/2020/09/07/which-gpu-for-d eep-learning/

Gupta, A. (2020, May 17). Evolution of Convolutional Neural Network Architectures. Medium.

https://medium.com/the-pen-point/evolution-of-conv olutional-neural-network-architectures-6b90d067e40 3 Hooker, S. (2021). "Moving Beyond 'Algorithmic Bias is a Data Problem." *Patterns*, 2 (4). https://doi.org/10.1016/j.patter.2021.10024

Huang, Y. & Chen, Y. (2020). "Autonomous Driving with Deep Learning: A Survey of State-of-Art Technologies." *ArXiv*. https://arxiv.org/abs/2006.06091

LeCun, Y., et al. (2015). "Deep Learning." *Nature*, 521. doi:10.1038/nature14539

Mayo Clinic. (2021). *Diabetic Retinopathy*. The Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/diab etic-retinopathy/symptoms-causes/syc-20371611

Mestre, D. (2018, September 14). *How to Solve the Common Problems in Image Recognition*. Medium. https://medium.com/empathyco/how-to-solve-the-common-problems-in-image-recognition-d519af322bea

O'Shea, K. & Nash, R. (2015). "An Introduction to Convolutional Neural Networks." *ArXiv*. https://drive.google.com/file/d/1UBBLcqWO9zvXuj NuVCaN3y9vUTb3L-M-/view?ts=6122a775

Shen, D., Wu, G., & Suk H. (2017). "Deep Learning in Medical Image Analysis." *The Annual Review of Biomedical Engineering*, 19. https://doi.org/10.1146/annurev-bioeng-071516-044442

Tandon, K. (2021). *Diabetic Retinopathy Balanced* (Version 1). [Data set]. Kaggle. https://www.kaggle.com/kushagratandon12/diabetic-r etinopathy-balanced

Utoikamanu, F. (2019). "Closing the Technology Gap in Least Developed Countries." *UN Chronicle*, 55 (4).

https://www.un.org/en/chronicle/article/closing-techn ology-gap-least-developed-countries

Zhou, V. (2019, March 5). *Machine Learning for Beginners: An Introduction to Neural Networks*. Towards Data Science.

https://towardsdatascience.com/machine-learning-for -beginners-an-introduction-to-neural-networks-d49f2 2d238f9

Screening P2P Loans Beyond Loan Grade: An Application of Machine Learning Algorithm XGBoost

Selina Sun^{1*}

¹Thomas Jefferson High School of Science and Technology, Alexandria, VA USA

Received October 8, 2021; Revised April 11, 2022; Accepted, April 19, 2022

Abstract

The credit risks of P2P loans fall onto the investors rather than the platforms, which provides a motivation for investors to screen loans beyond the suggestive loan grade provided by the platforms. This is made possible by P2P platforms sharing with the public the same information they use in screening. This paper studies whether screening models can be developed using machine learning algorithms to better screen P2P loans than loan grade. Based on the loans listed by one of the largest P2P platforms in the world, Lending Club, we find that machine learning algorithm XGBoost can improve default rank ordering by 10.3% in sample and 4.4% out of sample, as compared with loan grade. XGBoost also outperforms the conventional parametric Logistic regression model with a performance gain of 4.4% in sample and 1.1% out of sample.

Keywords: Marketplace Lending, Machine Learning, Logistic Regression

1. Introduction

Peer-to-Peer (P2P) lending is a practice of lending money to businesses and consumers facilitated by online lending platforms that connect borrowers and investors. Since its inception with Prosper in 2006 in the US, P2P lending has originated more than \$18 billion loans by 2015 and is expected to grow further to \$150 billion by 2025 (Prime Meridian, 2015).

P2P lending is also called crowdfunding, social finance, marketplace lending, or disintermediation finance, each emphasizing different perspectives of P2P lending. One characteristic that makes P2P lending different from conventional bank lending is the disintermediation, in that P2P platforms do not serve as the intermediators by holding loan originations on their balance sheets, and the credit risks of P2P loans fall directly onto the investors rather than the platforms. Therefore, the objectives of

online platforms may not be necessarily aligned with investors (Vallee and Zeng, 2019). On the other side, online platforms share with the public the same information that they use to screen loans. In the US, the P2P market regulator, the U.S. Securities and Exchange Commission (SEC), demands that online platforms disclose all decisioning information to the public. Therefore, sophisticated and active investors might be able to better predict credit risks than inexperienced and passive investors who rely on the screening tools in the form of loan grade assigned by online platforms.

The objective of this paper is to study whether screening models can be developed to better capture the default risks of P2P loans than the suggestive loan grades provided by the platforms. It hypothesizes that customized screening models can predict default risks more accurately than loan grades. We take the platform grade as the baseline and evaluate the

Advisor: Yan Zhang yan.zhang@occ.treas.gov

^{*} Corresponding Author selinateam14@gmail.com

benefits of utilizing the more granular raw data collected at loan application, including the borrower-provided soft information and the standard borrower credit bureau profile. Our study is based on Lending Club (LC) data. Being the second earliest P2P platform in the US and the first ever to register its offerings as securities, Lending Club is one of the largest P2P lending platforms in the world. It is the largest P2P lending platform in the US and took approximately 35%-50% of the US market share in 2015-2020 (IBISWorld, 2020). As of 2020, LC has reached a total loan volume of \$50 billion since its launch in 2006 ("Top 7 P2P lending sites to lend online from money US", https://crowdfunding-platforms.com/top-7-p2p-lendin g-sites-to-lend-money-online-from-us). Therefore, the results based on LC could be of broader interpretation.

In terms of modeling algorithms, we study both the traditional parametric statistical models and the innovative machine learning (ML) models. Since we measure loan risks by a binary classification of loan default, the statistical model used is the Logistic regression (LR) model and the ML model is eXtreme Gradient Boosting (XGBoost, or XGB). The latter is a ML algorithm that is gaining wide popularity in credit risk modeling (Chang et al., 2018) but has not been studied as a P2P screening algorithm in an empirical study of LC loans, based on our knowledge.

We find that our models perform better in predicting default risks than platform grades. Given the LC loans used in the paper, the XGB model achieves a 10.3% increase in rank ordering defaults in sample and a 4.4% increase out of sample. The XGB model is also able to classify defaults more accurately than the Logistic regression model with rank ordering gain of 4.4% in sample and 1.1% out of sample. While loan grade is found to be reasonably predictive, other loan and borrower variables can be utilized to screen loans further, including payment-to-income (PTI) ratio, debt-to-income (DTI) ratio, Fair Isaac Corporation (FICO) score, and number of accounts opened in the past 24 months.

The paper contributes to the strand of emerging literature studying the screening ability of lenders in marketplace lending. One sub-group of this literature identifies information that can be used to approximate platform grade or even further to screen loans on top of platform grade. Using Prosper data, Iver et al. (2015) show that investors can gain by leveraging the number of current delinquencies, debt-to-income ratio, amount delinquent, and number of credit inquiries in the last six months in addition to the FICO-based grade. Based on Lending Club data, Emekter et al. (2015) identify DTI, FICO, and revolving line utilization as contributing to loan screening alongside loan grade. Serrano-Cinca et al. (2015) suggest that besides loan grade, loan purpose. annual income, homeownership, credit history, and borrower's indebtedness are relevant to risks of LC loans. Another sub-group of the literature tests the use of financial technology in the form of ML algorithms in screening. Malekipirbazari and Aksakalli (2015) compare Random Forest (RF), Support Vector Machines (SVM), K-Nearest Neighbor (KNN), and Logistic regression in predicting loan defaults and find RF performs the best. Chang et al. (2015) conclude that Naïve Bayes performs better than SVM and Logistic regression in LC loan default prediction. This paper evaluates the ML algorithm of XGB, which has not been documented in existing literature for screening P2P loans. Our analysis is based on a LC dataset with a much more recent snapshot and therefore includes more loans. We assess model screening ability not only in sample but also out of sample.

The rest of the paper is organized as follows. Section 2 discusses the LC data and defines the analysis samples. Section 3 presents model specifications and performance measures. Section 4 reports the regression results. Section 5 concludes.

2. Data

The study uses a Lending Club dataset that was downloaded on July 3, 2021 from the Kaggle website (https://www.kaggle.com/wordsforthewise/lending-cl ub). It contains all the loans accepted by Lending Club during the period from June 2007 (the beginning of its loan issuing) to December 2018. In aggregate, there are 2.26 million loans with a total loan amount of \$34 billion. Each loan is reported as one data record with a rich set of data fields (151 variables in total) including information collected at the application (loan application submission and borrower credit pull), LC assigned risk grade and pricing, and loan performance read by the end of 2019Q1 (loan status, hardship program, settlement program, etc.).

LC loans are in the form of unsecured personal loans with fixed term and installment (i.e., monthly payment). The loan amount is in the range of \$500 to \$40,000, and the loan term includes 3-year and 5-year. The 5-year term loans were not rolled out until May 2010. Out of the two product terms offered by LC in the data, roughly 71% are of 3-year term. To obtain a more homogenous sample, we focus on the 3-year loans. Most (95%) of the loan applications are filed by individuals, so we exclude joint accounts to simplify the analysis without too much loss of generosity. As this paper evaluates loan credit risks by following up loan performance within two years of the origination date, we also drop loans that are originated after February 2017. They are right censored by the data read snapshot of 2019Q1 and therefore do not have the full 2-year performance window imposed. The LC loan volume has been growing at an accelerating pace and the originations from March 2017 through December 2019 take up 36% of the total loan volume. Imposing a shorter 1-year performance window allows us to include more loans into the analysis, but the performance window is too short to have a good read of loan performance. On the other side, if we go with the full 3-year performance window, even more loans will be excluded from our analysis, and the gain is marginal - not many loan defaults occur in the third year.

After the exclusion of 5-year termed loans, joint loans, and loans originated after February 2017, we have an analysis data sample of 982,245 loans. The application fields are well populated, however, some of the credit bureau fields have missing values. We impute the missing values with the median of the variable, so as not to distort the variable distribution. For variables with high missing rates, we also construct corresponding missing variable indicators to account for possible implications of the missing patterns to loan risks.

When borrowers submit their application to the LC online platform, they need to disclose loan

amount (loan amnt), loan purpose (purpose), annual income (annual inc), home ownership (home ownership), employment length and (emp length). In real time, the platform processes the applications and makes underwriting and pricing decisions using its screening models. The borrower credit bureau information is pulled and leveraged in the credit decisioning. The LC dataset discloses borrowers' FICO score in the data in the form of a range with upper bound and lower bound (fico range high and fico range low). The DTI ratio (dti) is calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income. Based on the date when the borrower's earliest reported credit line was opened (earliest cr line date), we can infer how long the borrower's credit history is (credit history). LC also makes public the detailed credit profile of borrowers in the form of credit bureau variables, such as the number of trades opened in past 24 months (acc open past 24mths), the number of inquiries in the past 6 months (inq last 6mths), and the percent of bankcards with utilization greater than 75% (percent bc gt 75). For approved loans, the installment is calculated and reported given the loan amount, term, and interest rate. We create a new variable, PTI ratio (pti), as the installment amount over the borrower's self-reported monthly income.

The loan status variable reports loan performance in the following nine status: 1) "Charged Off", 2) "Current", 3) "Default", 4) "Does not meet the credit policy. Status: Charged Off", 5) "Does not meet the credit policy. Status: Fully Paid", 6) "Fully Paid", 7) "In Grace Period", 8) "Late (16-30 days)", and 9) "Late (31-120 days)". The hardship program field flags borrowers who encounter hardship and receive assistance, and the settlement field identifies troubled borrowers who settle debts with lenders. This paper measures loan credit risks by a user defined technical "default" variable that takes the value of 1 if a loan is more than 30 days late or worse within 2 years since loan origination, and 0 otherwise. Specifically, a loan technically defaults if it reaches the status of 1) "Charged off", 3) "Default", 4) "Does not meet the credit policy. Status: Charged Off", 9) "Late (31-120 days)", or receives hardship assistance, and debt settlement.

Table 1 reports the definition and summary statistics of the data fields used in our analysis. Panel A lists the loan characteristic variables. On average, the default rate and interest rate are very close at 11.9% and 12.0%, respectively. The mean loan amount funded is \$12.6K. The applied loan amount and funded loan amount are very similar, as shown by the small loan amount difference (amnt_diff), suggesting that the requested loans are pretty much funded fully.

Table 1 Panel B provides information on the application variables. The annual income is reported to be \$74.3K on average. The data report detailed employment history in 12 categories of not reported, < 1 year, 1 year through 9 years with the increment of 1 year, and 10+ years. We translate the categorical employment length into a continuous employment length by coding "not reported" as 0, "< 1 year" as 0.5, "10+ years" as 10, and other employment length categories as they are. The average employment length is 5.5 years. Loan purpose has the following categories: Car, Credit card, Debt consolidation, Educational, Home improvement, House, Major purchase, Medical, Moving, Renewable energy, Small business, Vacation, Wedding, and Other, among which the top three categories are Debt consolidation, Credit card, and Home improvement, taking up 53.8%, 23.9%, and 6.6% of the total loans, respectively. Home ownership shows that 45.0% of the borrowers own homes with mortgage, 11.7% own homes paid off, and 43.2% rent. Types of income verification status include not verified, source verified, and verified, each of which takes up about a third of the total. LC verifies income if it determines that the borrower's requested loan amount is too high relative to the self-reported income. The income verification involves checking W-2 or paystub (source verified) or reaching out to the employers (verified). A loan can be listed as factional or as whole, the latter is to facilitate institutional investment. Factional loans consist 45.6% of the total and the remaining 54.4% are whole loans. The whole listing did not start until October 2012 and has since increased continuously to the level of close to 90% by the end of 2018.

As for borrower creditworthiness as reported in Panel C of Table 1, the mean DTI ratio is 17.8. LC imposes a maximum DTI restriction for loan approval. Most of the time the DTI cap is set at 39.99%, but in the first half of 2016, it rose to a concerning level of 50%. The lowest FICO score on average is 694. Similar to DTI, LC requires borrower FICO to be above a certain level to list a loan. With a few fluctuations in the earlier years, the FICO threshold has been 660 since November 2008. The credit history, which measures the number of years since the borrower initially establishes a credit profile, is 16.2 years on average. LC requires at least 3 years of credit history to approve a loan. Out of the many bureau variables, this paper focuses on the top 15 that are shown to be highly related to loan risks. They are selected by an XGB model with number of iterations = 500, maximum tree depth = 3, learning rate = 0.05, minimum loss split = 0.05, minimum child weight = 5, lambda = 1, alpha = 0, subsample ratio of columns = 0.8, and subsample = 1. The section "Research Methodology" provides a detailed explanation of XGB and its hyperparameters. Other variable selection methods like stepwise regression lead to very similar variable selections.

Table 1: Variable definition and summary statistics

Variable	Definition	Mean	SD
variable	Definition	wieum	50
default	loans become 30+ DPD or worse within 2 years of loan issuance	11.9	
int_rate	loan interest rate	12.0	4.0
funded_amnt	loan amount funded	12,551	8,079
amnt_ diff	difference between the listed loan amount and funded loan amount	4	175

Panel A: Loan characteristics

Panel B: Application information

Variable	Definition	Mean	SD
annual_inc	self-reported annual income provided by the borrower during registration	74,348	69,908
pti	ratio of the monthly payment of the loan (installment) to borrower self-reported monthly income	7.7	4.3
emp_ length	employment length in years	5.5	3.8

purpose: debt consolidation	loan purpose is to consolidate debts	53.8%	
purpose: credit card	loan purpose is to pay off credit card balances	23.9%	
purpose: home improvement	loan purpose is for home improvement	6.6%	
home: mortgage	home is owned with a mortgage	45.0%	
home: rent	home is rented	43.2%	
home: own	home is owned without a mortgage	11.7%	
verification: not verified	income is not verified	32.5%	
verification: source verified	income source is verified	38.2%	
verification: verified	income is verified	29.3%	
initial listing status: fractional	loan listed as fractional	45.6%	
initial listing status: whole	loan listed as whole	54.4%	

Panel C: Borrower creditworthiness

Variable	Definition	Mean	SD
dti	ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income	17.8	8.4
fico_range_low	lower boundary of the borrowers FICO range at loan origination	695	31
credit_history	credit bureau history in years 16.2		7.7
acc_open_past _24mths	number of trades opened in past 24 months	4.5	3.1
avg_cur_bal	avg_cur_bal average current balance of all accounts		15,593
bc_open _to_buy	total open to buy on revolving bankcards	9,397	14,490
bc_util	ratio of total current balance to high credit (or credit limit) for all bankcard accounts	60.7	27.1
delinq_2yrs	number of 30+ days past-due incidences of delinquency in the borrowers' credit file for the past 2 years	0.3	0.9
inq_last_6mths	number of inquiries in past 6 months (excluding auto and mortgage inquiries)	0.7	1.0
mo_sin_old _rev_tl_op	months since oldest revolving account opened	180.3	93.9

mort_acc	number of mortgage accounts	1.5	1.9
mths_since recent bc	months since most recent bankcard account opened	23.7	30.6
mths_since _recent_inq	months since most recent inquiry	6.8	5.5
num_tl_op _past_12m	number of accounts opened in past 12 months	2.1	1.8
percent_bc_gt_ 75	percentage of all bankcard accounts > 75% of limit	45.2	35.2
revol_util	revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit	52.2	24.1
tot_hi_cred _lim	total high credit or credit limit	158,580	172,866
total_bc_limit	total bankcard high credit or credit limit	20,469	20,912

Notes: The statistics are generated based on a sample of Lending Club loans of single applicant 3-year loans originated during June 2007 through February 2017 with performance snapshot of 2019Q1, total 982,245 loans.



Figure 1. Loan volume, pricing, and performance over time

Figure 1 plots the loan volume, pricing, and performance of the LC loans along the loan issuance date. The loan volume was quite small in the beginning with only 23 loans issued in the LC debut month of June 2007. The loan volume gradually picked up and reached 1,000 loan issuances in June 2011. LC business grew rapidly in the next 4-5 years. Since 2015 the monthly origination volume has been around 25,000. At the peak, 44,000 loans were originated in the single month of March 2016. The interest rate has been relatively stable, fluctuating in the range of 10%-14%. Conversely, the default rate

was quite volatile. It was high up to 2009, which could be due to the nascency of the business and the financial crisis of 2007-2009. Afterwards the default rate decreased and stayed below the interest rate line. In the extreme month, the interest rate was as much as 4% higher than the default rate. But starting from 2015 when the loan volume stabilized at the high level, the default rate became higher than the interest rate.



Figure 2. Loan volume, pricing, and performance by loan grade

Figure 2 presents the loan volume, pricing, and performance along the dimension of LC loan grade. LC assigns grades to inform investors of its assessed risks. The grade system contains 6 letter groups from A to G with increasing risks. Within each grade, there are 5 numeric subgrades from 1 to 5 with increasing risks. For simplicity this paper refers to the letter and number combined grade as grade unless otherwise noted. Figure 2 shows that the loan distribution is right skewed with grades A through D taking almost 96% of the loans and grade B being the most populated with 35% of the loans. The interest rate is the same for loans with the same grade at the same issuance time and is adjusted periodically. Figure 2 shows a very clear monotonic increasing trend of the interest rate with grade, confirming the mapping between grade and interest rate. The best grade A1 has an average interest rate of approximately 2%, and the worst grade G5 has an average interest rate as high as 40%. The default rate also monotonically increases with grade, suggesting that LC grade rank orders risks. However, the interest rate line and the default line have different slopes. For loans with good grades, the interest rate is higher than default

rate on average, so on aggregate investing highly rated loans returns positive cash flow. But for riskier loans, the interest rate is not high enough to cover the losses. This echoes the finding by Emekter et al. (2015) that the high risks of low-grade loans are not covered by the interest rate charged.

The dataset is then split into training and testing by issue year. In earlier years of LC business, the number of loans were few, the screening models were still at their learning stage, and the default rates were high. We decide to drop this small volume of loans and use loans issued during January 2011- December 2015 for training the screening models, and loans issued afterwards (January 2016 - February 2017) for testing. Our testing sample is not only out of sample but also out of time, enabling a more rigorous and meaningful testing. The training and testing datasets have 603,497 and 361,319 observations, respectively. To facilitate the tuning speed, we randomly sample 30% of the training sample to construct the hyperparameter tuning sample of the XGB models. The tuning dataset contains 181,049 loans, which is sufficiently large for the tuning purpose.

We develop screening models using two types of algorithms: the LR model and the XGB model. In predicting credit risk of retail loans with binary outcomes, Logistic regression is a commonly used conventional statistical model, while machine learning algorithms like XGB are gaining popularity. Variable contribution to screening is shown by variable importance. The model performance is evaluated by model rank ordering of loan defaults.

3. Research Methodology

3.1. Logistic Regression

The Logistic regression model is a parametric statistical model used to predict the probability of a binary outcome. It takes inputs (or independent variables, explanatory variables) and estimates a linear function of them, then the weighted sum of the inputs will be transformed through a Logistic function to output a continuous number between 0 and 1 as the predicted probability of the outcome (or dependent variable). The functional form of the Logistic regression can be written as

$$logit(Y) = log log \left(\frac{Y}{1-Y}\right) = \beta X$$

where *Y* is the binary dependent variable which equals 1 for an event and 0 for a non-event, and *X* is the vector of the independent variables. The β are the model coefficients which can be estimated by maximizing the log-likelihood function. With the estimated $\hat{\beta}$ the response probability can be predicted as

$$P(Y = 1) = \frac{e^{\hat{\beta}X}}{1 + e^{\hat{\beta}X}}$$

The modeling analysis of this paper is implemented in R 3.6.1. We fit the Logistic regression model using the *glm* function available in base R.

3.2. XGB

XGB (Chen and Guestrin, 2016) is a type of ensemble model. Arguing that one single learner might be unstable, ensemble models train multiple learners with a common objective to generate a more robust learner. XGB leverages the boosting technique under which learners are built sequentially until no further improvement can be made. Boosting creates multiple new training datasets via random sampling with replacement over weighted data and the previous learner informs the weight. XGB is shown to be effective in reducing variations and improving stability and is gaining wide popularity. This paper will calibrate and estimate XGB models using *R caret* package with the *xgboost* package running in the background.

In machine learning models, the dependent variable is often called the label and the independent variables are called the features. Corresponding to variable coefficients in parametric models, machine learning models are specified by hyperparameters. The XGB hyperparameters include nrounds (number of iterations), max_depth (maximum tree depth), eta (learning rate), gamma (minimum split loss), min_child_weight (minimum child weight), lambda (L2 regulation on leaf weights, equivalent to Ridge regression if equals 1), alpha (L1 regulation on leaf weights, equivalent to LASSO regression if equals 1), colsample bytree (subsample ratio of columns), and

subsample (subsample percentage). XGB models with larger number of iterations, deeper trees, smaller learning rate, smaller split loss, and smaller child weight can achieve better fitting of the development sample; but might overfit with less comparable performance on the testing sample. We calibrate the key hyperparameters to reduce variations of model performance across samples. A grid search is performed for combinations of the hyperparameter values to identify a set of hyperparameters leading to the best performance. In addition, this paper uses the 5-fold cross-validation to address potential overfitting issues. The training dataset is divided equally into five random samples and each time four of them are used for model training and the remaining one sample is used for validation. For a combination of hyperparameters, the overall performance is the average of the five performances on the validation sample as it takes one of the five samples. In the end, the combination of hyperparameters resulting in the best performance is selected as the optimal hyperparameters. The final model is then estimated using the optimized hyperparameters on the entire training sample.

Table 2: Hyperparameter tuning

Hyper parameters	max tree depth	eta (learning rate)	gamma (min split loss)
search range	5, 6, 7, 8	0.01, 0.05, 0.1	0.01, 0.05, 0.1
XGB-X0	5	0.05	0.05
XGB-X1	5	0.05	0.01

Notes: The XGB hyperparameters are tuned using the tuning sample of a 30% random sample of the training dataset with total observations of 181,049 observations.

We tune three key hyperparameters of XGB: maximum tree depth, learning rate, and minimum loss split. Table 2 Row 1 shows the search ranges of the hyperparameter tuning. We test maximum tree depth from 5 through 8, and learning rate and minimum loss split for the values of 0.01, 0.05, and 0.1, respectively. In total, 36 combinations of hyperparameters are evaluated for each model. We set the number of iterations to be 500, which is reasonably large. The minimum child weight is set fixed at the value of 5. The hyperparameter lambda is set at 1, alpha is set at 0, subsample ratio of columns is set at 0.8, and subsample is set at 1. Rows 2-3 of Table 2 report the optimized hyperparameters of the XGB model with feature set X0 and X1. Details on the feature sets are provided in the next sub-section.

3.3. Model Specification

We use the LC assigned interest rate as the baseline for default prediction. As the interest rate is assigned by the loan grade, the predictive power of interest rate reflects the quality of grade in signaling loan quality. To calibrate the difference between the 2-year default probability and the interest rate of 3-year maturity, we estimate a Logistic regression of default with interest rate as the only explanatory variable. We record the benchmark model as LR-X0. Similarly, we develop a XGB benchmark model, XGB-X0, with interest rate as the only feature. We expect interest rate is better calibrated under XGB-X0 than LR-X0 since ML can capture nonlinear relationships better than LR, which is a generalized linear regression.

The feature set used to develop the screening models contains interest rate, loan and application information (funded loan amount, amount difference between the applied and funded loan amount, PTI, employment length, loan purpose, home ownership, listing status, and verification status), key bureau variables (DTI, the lowest FICO), and the 15 selected bureau variables. We also control for loan issue year and residence state. The corresponding screening models are labeled as LR-X1 and XGB-X1.

3.4. Performance Measure and Model Interpretation

Model performance is evaluated for its ranking ordering. The performance metric used for model rank ordering is AUC. With a range from 0 to 1, the AUC reflects the area under the receiver operating characteristic (ROC) curve. The larger the AUC, the better the model classifies events. The ROC curve plots the true positive rate (TPR) on the y axis and false positive rate (FPR) on the x axis, so ideally the ROC curve should be pushed to the edge of the upper left corner with 100% TPR and 0% FPR and area under the curve is maximized to be 1.

Unlike parametric statistical models whose model specification can reveal the relationship between model inputs and model outputs, dubbed as a "black box", ML is more complicated and less transparent. We report variable importance of the models to help understand the contributions of variables to loan screening. The variable importance of LR is ranked by the magnitude of the standardized coefficients, and the variable importance of XGB is ranked by the impurity importance which is defined as the improvement in the performance measure (the area under the curve, AUC, is used in this paper) attributed to the splitting variable at each split in each tree, weighted by the number of observations the node is responsible for. The feature importance is then accumulated over all the trees in the model for each variable. The variable importance is scaled to the variable with the top importance with a value of 0 to 100.

4. Results

Table 3 reports the model performance in rank ordering defaults. If calibrated by LR, the interest rate has a training AUC of 66.0 and a testing AUC of 67.7. The performance is similar under XGB with a training AUC is 67.0 and a testing AUC of 66.9. If investors develop customized screening models supplementing interest rate with publicly available information on loan and borrower, their screening ability can be improved even more. The LR-X1 has a training AUC of 69.7 and a testing AUC of 69.9, and the XGB-X1 has a training AUC of 72.8 and a testing AUC of 70.7. Comparing XGB-X1 with LR-X0 shows that by leveraging both the full set of granular information and loan grade, and the more sophisticated ML algorithm, the performance of screening is improved by 10.3% (AUC 72.8 vs. 66.0) in sample and 4.4% (AUC 70.7 vs. 67.7) out of sample. In utilizing the full information set to develop the screening model, the XGB model performs better than the traditional statistical model, with an AUC gain of 4.4% (72.8 vs. 69.7) in sample and 1.1% (70.7 vs. 69.9) out of sample.

The better performance of machine learning algorithm XGB over LR could be attributed to its ability to fit complicated data patterns such as nonlinearities and interactions. For example, the current specification of LR assumes a linear relationship between default and FICO, but the relationship could potentially be nonlinear. Sourced from the same borrower credit profile, FICO could be related to other bureau variables, which is not accounted for explicitly in the LR model used. We may specify more complicated parametric models to account for possible nonlinearities or interactions, but ML models can learn the data patterns by themselves, resulting in more flexible and better model performance.

Table 3: Model performance comparison

	LR-X0	XGB-X0	LR-X1	XGB-X1
Training	66.0	67.0	69.7	72.8
Testing	67.7	66.9	69.9	70.7

Table 4: Model variable importance comparison

		1	
LR-X1		XGB-X1	
int_rate	100	int_rate	100
pti	49	pti	15
acc_open_past_ 24mths	43	acc_open_past_ 24mths	14
dti	42	dti	10
fico_range_low	28	issue_y2015	9
mths_since_ recent_bc	24	fico_range_low	8
percent_bc_gt_75	20	tot_hi_cred_lim	6
inq_last_6mths	18	emp_length	5
mo_sin_old_ rev_tl_op	18	total_bc_limit	5
emp length	18	avg cur bal	5

Table 4 lists the 10 variables with top importance. Both LR-X1 and XGB-X1 list interest rate as the most important variable. Both models list PTI, number of accounts opened in the past 24 months, and DTI as the next three most important variables. The FICO score is not ranked as high with consideration of more granular bureau variables. Overall, this paper finds that LC loan grade is highly effective in risk screening, however, application and borrower information can supplement loan grade to further improve loan screening. Such variables are PTI, number of accounts opened in the past 24 months, DTI, and FICO.

5. Conclusion

In P2P lending, the platforms do not serve as the intermediator and therefore do not bear the credit risks of the listed loans. Investors take on the credit risks directly and therefore the objectives of platforms and investors could potentially be misaligned. This creates a motivation for P2P investors to actively screen loans rather than depend on the suggestive grade provided by platforms. Platforms are required to share with investors the same information they used in approving loans and assigning grades, making it feasible for investors to screen loans beyond loan grade. This paper shows the possibility and value of active screening. Leveraging both the full set of granular information and loan grade, and under the more advanced ML algorithm, the performance of screening is improved by 10.3% in sample and 4.4% out of sample in rank ordering. also outperforms the conventional XGBoost parametric Logistic regression model with a performance gain of 4.4% in sample and 1.1% out of sample.

References

Chang, Y. C., Chang, K. H., & Wu, G. J. (2018). Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Applied Soft Computing Journal*, 73: 914–920. https://doi.org/10.1016/j.asoc.2018.09.029

Chang, S., Kim, S. D., & Kondo, G. (2015). Predicting default risk of lending club loans. *Machine Learning*, 1–5.

Chen, T., and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 785–794. ACM. https://doi.org/10.1145/2939672.2939785

Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Applied Economics*, 47(1): 54–70. https://doi.org/10.1080/00036846.2014.962222

IBISWorld (2020, December 17). *Peer-to-Peer Lending Platforms Industry in the US – Market Research Report.* https://www.ibisworld.com/united-states/market-rese arch-reports/peer-to-lending-platforms-industry/

Iyer, R., A. Khwaja, E. Luttmer, and K. Shue. (2015). Screening peers softly: Inferring the quality of small borrowers. *Management Science*, 62:1554–77. https://doi.org/10.1287/mnsc.2015.2181

Malekipirbazari, M., & Aksakalli, V. (2015). Risk assessment in social lending via random forests. *Expert Systems with Applications*, 42(10): 4621–4631. https://doi.org/10.1016/j.eswa.2015.02.001 Prime Meridian (2015, December 28). *Marketplace lending in 2015 – a year of performance and growth.* https://www.pmifunds.com/marketplace-lending-in-2 015-a-year-of-performance-and-growth/

Serrano-Cinca, C., Gutiérrez-Nieto, B., & López-Palacios, L. (2015). Determinants of default in P2P lending. PLoS ONE, 10(10), 1–22. https://doi.org/10.1371/journal.pone.0139427

Vallee, B., Zeng, Y., (2019). Marketplace lending: a new banking paradigm? *Review of Financial Studies*, 32(5): 1939-1982. https://doi.org/10.1093/rfs/hhy100



Huntington's Disease - A Breakthrough in Treatment using CRISPR/Cas9

Sanika Sharma^{1*}

¹Irvington High School, Fremont, CA USA

Received September 6, 2021; Revised May 16, 2022; Accepted, May 30, 2022

Abstract

Huntington's disease is classified as a rare neurodegenerative disease that affects patients' motor skills. The disease is passed on from generation to generation due to it being a dominant genetic disease. Some symptoms of Huntington's disease include chorea and cognitive decline. Although there is no current cure, scientists believe that gene editing methods such as CRISPR/Cas9 could be a possible way to combat Huntington's. A couple of methods scientists used to research treatments for Huntington's, include using hiPSCs with CRISPR/Cas9 and personalizing the CRISPR/Cas9 treatment using specific PAM sites. These methods have not been tested clinically yet, and there is still more research to be done on whether this is entirely safe, however, there is still hope that this technique could become a cure in the future.

Keywords: Huntington's Disease, Gene Editing, CRISPR, Biology, Genetics

1. Introduction

Huntington's disease (HD) is a dominant, genetic disorder that causes neural degeneration. This disease is relatively rare, occurring in around 1 in every 10,000 people in the United States (Huntington's Disease Overview, Incidence and Prevalence of HD, n.d.). Many patients usually start developing symptoms between the age of 30 and 50, however it can begin as early as the age of two (Roos, 2010). HD is characterized by unwanted choreatic movement as well as motor, cognitive, and psychiatric disturbances (Roos, 2010). Chorea is classified as jerky, involuntary movement that initially takes place in the fingers or toes and eventually moves up to facial muscles and other parts of the body. In addition to the chorea, patients experience a decline in motor skills such as slurred speech or balance issues (What is Huntington's Disease?, n.d.). Day-to-day activities become harder,

and patients have a difficult time walking or standing. In addition to the many physical symptoms, HD patients also experience significant cognitive issues such as depression, apathy, or dementia (What is Huntington's Disease?, n.d.).

HD is caused by a DNA error in the huntingtin gene which is important since it makes proteins that help brain development before birth and plays a role in synaptic function (Roos, 2010). It also has an anti-apoptotic function which is crucial so that a cell can get rid of itself in the case of abnormalities such as improper cell division. Normally, the gene has less than 26 CAG letter repeats, but the faulty gene consists of more than 40 CAG repeats, which cause the protein to be too long (UC Davis Health, n.d.). Over time, the mutant protein misfolds and forms clumps in neurons, eventually leading to cell death by stopping necessary functions (Finkbeiner, 2011). The reason why many treatments or cures have not been developed so far, is because it is difficult to

* Corresponding Author sanika2311@gmail.com Advisor: Maiko Kitaoka mkitaoka@berkeley.edu

treat the mutation without completely removing the gene. As the huntingtin gene is essential to have in the body, a patient cannot afford to lose it, making this option unviable.

A Huntington patient's life is divided into 2 stages, at-risk preclinical and clinical (Roos, 2010). If someone has a parent with Huntington, doctors have enough information about genetic inheritance to know that the child could potentially develop the disease, labeling them as at-risk preclinical. People in this stage do not show any symptoms and have not manifested the disease yet, so it is part of preventative care to let the patient and doctor know to look out for signs (Roos, 2010). This stage comes to an end once they confirm that the patient begins to display symptoms and carry the extra CAG repeats on the huntingtin gene (Roos, 2010). As a person approaches the clinical stage, they start developing more symptoms and signs associated with the disease.

As of now, there are no cures for Huntington's, however there are treatments to help patients. Genetic counselors can aid both patients and families with the mental health symptoms and help explain what exactly the patient is going through (Roos, 2010). Additionally, physical and occupational therapists work with patients and families to ease symptoms; there are also medicines to provide relief. For example, dopamine receptor blocking, or depleting agents are used to treat chorea (Roos, 2010). Additionally, many antidepressants are used to help with patients' depression as a result of the disease. However, scientists are currently doing intensive research to look for cures for Huntington's and believe that the answer lies in gene editing with tools like CRISPR/Cas9.

2. Gene Editing

Gene editing has become a beneficial tool for scientists over the past few decades. Specifically, a tool called CRISPR/Cas9 has become popular in recent years as it is simpler, faster, and more efficient to use compared to other gene editing tools. CRISPR, also known as clustered regularly interspaced short palindromic repeats, is a fairly new technology that allows scientists to precisely and accurately edit the DNA of an organism to remove mutations or create new beneficial ones. CRISPR is based on the use of very specific and programmable nucleases that produce changes in regions of interest in genomes using double-strand breaks (DSBs). CRISPR directs a nuclease called Cas9 to a specific location to create the DSB. Later, the cell repairs the DSB by its natural mechanism and corrects or adds new sequences in the DNA strand. (Schmidt, et al., 2021) Another important component of the CRISPR technology is the guide RNA (gRNA) that guides the Cas9 nuclease to the target site in order to edit the genome. gRNA is made up of two parts: crispr RNA (crRNA) and tracr RNA. Once these components join and guide Cas9 to the site, they perform a sequence-specific cleavage by recognizing the base pairings and target sequence. ("Full Stack Genome Engineering," n.d.)



Figure 1: This figure shows the general steps associated with CRISPR/Cas9. Once the Cas9 enzyme creates a DSB in the DNA strand, new nucleotide bases are either added or removed, and finally repaired by cellular mechanisms such as NHEJ and HDR. NHEJ is a mechanism that simply joins the two ends of the strand together, while HDR repairs the DNA strand by matching the nucleotide pairs. This results in an addition or deletion of certain genes. (Rodríguez-Rodríguez, et al., 2019)

In addition to gRNA, CRISPR requires protospacer adjacent motif, or PAM, sites to accurately find the intended target site. PAM sequences are groups of around three to five nucleotides that are close to the target which help Cas9 identify where to bind to. The length of the PAM sequence helps determine the frequency of the target sites; longer sequences are found less frequently in the genome than shorter ones, while shorter sequences are more frequent. The length also helps figure out how persistent unintended off-target cuts will be. If a PAM site is long, there are less chances for there to be off-target cuts since the more specific the sequence is, the less likely it is for Cas9 to mess up while reading the nucleotides. However, if a PAM site is short, that means the sequences occur more frequently on the genome so the Cas9 might pick one that is not intended. Although a lot of research has been done on CRISPR, scientists are still not sure about the full effects off-target cuts might have on organisms. Nevertheless, there are many methods that are used to perfect gRNA design as much as possible to limit these cuts. One way the gRNA design is enhanced is to use the mutant version of Cas9 called nSpCas9. This mutant enzyme requires two nSpCas9 to make a DSB, each with its own sgRNA. Off-target activity is reduced by 50 to even 1,500 times using this method (Roos, 2010; Rodríguez-Rodríguez, et al., 2019).



Figure 2: Here, the sgRNA, modeled with the green, guides the Cas9 molecule to the target site near the PAM sites in blue and creates a DSB in the strand. (Full Stack Genome Engineering, n.d.)

Even though CRISPR has a lot of potential, it is important to be aware of its risks as well. In a study done at Columbia University, researchers created 40 embryos with a mutation responsible for blindness ("Lab Tests Show Risks of Using CRISPR Gene Editing on Embryos," 2020). They edited these embryos to get rid of the mutation, however they found that instead of editing the mutation, the chromosome the mutation with disappeared altogether. This change is extremely fatal for embryos and can alter many important functions. Furthermore, there is not enough evidence from research showing any long-term effects that CRISPR could have on a genome or person ("Lab Tests Show Risks of Using CRISPR Gene Editing on Embryos," 2020).



Figure 3: This figure shows the huntingtin gene on chromosome 4. The blue sequences model the pam sites and the orange triangles are the cut site. The green guide RNA is guiding the Cas9 to the specific target site to correct the mutation. (Reproduced from "Full Stack Genome Engineering," n.d.)

Despite its risks, CRISPR could still be extremely useful and promising. Many therapies and treatments for diseases such as sickle-cell, cancer, down syndrome, and Huntington's include the use of CRISPR. Although these therapies are still being researched, some are going through clinical trials and might become widespread treatments. Recently, some scientists conducted experiments with CRISPR to see if they could suppress the mutation in Huntington's disease and found promising results.

3. New Potential Therapies

As Huntington's disease does not have any cures right now, scientists are doing all they can to research potential ones. However, there are not any clinical trials right now for HD because of the limitations of CRISPR. Currently, CRISPR is mostly used for blood disorders such as sickle cell anemia or thalassemia because of how accessible and easy it is to treat blood cells. But, since this disorder mainly affects the brain and neurons, it is much more difficult to do treatments with gene editing in the brain due to its inaccessibility. Additionally, there is the issue of getting the CRISPR treatment to every affected cell in the human body. There are millions of cells in a person, and current technology is not equipped to deliver the treatment to every single cell yet. Nonetheless, there is still research going on in vitro related to treatments for Huntington.

An example is an experiment that a couple of scientists conducted in 2017 where they used human induced pluripotent stem cells (hiPSCs) and edited their DNA using CRISPR/Cas9. hiPSCs are special types of stem cells that are taken from a person and treated with different transcription factors that maintain their pluripotency, which is the ability of a stem cell to make other types of cells in the body (Xu, et al., 2017). hiPSCs are particularly useful since they are easy to culture in a lab and can become any desired cell type. Since Huntington affects neurons and muscle cells, scientists can use hiPSCs to create these different types of cells while having them remain genetically identical.

Because these cells are being cultured outside of a natural environment there are chances that they can pick up different types of undesirable mutations or changes, so remaining genetically identical is crucial to get accurate results throughout the study. After creating their gRNA, the scientists made different types of selections to see which cells received the correct CRISPR editing, did not receive it, received a mutant version, or received it and died (Xu, et al., 2017). This marker is useful when they go back to study the population of cells that CRISPR successfully altered. Once the CRISPR/Cas9 reached the cells, the researchers concluded that it successfully corrected the HD mutation in the cell population. They compared the corrected cells with normal ones and reported that the corrected ones were functioning normally (Xu, et al., 2017).

In another study, a group of different scientists attempted to do the same thing, just using a different approach. Here, the researchers used a personalized CRISPR/Cas9 strategy based on SNPs to target specific CRISPR/Cas9 sites targeting the mutant huntingtin gene (Shin, et al., 2016). First, the researchers identified eight of the most frequent huntingtin gene haplotypes. Upon identification, they looked for pairs of PAM sequences that were present in the mutant chromosome but not from the normal chromosome. Afterwards, the scientists used CRISPR/Cas9 to target the PAM sites and eliminate the promoter region, transcription start site, and CAG repeats. Conclusions and results showed that performing these actions permanently inactivated the CAG repeats in the mutant Huntingtin gene, thus getting rid of the disease. Something unique about this approach is that it can be personalized for every patient which is important since one technique might work on someone's body but be rejected by another's. Therefore, this approach might be more favorable than the previous one (Shin, et al., 2016).



Figure 4: This figure shows the general steps that a group of scientists took to research potential Huntington cures. After taking out hiPSCs from patients, as shown on the left, they used CRISPR/Cas9 to correct the mutation. The right part of the figure shows the neural cells when they are affected with Huntington's disease compared to healthy neural cells. (Reproduced from Xu, et al., 2017)

4. Conclusion

Huntington's is a neurodegenerative disease that

mainly affects neurons. This disease is caused by the repeat of the nucleotide letters CAG in the huntingtin gene, causing the huntingtin protein to be abnormally large which eventually damages the cell. As of now there is no cure for the disease, but there are treatments available to ease symptoms of patients that are affected by it. In the search of a cure, scientists are researching gene editing methods to try to silence the mutation. The CRISPR/Cas9 system has been tested on cells and shows that it is possible to alter the gene and potentially cure the disease by using stem cells and targeting patient specific PAM sites. Clinical trials have not started yet since it is much more difficult to use CRISPR in patients due to the hardship of editing every cell in the body. However, there is hope that in the future this treatment could be used to help Huntington patients.



Figure 5: This figure shows the various PAM sites that were used in this study to narrow down the specific target sites. (Reproduced from Shin, et al., 2016)

Disclosure

S.S originated the research, did the literature searches, and wrote the paper with figures. M.K supervised the research and writing as well as project direction.

References

Finkbeiner, S. (2011). Huntington's disease. *Cold Spring Harbor Perspectives in Biology*, 3(6), doi: 10.1101/cshperspect.a007476

"Full Stack Genome Engineering." (n.d.). Synthego, www.synthego.com/guide/how-to-use-crispr/sgrna.

"Huntington's Disease Overview, Incidence and Prevalence of HD," (n.d.). Center for Neurological Treatment and Research.

www.neurocntr.com/huntingtons-disease.php

"Lab Tests Show Risks of Using CRISPR Gene Editing on Embryos." STAT, 29 Oct. 2020, www.statnews.com/2020/10/29/lab-tests-show-risksof-using-crispr-gene-editing-on-embryos/.

Rodríguez-Rodríguez, Diana Raquel et al. (2019). Genome editing: A perspective on the application of CRISPR/Cas9 to study human diseases (Review). International journal of molecular medicine. 43(4). 1559-1574. doi:10.3892/ijmm.2019.4112

Roos R. A. (2010). Huntington's disease: a clinical review. *Orphanet J Rare Dis.* 5(40). doi:10.1186/1750-1172-5-40

Schmidt, M.J., Gupta, A., Bednarski, C. et al. (2021). Improved CRISPR genome editing using small highly active and specific engineered RNA-guided nucleases. *Nat Commun* 12, 4219. doi.org/10.1038/s41467-021-24454-5

Shin W, J., Kim, K. Chao J, M. et al. (2016). Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9, *Human Molecular Genetics*, 25(20), 4566-4576, doi.org/10.1093/hmg/ddw286

UC Davis Health, Department of Neurology. "The Huntington Gene." UC Davis Health, health.ucdavis.edu/huntingtons/genetic-change.html.

"What is Huntington's Disease?" (n.d.). Huntington's Together.

www.huntingtonsdisease.com/understanding-huntingt ons-disease.html?c=hta-16b2ec905d3&gclid=Cj0KC QjwyYKUBhDJARIsAMj9lkGz2EGohTc6Fi_EPhH F2NWV1myKN9WtxDjR4_PuD_SDw59cI4e6xFka AhAQEALw wcB&gclsrc=aw.ds

Xu, X., Tay, Y., Sim, B. et al. (2017). Reversal of Phenotypic Abnormalities by CRISPR/Cas9-Mediated Gene Correction in Huntington Disease Patient-Derived Induced Pluripotent Stem Cells, *Stem Cell Reports*, 8 (3), 619-633, doi.org/10.1016/j.stemcr.2017.01.022.