Journal of Research High School

Vol. 2026 (6) 457 – 465
ISSN 2688-3651 [online]

The Use of 1-Dimensional Convolutional Neural Networks on Detecting Alzheimer's through Handwriting Data

SaiPranav Chamarthy[1] *

[1]Oak Park High School, Oak Park, CA, USA
*Corresponding Author: saipranavchamarthy@gmail.com

Advisor: Linda Banh, lbanh@alumni.stanford.edu

Received June 29, 2025; Revised January 22, 2026; Accepted March 2, 2026

Abstract

Alzheimer's disease (AD) is a progressive neurodegenerative disorder affecting millions of individuals worldwide, and early diagnosis remains challenging due to the high cost and complexity of current clinical methods. Recent research indicates that cognitive and motor impairments associated with Alzheimer's disease can manifest as measurable changes in handwriting, making handwriting analysis a potential low-cost, non-invasive diagnostic aid. The aim of this study is to evaluate the effectiveness of supervised machine learning models for classifying Alzheimer's disease using handwriting data from the DARWIN dataset. Multiple classical and deep learning models were trained and compared, including logistic regression, decision tree, random forest, support vector classification, k-nearest neighbors, ridge classifier, a multi-layer perceptron (MLP), and a one-dimensional convolutional neural network (1D-CNN), using numerical features extracted from digitally recorded handwriting tasks. Model performance was assessed with accuracy, precision, recall, and F1 score. Among all evaluated approaches, the 1D-CNN achieved the strongest performance, reaching a testing accuracy of 0.9714. These results may suggest that convolutional neural networks are particularly effective at capturing important relationships between variables in tabular handwriting data and support the potential of handwriting-based deep learning models as a supplementary, accessible tool for Alzheimer's disease detection.

1. Introduction

1.1 Background

Alzheimer's disease (AD) is a highly prevalent neurodegenerative disorder, affecting nearly 6 million individuals aged 65 and older in the United States and manifesting with progressive cognitive and functional decline, including memory loss and impairments in daily activities (Alzheimer's Disease Fact Sheet, 2023). AD pathology disrupts multiple brain regions involved in motor, linguistic, and visuospatial processing, leading not only to memory issues but also to detectable changes in fine motor control and written language. Handwriting reflects an interaction between planning, visual processing, and fine motor execution which are disrupted early in AD progression due to degeneration. AD patients exhibit greater variability in motor execution, altered writing velocity and timing, increased pauses, and degraded fluency or legibility compared to healthy older adults, reflecting impairments across motor, linguistic, and visuospatial domains. (Fernandes et al., 2023). These handwriting alterations have been shown to differentiate AD from mild cognitive impairment (MCI) and normal aging, supporting their potential as early behavioral markers of disease progression (Schröter et al., 2003).

Advances in digital capture technologies have enabled precise recording of pen position, pressure, and temporal dynamics during handwriting tasks, yielding rich datasets that reveal subtle kinematic differences between individuals with AD and healthy controls (Qi et al., 2023). While traditional diagnostic approaches—such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and cerebrospinal fluid analysis—remain the clinical gold

Vol. 2026 (6) 457 – 465
ISSN 2688-3651 [online]

standard, they are costly, time-intensive, and dependent on specialized equipment and expertise, often delaying diagnosis by months to a year (Bružaitė, 2023). In contrast, handwriting assessment offers a rapid, non-invasive, and potentially low-cost additional tool that could augment early detection and monitoring of AD.

This paper evaluates the performance of multiple supervised learning models—including 1D-CNN, multilayer perceptron (MLP), random forest, logistic regression, decision tree, support vector classification (SVC), k-nearest neighbors (KNN), and ridge classifier—in classifying Alzheimer's disease status using digital handwriting features. We aim to determine the feasibility and accuracy of machine-learning based handwriting diagnosis in potential AD screening through analysis of these features.

1.2 Literature Review

Classifying AD through unconventional means continues to be a promising problem for researchers, with many studies discussing the use of handwriting analysis to separate patients with AD from healthy control subjects. Garre-Olmo et al. have used kinematic and pressure-related characteristics of handwriting from participants taking part in a series of handwriting tasks to differentiate between AD patients, MCI patients, and healthy control subjects. The patients were classified using the Kruskal-Wallis test on data from a single handwriting task at a time. Depending on the exact task the methodology used resulted in a specificity ranging from 64.2% to 100.0% and a sensitivity ranging from 71.4% to 100% (Garre-Olmo et al., 2017). This study demonstrates the proof of concept that an analysis of handwriting can successfully classify individuals with AD against healthy subjects. The study was limited by a small sample size and possibly the statistical algorithm used to analyze the handwriting.

Dao et al. have worked with data acquired by digital pens, treated it as one-dimensional time-series data, and employed a 1D-CNN in classifying it (Dao et al., 2022). They augmented data from real patients creating loop patterns with synthetic data generated by a generative adversarial network (GAN) and attained 87.04% accuracy, 85.19% sensitivity and 88.89% specificity. The authors claim that their model outperforms earlier state-of-the-art methods. Specifically, they outperformed Kahindo et al., whose model used K-medoids and a Bayesian method on data regarding the velocity of the pen (Kahindo et al., 2018; Dao et al., 2022). Similarly, Schröter et al. were also able to classify Alzheimer's patients from controls and those with MCI with statistical analysis of handwriting data collected from participants (Schröter et al., 2003). With a K-medoids and a naive Bayesian approach, the authors achieved classification accuracy of 77%, specificity of 76%, and sensitivity of 79.4%; the results were promising, although more data or other tests may yield higher outcomes. These findings suggest that implementing machine learning techniques instead of classical statistical algorithms and using a more established dataset could improve the accuracy of classification.

Recent work has continued to build on handwriting-based approaches to Alzheimer's disease classification by addressing limitations in earlier statistical and convolutional models. Leveraging the DARWIN (Diagnosis AlzheimeR WIth haNdwriting) dataset, which consists of 25 distinct handwriting tasks, researchers have proposed a self-attention–based framework to capture both motor and cognitive impairments associated with AD. Unlike prior studies that analyzed individual tasks in isolation or relied on classical statistical tests, this approach integrates information across multiple handwriting tasks and models long-range temporal dependencies within the handwriting signals. The self-attention model achieved an accuracy of 94.3% and an F1-score of 94.5%, outperforming traditional machine learning methods as well as deep learning approaches such as convolutional neural networks by approximately 4%. In addition, the model demonstrated balanced performance with a sensitivity of 94.5% and a specificity of 94.1%, indicating robust discrimination between Alzheimer's patients and healthy controls. These findings further support the viability of handwriting analysis as an efficient and scalable alternative to MRI-based diagnostic techniques for early Alzheimer's disease detection.

2. Materials and Methods

2.1 Dataset

The DARWIN dataset contains 174 subjects' data, 89 AD patients, and 85 healthy control subjects It comprises

Vol. 2026 (6) 457 – 465
ISSN 2688-3651 [online]

of data from 25 tasks typically related to Alzheimer's diagnosis, where data points such as mean acceleration, grand mean tremor (GMT), time on paper, average pressure, and other handwriting features are recorded by digital pen and tablet for each subject. A Wacom Bamboo tablet and pen were used to record pen movements and timing as participants performed these tasks, capturing both in-air and on-paper behavior as well as timestamps and pressure information. Tasks used included those where the subject copied given letters, words, or phrases; those where they wrote down spoken phrases; those where they drew complex shapes or lines to assess motor control; and tasks where they wrote items from memory. From this raw data, 18 distinct numerical features were extracted for each task, resulting in a total of 450 features per participant (25 tasks × 18 features) that represent characteristics such as time, speed, acceleration, pressure, and pen displacement (Cilia et al., 2023). Each data point is categorized as categorical, continuous, or an integer, and 25 records for each feature exist per individual task. For any tasks that required numerical data on categorical values, one-hot encoding was used where each possible value for a categorical variable is represented as a one in one coordinate of a vector and a zero in all other coordinates.

There were no missing values in the dataset, so no data imputation was done. Some important variables in the dataset are air time (time spent in the air during the execution of tasks), disparity index (health compared to the control group), and GMT (level of tremor in writing). Other points include mean acceleration, pressure, jerk, and time in the air. These features are important because variations in them may indicate changes in the participant's condition. To explore patterns and look at boundaries between the classes, principal component analysis (PCA) was performed using the scikit-learn library. PCA works by computing the eigenvectors (principal components) of the covariance matrix, which represents the covariance between each pair of variables in the dataset. The direction of the principal component with the largest eigenvalue corresponds to the direction containing the greatest variance in the data, with following components capturing less variance and being orthogonal to the others. There were two principal components containing 95% of the variance within the data. After this, outliers were removed to allow for a clearer visualization. To find outliers nine points with the highest distance from the mean were removed, followed by four points removed for having extreme y-coordinates on the PCA plot. PCA was chosen for its simplicity and effectiveness in reducing dimensionality for visualization purposes. However, the PCA-transformed data was not used for training the models. Since the original dataset was relatively small, it was possible to train models directly. PCA was therefore used solely to visualize class boundaries and structure in the data. As seen in Figure 1, the two clusters overlap hinting that the boundaries between the two classes are not simple, suggesting that a deep learning approach would perform better. The overlap indicates that no small subset of features cleanly separates AD patients from healthy controls.

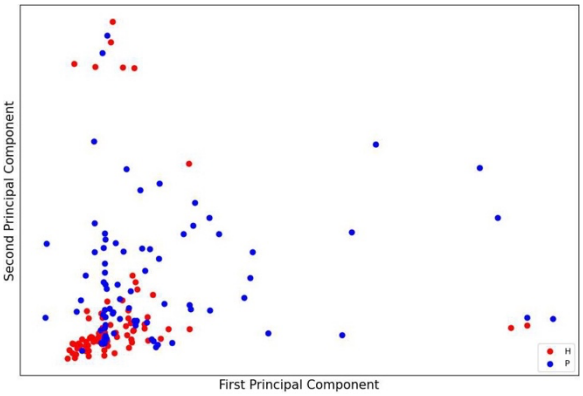

Figure 1. 2-component PCA analysis, containing more than 95% of the variance, of the dataset where P labels Alzheimer's patients and H labels healthy control subjects. 13 Outliers are removed to make viewing easier.

To quantify how unclear the boundaries between the healthy control datapoints and the AD patients datapoints are the average silhouette score for the dataset was calculated. Silhouette score is a measure of how well separated two clusters are by comparing the distances between a point and its own cluster and the distances between the same point and the nearest possible other cluster.

For any point in the dataset (the features or values for each task are the coordinates) x_i its silhouette score is given by

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (1)$$

In this expression cohesion or a_i is the average distance from x_i to any point in the set of all points in the same class as x_i excluding itself. A smaller value of a_i means that the point lies close to other points from its own group. Separation or b_i is the average distance from x_i to any point in the set of all points in the opposite class as x_i. A larger

Vol. 2026 (6) 457 – 465
ISSN 2688-3651 [online]

value of b_i means that the point is far from the other group (1). A high silhouette score occurs when a_i is much smaller than b_i which indicates very clear boundaries or high clustering as the distance from the point to its class is low compared to the distance from the point to the opposite class. As we can see in Figure 2 that is not the case as no data-point has a silhouette score higher than 0.3 indicating unclear boundaries and low clustering. As boundaries between the classes are unclear, it is harder to classify those points into their respective classes with a simpler model relying on a linear relationship or boundary between the two classes. A model which does not rely simple relationships in the data and any simple boundaries in the data will therefore perform better.

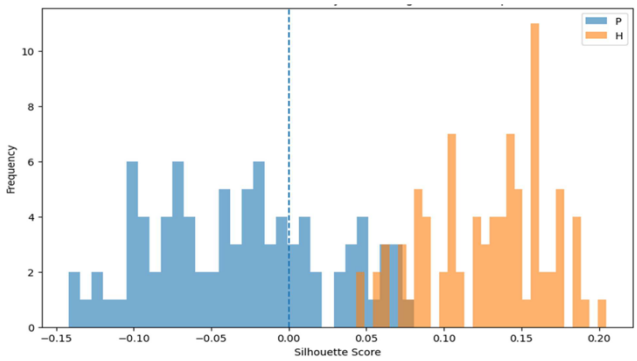

Figure 2. Histogram of Silhouette Scores with classes indicated by color.

2.2 Models

SVC, logistic regression, ridge classifier, KNN, decision tree, random forest, MLP, and 1D-CNN models were evaluated. SVC, logistic regression, ridge classifier, KNN, and decision tree models are simpler models and will most likely have a disadvantage with the complex data set. Logistic regression and ridge classifier use linear decision boundaries in the feature space to classify the data by a specific variable denoted as class (in this case AD patients vs healthy controls). It is expected that both of these models would be weak at handling datasets that have non-linear or weaker boundaries between the classes.

The SVC approach maps the dataset onto a higher dimension (a kernel function is used to accomplish this) and finds a hyperplane that separates the dataset by the class. A radial basis function kernel (RBF) was used because of the non-linear nature of the dataset. Due to the complexity of the dataset, it is expected that this model would still be ineffective.

KNN and decision tree models are predicted to perform worse because of their sensitivity to noise in data and inability to capture complexity. The KNN approach classifies data based on proximity to neighbors in the training data. As seen in Figure 2, the distance between the points is not very useful for splitting the data into clusters, as evidenced by the low silhouette scores for all points, making KNN a bad fit. A decision tree model works by recursively splitting data based on most informative features. These splits can be viewed as a tree graph where each node asks a question and each branch is a specific answer to the question. The final nodes at the bottom of the decision tree determine the final classification. The performance of a decision tree might be inconsistent when the classes overlap spatially such as our example.

Random forest models are predicted to perform better. They consist of multiple decision trees. Each decision tree has an individual outcome. The total outcome is based on the majority of the decision trees' individual outcomes. This approach accommodates the complexity of the data set.

Neural network models are expected to perform better because they can capture complexity and consider many features in tandem with each other. They are inspired by the human brain and use connected layers of neurons to learn potentially complex patterns from data without being programmed in a specific way. In this study a 1D-CNN, a type of neural network that employs convolution layers (i.e. learned filters that can be used to extract specific features) is used. The 1D-CNN is anticipated to perform the best, since it can use convolution layers on the data to find relationships more explicitly. The tabular nature of the data set is however expected to be a challenge as it does not have inherent context like a smooth image or a continuous time-series would.

2.3 Methodology

Our first step was to split the data randomly into a training and testing set in a four to one ratio. Logistic regression,

Vol. 2026 (6) 457 – 465
ISSN 2688-3651 [online]

decision tree, random forest, SVC, KNN, and ridge regression were trained on the training set. For models that required hyperparameter optimization, a fraction of the training set was set apart for validation creating a train validation test ratio of seven to one to two. Hyperparameter optimization for the KNN, 1D-CNN, random forest was done through a grid search on all relevant parameters on the validation set performance. The MLP was implemented with cross-validation, which split the training data into two sets with a much smaller one (one-fifth the size of a training set)used for validation during the training process to reduce overfitting. To increase the efficiency of the grid search on hyperparameters for the MLP, while minimizing computation power, we attempted a Bayesian search. The Bayesian search checks the samples performance of a few hyperparameter values and then constructs a function that approximates the performance of hyperparameter values creating a gaussian process which assumes that the performance of a set of hyperparameters is normally distributed. It then uses a function to find the next best set of guesses, records the results, and then uses the new samples to repeat the process. This can accelerate the hyperparameter training, achieving better results in shorter periods of time.

Refer to Figure 3 for the model architecture used for the 1D-CNN. The 1D-CNN was implemented with a convolutional layer having 64 kernels of size 3x1 (filters used in the convolution) and an activation function of rectified linear unit (ReLu). ReLu acts on the final output of a neuron or an entire layer. It returns 0 if the input is a negative number and leaves the input unchanged otherwise. ReLu is used to introduce non-linearity and is a faster alternative to other activation functions. The output is subsequently processed by a max pooling and a flattening layer which downscales the data and turns it into a one-dimensional vector.

The output of the previous layer is then processed by a 100-neuron dense layer with the ReLu activation function. Each neuron in a dense layer is connected to the outputs of the previous layer fully, and takes a linear combination of those outputs, adds a bias value, and applies the activation function. The bias value and the coefficients or weights by which it multiples each input are learned and can be fine-tuned to achieve better classification.

The final layer is a 2-neuron dense layer representing a value for each of the two classes which is then processed by the softmax activation function, converting the raw values into probabilities for an input belonging to a specific class. The model will set checkpoints at the best version each round of training or epoch and default to returning the previous version if the new one performs worse on the validation portion of the training data. The model uses the adam optimizer to optimize weights, uses categorical cross-entropy to calculate the loss in order to assess performance during training, and was trained for 50 epochs.

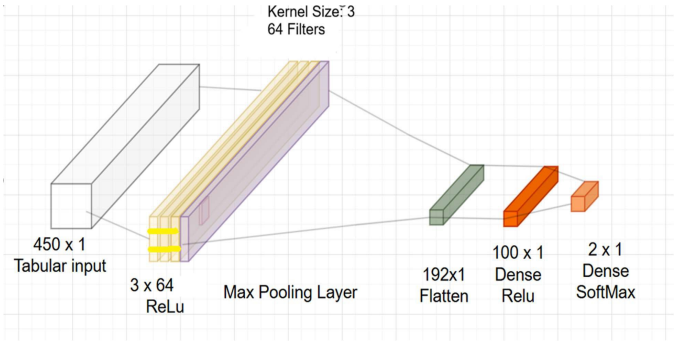

Figure 3. A diagram of the architecture of the model, note that the purple layer is a max pooling layer.

3. Results

3.1 Models

The models were evaluated based on their accuracy, precision, recall, and F1 score. The accuracy metric is the percentage of data points that were classified correctly or the true positives and true negatives over the total number of data points (2). The precision metric is the percentage of patients classified with Alzheimer's disease that were classified correctly (3). The recall metric is the percentage of patients with Alzheimer's disease patients that are classified correctly (4). The F1 Score is the harmonic mean of the precision and recall scores (5). The equations for each metric are provided below. True Positives are denoted as TP, True Negatives are denoted as TN, False Positives are denoted as FP, and False Negatives are denoted as FN.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Vol. 2026 (6) 457 – 465
ISSN 2688-3651 [online]

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1\ Score = \frac{2 * precision * recall}{precision + recall} \quad (5)$$

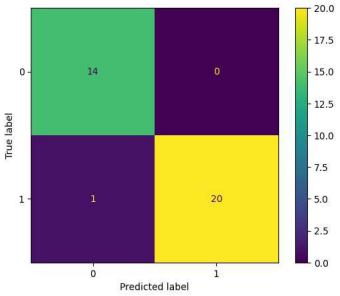

Figure 4. Confusion Matrix for 1D-CNN on the Validation Set.

The validation metrics are shown below in Table 1. The 1D-CNN was the most efficient model, with a 97.14% validation accuracy. The confusion matrices are shown in Figure 4 (1D-CNN's) and Figure 5 (KNN, random forest, ridge classifier, SVC, logistic regression, and decision tree). It required less time to train and achieve these metrics than the MLP which took the most time due to an extensive search for optimal hidden layer sizes.

Table 1. Validation metrics for each model to the number of significant figures provided by calculation capped at 4.

Model	1D-CNN	KNN (3)	Random Forest (11)	Ridge Classifier	SVC(RBF)	Logistic Regression	Decision Tree	MLP (7,27)
Accuracy	0.9714	0.7714	0.9429	0.6286	0.6286	0.8571	0.8286	0.8286
Precision	0.9524	0.7679	0.9333	0.625	0.5667	0.8583	0.8333	0.8302
Recall	1	0.675	0.95	0.6234	0.8030	0.8536	0.8268	0.8167
F1 Score	0.9655	0.7184	0.9416	0.6237	0.6645	0.8553	0.8273	0.8214

The parenthesis next to the name of the KNN is the number of neighbors, for random forest it is the number of Decision Trees, for SVC it is the kernel used, and for the MLP it is the hidden layer sizes.

4. Discussion

As shown in Table 1, logistic regression, SVC, decision tree, KNN, MLP, and ridge classifier all had test accuracies below 90%. The 1D-CNN and random forest model performed better in all the testing metrics. The hyperparameter, number of decision trees for the random forest model was evaluated between 1 and 100 with the training and validation set. The number of decision trees at 11 gave the best validation accuracy and was chosen for further training.

Figure 5 shows a histogram of the testing accuracies of each of the random forest models trained on the full training set for 1000 replicates. The random forest model achieved a maximum testing accuracy of 0.9429, with a considerably lower median testing accuracy of 0.8.

As seen in Figure 1 and Table 1, the simpler models (logistic regression, ridge regression, SVM, KNN, and decision tree) were unable to differentiate between the AD patients and healthy controls effectively. Random forest is expected to do well with tabular data, as tree-based models are known to be adept at classifying such data. Having the input of many decision trees allows the random forest model to classify the less-clear boundaries better than the simpler models. This is consistent with the fact that random forest worked well, although its performance varied a lot. Though

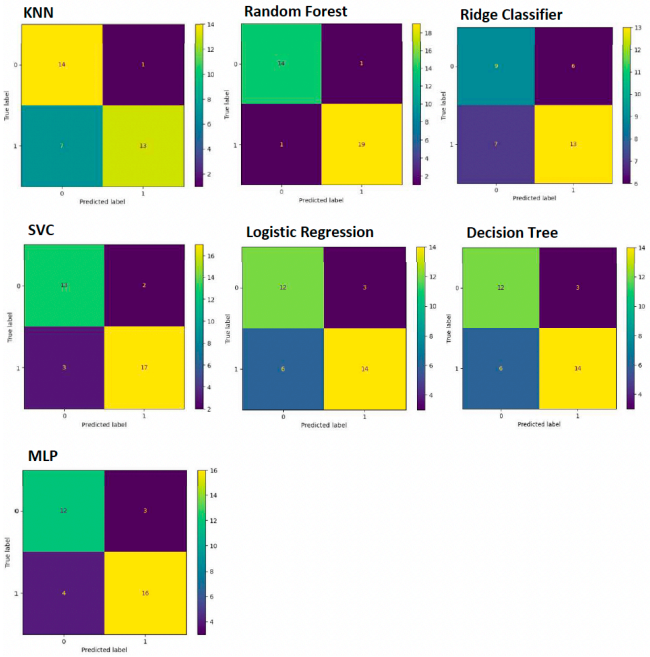

Figure 5. Confusion matrices for the KNN, random forest, ridge classifier, SVC, logistic regression, decision tree, and MLP in that order read from left to right.

Vol. 2026 (6) 457 – 465
ISSN 2688-3651 [online]

neural networks are considered to be less-suited for tabular data the 1D-CNN outperforms the random forest model, however the MLP does not.

The MLP took much longer than any other model attempted in this study to train due to testing many hidden layer sizes in an exhaustive grid search. Although the training accuracy of the model was 0.9677, the testing accuracy was not as high as seen in Table 1. The MLP performed worse than the random forest model on the testing set as seen in Table 1. The MLP overfitted on the training set, focusing on the random noise within the training set instead of the trend. While the MLP was trained using cross validation to avoid overfitting, the training accuracies always outdid the testing accuracies by more than 0.1, indicating otherwise. The MLP got similar training accuracies to the 1D-CNN but still got testing results which were significantly worse. To reduce overfitting in future implementations, several modifications could be applied. These include limiting model complexity by constraining the number of hidden layers (layers not directly interacting with the input or output) and neurons, incorporating techniques such as dropout, and stopping training early based on the model's performance on the validation set. Dropout refers to a technique where neurons are randomly ignored for a portion of time during training, forcing neurons to rely on many others. This creates a more robust network and potentially avoids overfitting.

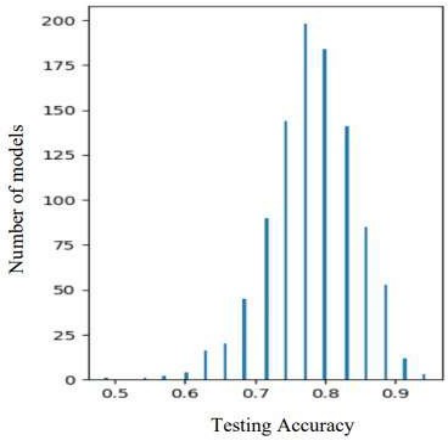

Figure 6: Histogram of the testing accuracies of one thousand random forest models with 11 trees.

The 1D-CNN was able to use the convolution operation to consider multiple features in context with each other in the data. Training the 1D-CNN model took considerably less computational power than MLP, since the 1D-CNN's training did not require explicit hyperparameter tuning. As seen in Table *2* and Figure 7, the training accuracy and F1 score went up as a whole and correlated positively with each other during training. However, most increases in precision had a decrease in recall and vice versa leading to their correlation being considerably lower at 0.2912. To improve recall by catching all the positives, typically a model will start misclassifying by creating false positives lowering precision and vice versa. This trade-off is observed in many models, but ultimately the correlation was positive due to the model improving as it trained. Ultimately, the testing precision never reached one hundred percent while the testing recall did. This indicates that there may be an outlier within the testing set that no models can classify, as that would have an effect on the recall and accuracy which would make the model worse. High recall with an imperfect precision is acceptable for screening tasks where a false negative is more costly than a false positive.

Table 2: Correlation between training accuracy, precision, recall, and F1 Score with four significant figures.

Metric	Accuracy	Precision	Recall	F1-Score
Accuracy	1	0.8571	0.6958	0.9505
Precision	X	1	0.2912	0.7007
Recall	X	X	1	0.8657
F1-Score	X	X	X	1

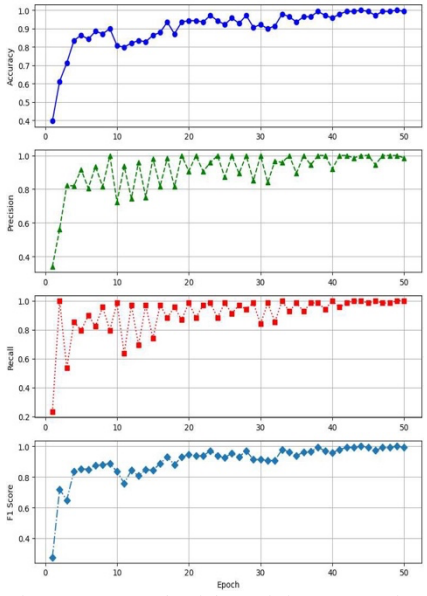

Figure 7. A graph of the training accuracies, precision, recall, and F1 Score of the 1D-CNN over each epoch trained.

Although 1D-CNNs are mainly considered for time-series data, it can use the convolution to also find relationships between the categories in a tabular dataset, leading to better performance. The boundaries between the classes are complex and the true challenge lies in the model's ability to interpret the. Thus, 1D-CNN's ability to extract useful information that is not highly structured, nor strongly following an obvious trend, is what made it perform best.

A major weakness in this study is the fact that the data is relatively

Vol. 2026 (6) 457 – 465
ISSN 2688-3651 [online]

small and cross-sectional, which might limit our ability to generalize the findings. Moreover, since the study used structured handwriting tasks and features, this approach may not work in analyzing free form handwritten text. Also, there is still some uncertainty in how other factors such as the level of educational and handwriting practices might influence handwriting even if there is no cognitive impairment. In the end, the 1D-CNN model, even though performing best, might end up being data-specific owing to the limited amount of data, and this highlights the need of larger and more diverse datasets in the future.

5. Conclusion

Through our research, we were able to test various models' effectiveness in classifying Alzheimer's disease patients using handwriting data. We formatted data to train and test each model, and varied time and resources for each entry. Although the MLP exemplified flaws in a deep learning approach, the 1D-CNN performed best with 0.9705 accuracy for training and 0.9714 on the testing set. This is due to a convolutional neural network's ability to assess more than one data point with convolutions. Though borders between Alzheimer's patients and healthy control subjects were blurry; data with such higher complexity, like ours, couldn't be fully analyzed by models which relied on well-defined and clear boundaries between the classes, such as logistic regression, ridge regression, and SVC. Our future research direction is to gather more data, try other model architectures, and perform data augmentation with a GAN to improve pre-processing and model accuracies.

Acknowledgment

Firstly, I would like to thank my parents (Sai Prasanth Chamarthy and Sabitha Katpally) for providing me with the tools to achieve this. Secondly, I would like to thank my mentors Linda Banh and Raghav Ganesh from Inspirit AI for helping me with the implementation details and guiding me throughout the entire process. I would like to thank the Olga Radko Math Circle for providing me with the mathematical foundations and critical thinking skills necessary for this. Finally, I would also like to thank the Oak Park High School research club for providing a helpful environment, and Isaiah Sohn for providing advice on the revisions.

References

Alzheimer's Disease fact Sheet. (2023, April 5). National Institute on Aging. Retrieved May 28, 2025, from https://www.nia.nih.gov/health/alzheimers-and-dementia/alzheimers-disease-fact-sheet

Bružaitė, M. (2022). Alzheimer's disease: symptoms, risk factors, diagnosis, treatment and future treatment options. *Medical Sciences*, 10(2). https://doi.org/10.53453/ms.2022.05.4

Cilia, N. D., et al. (2022). Diagnosing Alzheimer's disease from on-line handwriting: A novel dataset and performance benchmarking. *Engineering Applications of Artificial Intelligence*, 111, 104822. https://doi.org/10.1177/13872877241283920

Dao, Q., El-Yacoubi, M. A., & Rigaud, A. (2022). Detection of Alzheimer disease on online handwriting using1D convolutional neural network. *IEEE Access*, 11, 2148–2155. https://doi.org/10.1109/access.2022.3232396

El-Assy, A. M., et al. (2024). A novel CNN architecture for accurate early detection and classification of Alzheimer's disease using MRI data. *Scientific Reports*, 14(1). https://doi.org/10.1038/s41598-024-53733-6

Fernandes, C. P., et al. (2023). Handwriting changes in Alzheimer's disease: A systematic review. *Journal of Alzheimer's Disease*, 96(1), 1–11. https://doi.org/10.3233/JAD-230438

Garre-Olmo, J., et al. (2017). Kinematic and Pressure Features of Handwriting and Drawing: Preliminary Results Between Patients with Mild Cognitive Impairment, Alzheimer Disease and Healthy Controls. *Current Alzheimer Research*, 14(9). https://doi.org/10.2174/1567205014666170309120708

Vol. 2026 (6) 457 – 465
ISSN 2688-3651 [online]

Kang, L., et al. (2024). Early Alzheimer's disease diagnosis via handwriting with self attention mechanisms. *Journal of Alzheimer's Disease*, 102(1),173–180. https://doi.org/10.1177/13872877241283920

Matias, Ana. (2023, December 2 – 2024, July 30). Handwriting as an objective tool to support the identification of people with Alzheimer's disease: Suitability of an assessment protocol. Identifier NCT06483438. https://clinicaltrials.gov/ct2/show/NCT06483438

Palmis, S., et al. (2017). Motor control of handwriting in the developing brain: A review. *Cognitive Neuropsychology*, 34(3–4), 187–204. https://doi.org/10.1080/02643294.2017.1367654

Schröter, A., et al. (2003). Kinematic Analysis of Handwriting Movements in Patients with Alzheimer's Disease, Mild Cognitive Impairment, Depression and Healthy Subjects. *Dementia and Geriatric Cognitive Disorders*, 15(3), 132–142. https://doi.org/10.1159/000068484

Qi, Hengnian, et al. (2023) "A Study of Auxiliary Screening for Alzheimer's Disease Based on Handwriting Characteristics." *Frontiers in Aging Neuroscience*, 15, https://doi.org/10.3389/fnagi.2023.1117250

Yiannopoulou KG, & Papageorgiou SG. (2020). Current and Future Treatments in Alzheimer Disease: An Update. *Journal of Central Nervous System Disease*, 12. doi:10.1177/1179573520907397