A Novel Approach to Early Detection of Dysgraphia Using Deep Learning Neural Networks

Nathan Guan

Abstract

Dysgraphia is a neurological disorder that impairs writing and fine motor skills, making diagnosis challenging and often leading to misdiagnosis [1]. It frequently coexists with other neurodevelopmental disorders, emphasizing the need for accurate early detection. This paper reviews current AI-based diagnostic systems for dysgraphia and proposes a deep-learning neural network framework that achieves over 80% accuracy and 100% sensitivity, significantly improving upon previous methods [2].

Index Terms

Dysgraphia, deep learning, neural networks, early detection, machine learning

I. Introduction

Handwriting is a critical skill in a child's education, often constituting up to 60% of classroom time. Research ties handwriting proficiency to academic achievement in young children, particularly in composition and literacy skills [3]. While most children develop these skills naturally, those with learning disorders face significant challenges, affecting their ability to acquire necessary skills. These difficulties, often labeled as "special education needs," reflect differences in brain function rather than intelligence, potentially undermining a child's self-esteem and confidence [4].

Dysgraphia is one such learning disorder, interfering with nearly all aspects of writing, including spelling, legibility, and expression [5], [6]. It frequently coexists with other neurodevelopmental disorders, such as dyslexia, ADHD, ASD, and cerebral palsy, with co-morbidities over 30% [6]. Studies show that approximately 60% of individuals with ADHD or ASD also have dysgraphia [3], [5]. This percentage could be even higher, as over 90% of children with writing disorders are commonly dismissed and undiagnosed due to the belief that it simply is poor handwriting [1]. Table I offers a concise breakdown of the different categories and abilities that dysgraphia affects.

TABLE I: Types of Dysgraphia [4]

Type	**Characteristics**
Dyslexic	Spontaneously written work: illegible. Copied work: good. Spelling: bad.
Motor	Slow muscle tone, poor posture, decreased fine motor skills, decreased visual motor skills. Written often inconsistent slant or size due to unusual grip. Breakdown between the language center of the brain and the motor map signals. Spontaneous work and copy work: poor or illegible. Difficult with motor memory.
Spatial	Drawing and writing task: difficult. Understanding space: weakness. Spontaneous and copy work: illegible. Sensory processing or body awareness deficit.
Phonological	Writing and spelling disturbances. Trouble blending phonemes to create words. Trouble spelling unfamiliar words and non-words.
Lexical	Use sound to letter patterns with misspelling or irregular words.

In our digital age, dysgraphia remains a pressing issue, impacting more than just academic performance by affecting self-esteem, self-expression, and communication skills. Early diagnosis is crucial for intervention, allowing children to overcome obstacles and develop compensatory techniques [7]. Additionally, early identification can enable teachers to tailor instruction and may lead to the early detection of other conditions.

Nathan Guan is with Lake Washington High School, Kirkland, WA, (e-mail: nguan@lwsd.org).

II. Previous Research Work

In this section, we will first explore psychological techniques for diagnosing dysgraphia and for manual assessment of handwriting [8]. Subsequently, we will examine noteworthy advancements in automated analysis through machine learning methods, as well as diagnosis systems based on non-machine learning approaches, all aimed at detecting dysgraphia.

A. Psychological Approach

The conventional approach to dysgraphia diagnosis involves a multifaceted method comprising academic skills assessment and cognitive analysis. Dysgraphia is typically characterized by noticeable challenges with handwriting, letter formation, spatial perception, fine motor skills, spelling, and composing written work. While students with dysgraphia often excel in verbal expression, they struggle to translate their thoughts onto paper. Occupational therapists or psychologists assess various skills, including constructional ability, executive function, writing and spelling, phonological awareness, working memory, and intelligence. These psychological evaluations provide a comprehensive framework for assessing the cognitive and motor skills necessary for effective handwriting. Various standardized tests, such as the Beery Visual Motor Test, Rey Complex Figure Test, and Comprehensive Test of Phonological Processing, are employed to assess these aspects. Additionally, the Concise Evaluation Scale for Children's Handwriting (BHK) quantifies writing speed and quality [4], [9], also serving as a valuable tool in researching, developing, and evaluating machine learning-based dysgraphia diagnosis systems.

B. Computational Approach

Gargot et al. [10] conducted a study involving 280 children recruited from regular schools and specialized clinics to participate in the BHK assessment using digital tablets. Of these, 62 children were identified with dysgraphia, forming the experimental group, while 218 children with no dysgraphia comprised the control group. The researchers extracted twelve digital features of handwriting, including static, kinematic, pressure, and tilt aspects, and developed linear models to illustrate the evolution of handwriting skills. They also employed K-means clustering, which categorized the children into three clusters and revealed distinct developmental patterns in handwriting features. Children in Cluster 1 (C1) often displayed mild dysgraphia, while Clusters 2 (C2) and 3 (C3) exhibited moderate to severe dysgraphia, with specific struggles in kinematics, pressure, and tilt features. These findings highlight the potential for an automated dysgraphia identification process and suggest possible benefits from implementing the pressure and tilt data when training.

C. Conventional Machine Learning Approach

The general flow of a machine learning-based dysgraphia diagnosis system involves several key stages: data collection (both offline and online), preprocessing, feature extraction, feature selection, application of a machine learning algorithm, and finally, prediction [11]. The conventional approach for dysgraphia diagnosis mirrors the structure of other machine learning applications, emphasizing the sequential process from data acquisition to predictive output.

Drotar and Dobe's research [2] focused on creating a binary classification model to distinguish handwriting samples of typically developing children from those with dysgraphia. Other researchers turned to explore several machine learning algorithms, using Python's sci-kit-learn module and TPOT for optimization [10]. After experimentation, they manually selected Adaptive Boosting (AdaBoost), Random Forest (RF), and Support Vector Machines (SVM) as the most promising classifiers. The classifiers exhibited similar performance, with SVM showing the highest accuracy at 78.8%, followed by RF at 77.6% and AdaBoost at 79.5% accuracy in some cases. SVM also demonstrated the highest sensitivity at 82.4%, while RF had the best specificity at 83.3%. These results suggest that SVM, RF, and AdaBoost are all viable options for predicting dysgraphia, each with strengths in different aspects of classification.

D. 1D Convolutional Neural Network for Time Sequences

While 2D Convolutional Neural Networks (CNNs) are commonly used for image recognition, 1D CNNs offer significant potential for applications beyond Natural Language Processing (NLP). Despite a lack of comprehensive guides on 1D CNNs, Nils [11] provided an insightful resource on their implementation. 1D CNNs are effective for analyzing time-based sensor data, such as gyroscope or accelerometer data, and are particularly useful for processing shorter, fixed-length data segments where spatial placement is less crucial.

In this study, 1D CNN was utilized to diagnose dysgraphia by segmenting time sequences, extracting relevant features, and training the model to identify handwriting affected by dysgraphia. The model demonstrated high accuracy, even with a heterogeneous dataset, through both segment-level and aggregated person-level evaluations.

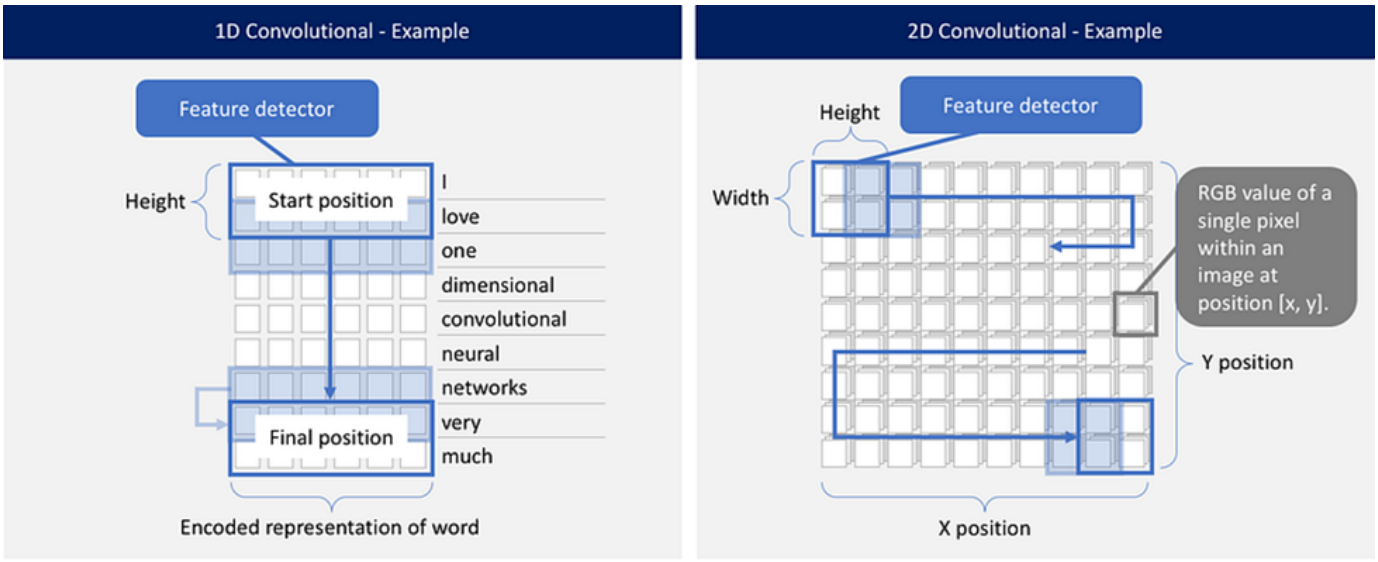

Fig. 1: Comparison Between 1D and 2D Convolutional Neural Network [11].

III. Proposed Methodology

A. Participants and data collection

This study uses raw data from Drotar and Dobe's research [2], pre-processed for neural network model fitting and validation. The dataset consists of 120 schoolchildren, with a fairly even age distribution confirmed by a one-sample Kolmogorov–Smirnov test (significance = 0.055). A t-test revealed no significant age difference between children with and without dysgraphia. Among the participants, 80 were boys and 40 were girls; 57 of the children were diagnosed with dysgraphia. Of the participants, 16 were left-handed, and 104 were right-handed.

The data was collected at the Centre for Special Needs Education under professional supervision, with informed consent obtained from the parents or guardians of each child. The dataset, available in a public repository [8], was recorded using a WACOM Intuous Pro Large Tablet [12]. The children wrote with a pen on paper placed on the tablet, which captured variables including pen movement in x and y directions, pressure, azimuth, and altitude, as well as on-surface and in-air movements.

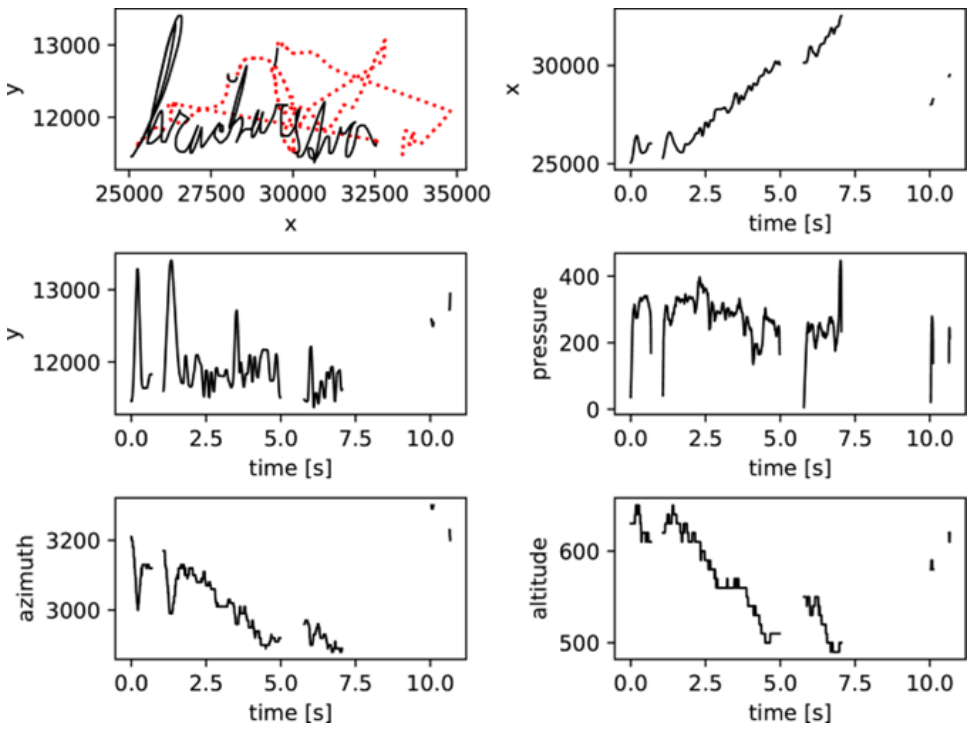

Fig. 2: Handwriting Sample of a Participant with Dysgraphia and All Values Recorded by the Tablet [2].

B. Pre-processing

The raw data is saved as time sequences per participant. Under each participant's session folder, a plain text file is used to save all the time-sampled data records.

TABLE II: Data After Preprocessing of Participant ID 6

ID	Dysg.	X	Y	Timestamp	Surf.	Azim.	Alti.	Pres.
6	1	11888	32154	727745086	1	1370	470	7
6	1	11888	32156	727745093	1	1370	470	18
6	1	11888	32156	727745101	1	1370	470	31
6	1	11888	32157	727745108	1	1370	470	42
6	1	11888	32159	727745116	1	1370	470	49

Table II provides an excerpt from the consolidated post-processing data file that includes all participants. The first column lists the participant IDs, with participant ID 6 shown as an example. The second column indicates the presence or absence of dysgraphia, where '1' represents a diagnosis of dysgraphia and '0' indicates no dysgraphia. The subsequent six columns retain the original raw data, showing the X and Y coordinates of the pen's position, a timestamp in 100 microseconds, a Boolean value indicating whether the pen is in the air (0) or on the drawing tablet (1), and the pen's azimuth, altitude, and pressure. In total, 2,176,420 data records were collected from all participants.

C. Data visualization

Among the 2,176,420 data records, the distribution of whether the participant has dysgraphia or not is shown in Figure 3 on the left. The distribution of data records from each participant, detailed in descending order, is shown in Figure 4 on the right.

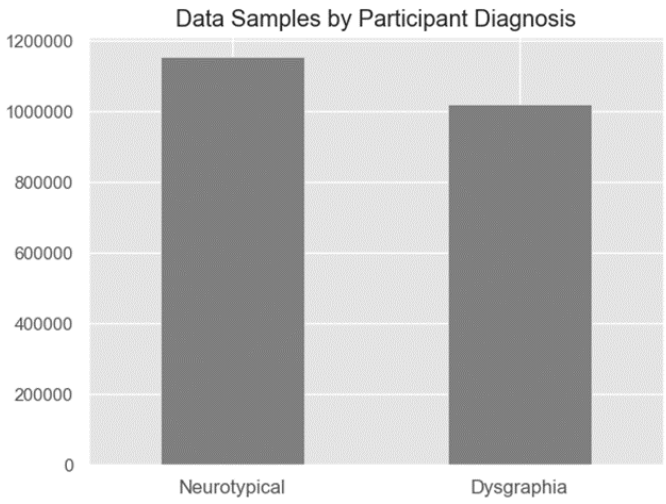

Fig. 3: Participant Distribution - Dysgraphia and Neurotypical

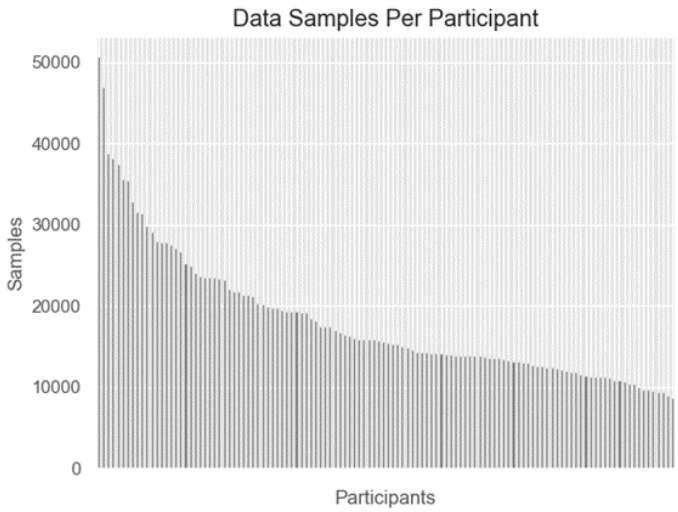

Fig. 4: Data Records Distribution by Participant

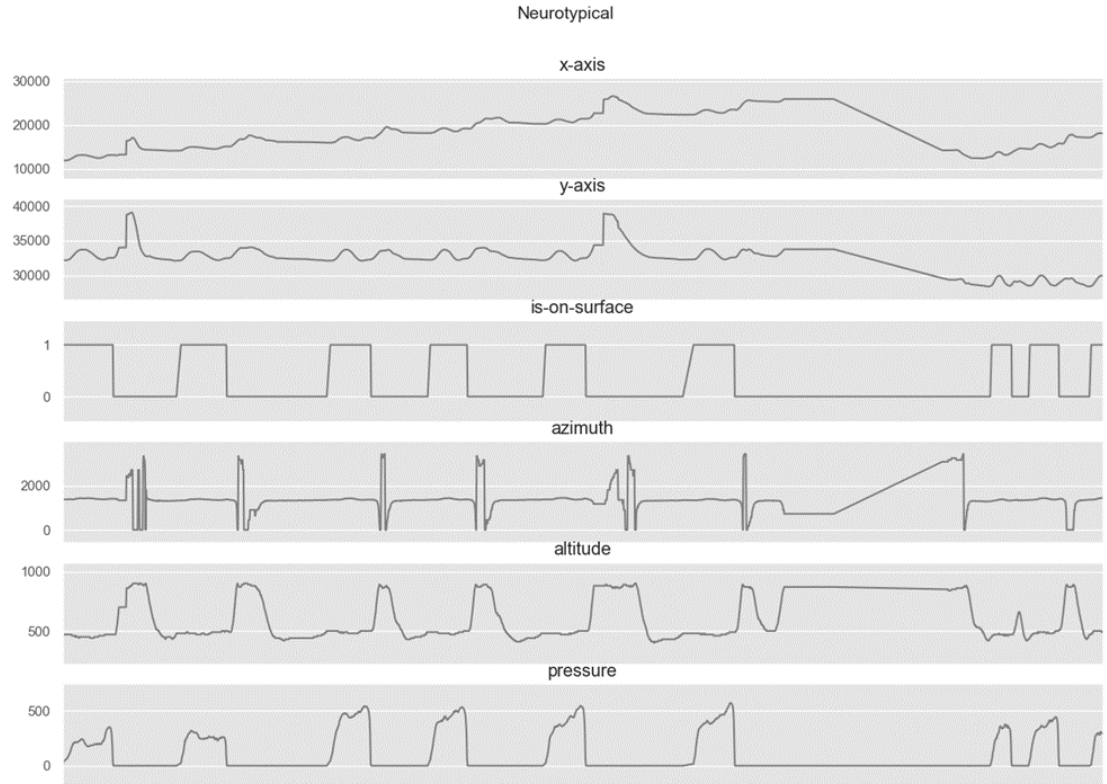

Fig. 5: Data Plot of Neurotypical Participant

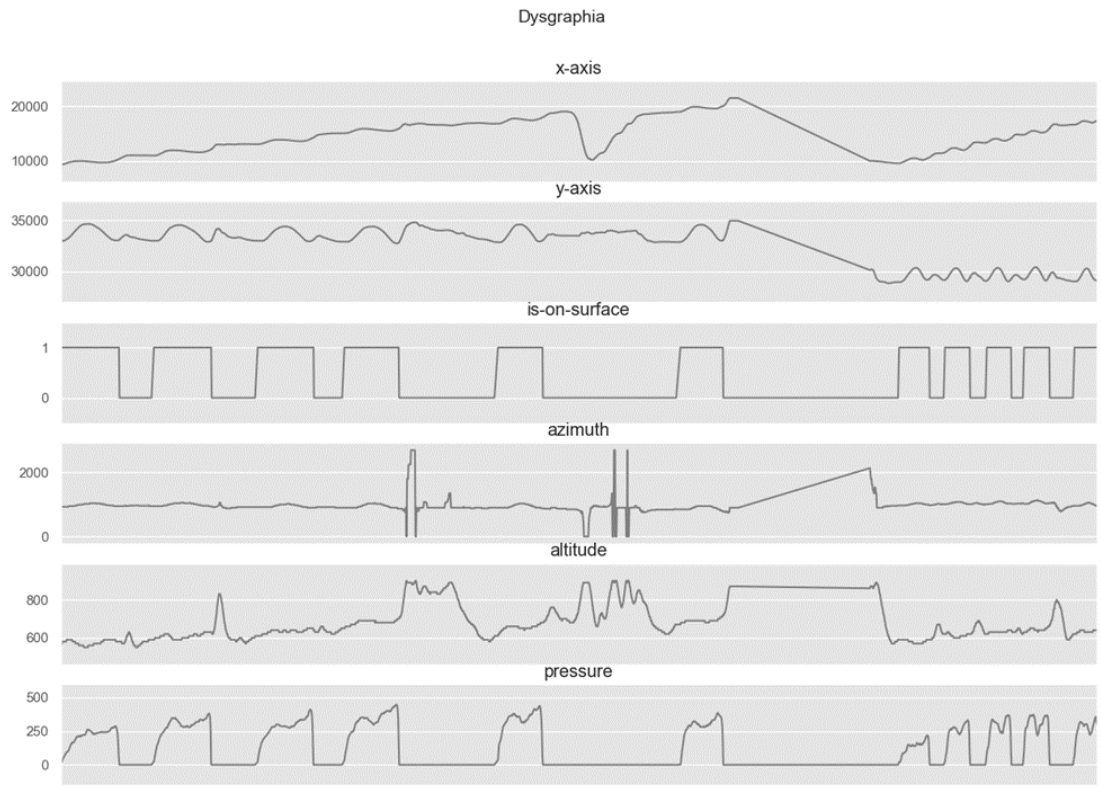

Fig. 6: Data Plot of Participant with Dysgraphia

IV. Primary Data Analysis

A. Data Normalization

The first step in the experimental study is to prepare the data for the model so it's able to reliably interpret and learn each input. This is important because the data might have different ranges and units which could lead to unbalanced analysis. In this model, Z-Score Standardization was used to transform the data to have a mean of 0 and a standard deviation of 1.

B. Training vs. Test Data Separation

In this study, the data consists of time sequences; therefore, selecting individual data records to divide into training and test datasets may yield inaccurate results. A more effective approach involves using each participant's ID to segment the dataset. This is because each data record receives a prediction for dysgraphia diagnosis, which might not accurately represent the participant's overall diagnosis, as it only reflects a small snippet of the writing sample. For the dataset division, participant IDs were randomly chosen to create the training and testing datasets, while ensuring the proportion of dysgraphia to neurotypical cases remains consistent. Consequently, this split resulted in 1,746,903 records for the training dataset and 429,516 records for the test dataset, establishing a test-to-training ratio of approximately 1:4.

C. Data Segmentation

Given the use of consecutive time sequence data in determining dysgraphia, it is crucial to consider segments as a whole, as discerning dysgraphia from individual mere microseconds of data is impractical. Data segmentation is employed to reorganize the data into manageable segments, effectively addressing this challenge. In this model, the segments were configured with a length of 800 records, and each step spanned 400 records, allowing for a 50% overlap between segments. This overlap ensures continuity,

linking each data record to its preceding and succeeding records. Additionally, each segment is processed to yield a unique binary value (0 or 1), indicating whether the segment displays characteristics typical of dysgraphic handwriting.

TABLE III: The shape of the segmented dataset

Data Set	Number of Segments	Segment Length	Total Features	Percentage of Dataset
Training (x data)	4366	800	4800	80%
Training (y data)	4366	-	1	80%
Testing (x data)	1072	800	4800	20%
Testing (y data)	1072	-	1	20%

As shown below in Table III, the initial data is divided into a training dataset (80%) and test dataset (20%).

D. 1D CNN Model Training and Testing

The 1D CNN model includes a series of convolutional layers with varying kernel sizes, followed by max pooling layers to reduce dimensionality. The model then applies a dropout layer for regularization, followed by a dense layer with a sigmoid activation function for binary classification of dysgraphia.

An epoch represents a full pass through the entire training dataset, while 'batch size' refers to the number of data segments processed simultaneously by the model. To manage high memory demands, the dataset is divided into smaller, more manageable batches. We experimented with different epoch sizes, specifically 50 and 100, and found that while increasing the epoch count improved accuracy during the training and validation phases, it did not significantly enhance accuracy on the test dataset. Consequently, we selected 50 epochs to balance performance and avoid overfitting. Additionally, we incorporated an early stopping mechanism that activates if there is no improvement in accuracy over three consecutive epochs. After testing various batch sizes—32, 64, 128, and 256—we determined that a batch size of 64 produced the best results. Thus, we settled on 50 epochs and a batch size of 64, with early stopping to ensure optimal model performance.

E. Person-Level Evaluation

The Convolutional Neural Network (CNN) model is trained using individual data records. For processing time sequence data, we employ data segments that encapsulate timing information, thereby enabling the model to effectively understand and interpret temporal dynamics. However, the crucial determination of dysgraphia diagnosis is made at the person level, not the data segment level. A single participant has many data segments, some predicted as dysgraphia and others as neurotypical. By aggregating all data segments per participant to evaluate at the person level, we aim to enhance model performance. To this end, a custom callback function was developed to evaluate the model at the person level, diverging from the more typical record-level (or segment-level) evaluation. This is important since multiple data segments correspond to individual subjects, with the objective being to assess the model's efficacy based on person-specific outcomes.

After each training epoch, this callback function calculates predictions for the validation data and then performs aggregation at the person level. The aggregation methodology involves compiling predictions for each individual and determining the final classification based on a set threshold. If the proportion of 'positive' predictions (labeled '1') for an individual surpasses this threshold, the individual is categorized as '1'; otherwise, they are marked as '0'. This approach offers a more comprehensive evaluation than standard segment-level accuracy, as it reflects the model's capability in classifying individuals based on an aggregate interpretation of their data segments. The resulting person-level accuracy is then calculated and reported, offering a deeper understanding of the model's efficacy in accurately classifying entire individuals, rather than solely isolated data segments.

V. Results

Hyperparameter Optimization and Threshold Determination:

- In our pursuit of optimal performance, we fine-tuned the hyperparameters for person-level evaluation, as detailed in Section IV-E.
- A crucial aspect of our model's design was the implementation of an "optimal" threshold to discern dysgraphia in individuals, based on aggregated data segment predictions. Various thresholds, ranging from 0.30 to 0.70, were rigorously tested.
- The effectiveness of these thresholds was assessed using a Receiver Operating Characteristic (ROC) curve and quantified via the Area Under the Curve (AUC) metric, providing insights into the binary classifier's distinction capabilities (Figure 7).

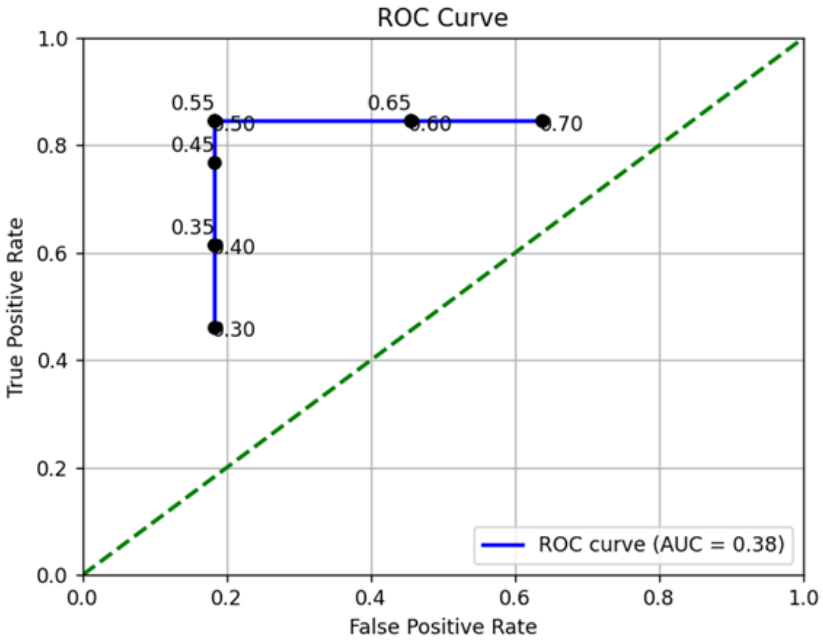

Fig. 7: ROC/AUC Graph with Different Threshold

- Our analysis revealed optimal performance at thresholds of 0.5 or 0.55, striking a balance between a high true positive rate (83%) and a low false positive rate (18%). Consequently, we adopted a simple majority vote threshold of 0.5, which efficiently classified individuals with dysgraphia.

Training Data Performance:

- Our model demonstrated a high level of accuracy in detecting dysgraphia. Specifically, we achieved a segment-level accuracy of 91.84%.
- More impressively, when using the person-level evaluation method, the model reached a perfect accuracy rate of 100%.

Test Dataset Performance:

- On the test dataset, the model's performance, while lower than on the training data, was still notable. The segment-level accuracy stood at 65.30%.
- Employing the same person-level aggregation method on the test dataset yielded an accuracy of 83.33% (Equation 1 below).

Confusion Matrix: The confusion matrix (Table IV) provided a detailed breakdown of the model's performance, distinguishing between actual cases of dysgraphia and neurotypical instances.

Performance Metrics:

Based on this matrix, we further measured the model performance and calculated the following metrics: Accuracy considers both positive and negative cases and measures the overall correctness of the classifier;

TABLE IV: Confusion Matrix

Predicted	Dysgraphia	Neurotypical
Actual Dysgraphia	13	0
Actual Neurotypical	4	7

while Sensitivity focuses on correctly identifying positive cases; and Specificity focuses on correctly identifying negative cases.

- Person Level Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = 83.33\% \tag{1}$$

- Person Level Specificity:

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{7}{7 + 4} = 63.64\% \tag{2}$$

- Person Level Sensitivity:

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{13}{13 + 0} = 100\% \tag{3}$$

Here, true positive (TP) and false positive (FP) represent the number of correctly identified dysgraphic subjects and the number of subjects diagnosed as having dysgraphia but are normally developing. Similarly, true negative (TN) represents the total number of correctly identified normally developing subjects, and false negative (FN) represents subjects with dysgraphia but evaluated as being normally developing.

Significance of Results:

- The model's high sensitivity rate (100%) is particularly significant, indicating effective early detection of dysgraphia. Such precision ensures fewer children with dysgraphia are misdiagnosed as neurotypical, vital for early intervention.
- The majority vote rule not only provides balanced specificity and sensitivity but also aligns with our goal of prioritizing the identification of dysgraphia, accepting a higher tolerance for false positives in neurotypical cases.

VI. Conclusion

Dysgraphia is a complex neurological disorder that often goes underdiagnosed or misdiagnosed. Particularly because it frequently co-occurs with other neurodevelopmental conditions like dyslexia and ADHD, early detection is crucial for effective intervention. However, current diagnostic methods lack the objectivity and consistency needed for accurate identification.

This study demonstrates the potential of 1D Convolutional Neural Networks (CNNs) in accurately diagnosing dysgraphia by analyzing time series data from handwriting samples. Our proposed method significantly improves upon previous machine learning techniques, achieving a diagnostic accuracy of 83.3% and sensitivity of 100%. This advancement not only enhances early diagnosis but also paves the way for more targeted interventions, offering a transformative approach to managing dysgraphia in children across educational and clinical settings.

Acknowledgment

Completion of this study was made possible through the resources provided by Carnegie Mellon University, the Scholar Launch Research Practicum, and valuable mentorship from Dr. Ganesh Mani of the Software and Societal Systems Department in the School of Computer Science.

REFERENCES

[1] Cleveland Clinic, "Dysgraphia: What It Is, Symptoms, Diagnosis & Treatment," 2022. https://my.clevelandclinic.org/health/diseases/23294-dysgraphia

[2] P. Drotár and M. Dobeš, "Dysgraphia detection through machine learning," Scientific Reports, vol. 10, no. 1, 2020. doi: 10.1038/s41598-020-78611-9

[3] S. M. Cahill, "Where Does Handwriting Fit In?," Intervention in School and Clinic, vol. 44, no. 4, pp. 223–228, 2009. doi: 10.1177/1053451208328826

[4] J. Kunhoth, S. Al-Maadeed, S. Kunhoth, and Y. Akbari, "Automated Systems for Diagnosis of Dysgraphia in Children: A Survey and Novel framework," arXiv, 2022. doi: 10.48550/arxiv.2206.13043

[5] D. Frye, "What is dysgraphia?," 2023. https://www.additudemag.com /what-is-dysgraphia-understanding-common-symptoms/

[6] F. Ashraf and N. Najam, "Prevalence and comorbidity of non-clinical Dyslexia, Dysgraphia and Dyscalculia in Pakistan," Pakistan Journal of Medical Sciences, vol. 36, no. 7, 2020. doi: 10.12669/pjms.36.7.2486

[7] REL Appalachia, "What does the research say about the association between handwriting instruction and reading development in early childhood (PreK–grade 3)? What interventions exist to improve handwriting for young children?," Jun. 1, 2019. [Online]. Available: https://ies.ed.gov/ncee/rel/Products/Region/appalachia/Ask-A-REL/-89941

[8] Peet, "Dysgraphia-detection-through-machine-learning," 2022. https:// github.com/peet292929/Dysgraphia-detection-through-machine-learning

[9] P. J. Chung, D. R. Patel, and I. Nizami, "Disorder of written expression and dysgraphia: definition, diagnosis, and management," Translational Pediatrics, vol. 9, no. S1, pp. S46–S54, 2020. doi: 10.21037/tp.2019.11.01

[10] T. Gargot et al., "Acquisition of handwriting in children with and without dysgraphia: A computational approach," PLOS ONE, vol. 15, no. 9, 2020, Art. no. e0237575. doi: 10.1371/journal.pone.0237575

[11] Nils, "Introduction to 1D convolutional neural networks in KERAS for time sequences," 2018. https://blog.goodaudience.com/introduction-to-1d-convolutional-neural-networks-in-keras-for-time-sequences-3a7ff801a2cf

[12] Amazon.com, "Wacom Intuos Pro Large Bluetooth graphics drawing tablet". https://www.amazon.com/Wacom-digital-graphic-drawing-PTH860/dp/B01MTZ9CCD