Dysgraphia Detection using Machine Learning

Victor Micha, Supervisor: Mounim El Yacoubi

***Abstract*—Dysgraphia, a neurological disorder affecting handwriting and motor skills, requires costly clinical evaluations for diagnosis. This paper re-examines the paper https://www.nature.com/articles/s41598-020-78611-9 and continues an AI-based approach using stroke-based handwriting analysis to automate dysgraphia detection, offering a scalable, cost-effective alternative. By comparing shallow machine learning, deep learning, and clustering methods, we demonstrate the potential of AI to enhance diagnostic accessibility and support early intervention.**

I. Introduction

Dysgraphia, a learning disability impairing handwriting and fine motor coordination, impacts academic and professional performance. Diagnosis often involves expensive specialist assessments, costing hundreds to thousands of dollars, limiting access for many. An automated AI diagnostic tool could lower barriers, enabling timely support. This study compares shallow machine learning, deep learning, and clustering approaches to detect dysgraphia from handwriting strokes, evaluating their effectiveness for accurate, accessible diagnosis.

II. Background

Handwriting data was collected from children using digital pens on tablets, capturing seven-dimensional points (x, y coordinates, pen-down status, timestamps, pressure, azimuth, altitude) for each student, resulting in a list of 7D points per participant. This project explores multiple approaches to analyze this data, comparing methods like aggregating each list into a single feature vector for classification versus segmenting the data into sequences of strokes for clustering, to evaluate their effectiveness in diagnosing dysgraphia.

III. Methodology/Approach

This section details three distinct approaches for dysgraphia diagnosis: shallow machine learning, deep learning, and stroke-based clustering.

A. Shallow ML

To diagnose dysgraphia using shallow machine learning, we aggregate each student's handwriting data into a single feature vector by extracting statistical features (e.g., mean, standard deviation, max, min of speed, stroke length, pressure variability) from CSV files, considering only pen-down data, and optionally incorporating demographic features (sex, hand, age). These vectors form a feature matrix with a binary target variable indicating whether or not they have dysgraphia. We split the data into training and test sets and train three models—Support Vector Machine (SVM), Random Forest (RF), and AdaBoost—using 10-fold stratified cross-validation repeated 10 times, as per the referenced study, to evaluate their performance in classifying students based on handwriting patterns.

B. Deep ML

For the deep learning approach, we first train a Variational Autoencoder (VAE) on a feature matrix (120 students, about 20 features) to learn the distribution of aggregated handwriting features, generating synthetic data (Xfake) by sampling the VAE's latent space, with a 20:1 ratio of fake to real data. A tuned SVM, trained on real features (X) and dysgraphia labels (y), predicts labels (yfake) for the synthetic data, matching the real dysgraphia prevalence. We then train a deep neural network (NN) for binary classification using real data, real plus synthetic data, and synthetic data alone, evaluating performance via 10-fold stratified cross-validation repeated 10 times, consistent with the referenced study, to assess the impact of synthetic data augmentation on classification accuracy.

C. Clustering

In the clustering approach, we segment handwriting time series (x, y, pen-down status, timestamps, pressure, azimuth, altitude) from 120 users into 7628 individual strokes based on pen-down transitions, preserving temporal dynamics. Using K-medoids with Dynamic Time Warping (DTW) as the dissimilarity measure, we cluster a training subset of strokes to identify homogeneous groups reflecting dysgraphia-specific patterns, employing 10-fold cross-validation repeated 5 times. For prediction, a new stroke is assigned to the nearest cluster (via DTW to medoids), with the cluster's dysgraphia proportion as the diagnosis; multiple strokes per user are averaged for a final score, providing a fine-grained, sequence-based diagnostic method.

IV. Results and Discussion

A. Shallow ML

In this section, we compare the performance of SVM, AdaBoost, and RF while analyzing whether or not including SHA (sex, hand, age) in our dataset improves diagnosis accuracy or not.

The performance of shallow machine learning models (SVM, Random Forest, AdaBoost) for dysgraphia classification is summarized below, comparing cross-validation (CV) and test accuracies with and without specific handwriting attributes (SHA). All models were evaluated using 10-fold stratified cross-validation repeated 10 times, with results reported as mean accuracy $\pm$ standard deviation.

Table IV summarizes the mean CV accuracies and differences across classifiers, highlighting the impact of including specific handwriting attributes (SHA).

A positive CV Diff or Test Diff indicates higher accuracies without sex, hand, age (SHA). The differences are mixed (SVM: 0.94%, RF: -0.83%, AdaBoost: 0.53% for CV)—so the improvement is inconsistent and minimal, with SVM and AdaBoost slightly better without SHA, while Random Forest performs better with SHA included. Overall, the results show that there is no meaningful performance change with or without SHA.

B. Deep ML

In this section, we present the results of deep machine learning approaches using a neural network (NN) trained on real, fake, and combined real+fake data for dysgraphia classification.

Table V shows that the neural network (NN) trained on combined real and synthetic data achieves slightly higher cross-validation accuracy than the best shallow machine learning models, indicating that synthetic data generation via variational autoencoder and deep NN training enhances performance. The real+fake NN outperforms both the real-data-only and fake-data-only NNs, with the fake-data NN performing closely. As anticipated, the NN trained solely on real data exhibits poor accuracy due to insufficient training data, which is the motivation for synthetic data generation.

C. Clustering

In this section, we present the results of the clustering approach using K-medoids with dynamic time warping (DTW) on 7628 handwriting strokes from 120 users. The clustering model achieved a test Continuous Accuracy of 85.60% and an average cross-validation (CV) continuous accuracy of 85.47% $\pm$ 0.97% over 10-fold CV repeated 5 times. This performance surpasses both shallow machine learning (best: 74.89%) and deep learning (best: 75.74%) approaches by approximately 10%, highlighting the effectiveness of preserving temporal dynamics in stroke-level analysis for dysgraphia detection.

This approach enables visualization of cluster medoids, enhancing explainability and reliability of the dysgraphia detection model. Figure 2 illustrates the medoid stroke of a cluster containing 835 strokes with a 100% dysgraphia diagnosis rate. The stroke exhibits a distinct speed variation, transitioning from slower movements in the bottom left to significantly faster ones in the top left, suggesting that dysgraphia patients may display faster or shakier handwriting patterns.

Figure 1 depicts the medoid stroke of a cluster with 1437 strokes and a 10.44% dysgraphia diagnosis rate. The stroke demonstrates minimal speed variation, indicating smoother and more consistent handwriting, which may characterize individuals without dysgraphia.

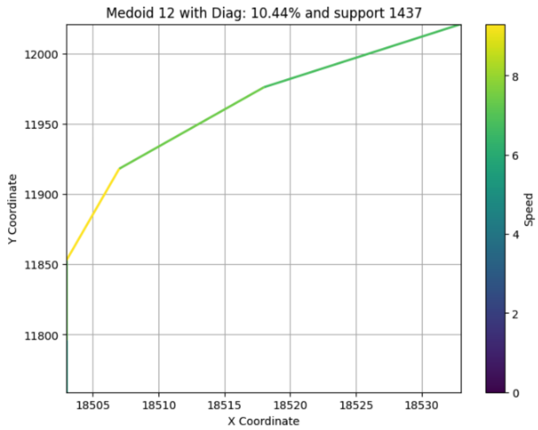

Fig. 1: Visualization of Stroke (Medoid)

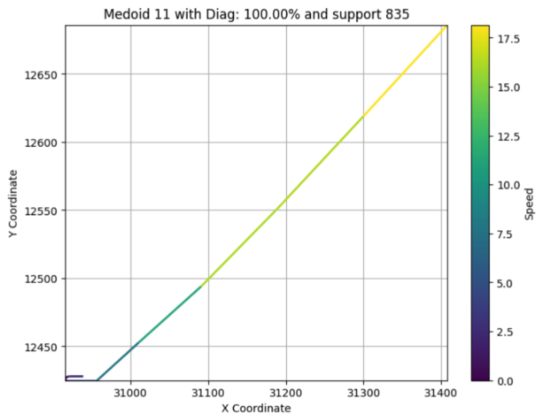

Fig. 2: Visualization of Stroke (Medoid)

V. Conclusion

This study evaluated three machine learning approaches for dysgraphia detection using handwriting stroke data: shallow machine learning, deep learning with synthetic data, and clustering. Shallow models such as support vector machines, random forests, and boosting methods showed limited performance, likely due to the loss of temporal dynamics when aggregating strokes into single feature vectors. The deep learning approach, which combined a variational autoencoder for synthetic data generation with a neural network classifier, demonstrated moderate improvement, suggesting the value of data augmentation in scenarios with limited real data. However, the clustering method, based on K-medoids with dynamic time warping applied directly to individual strokes, achieved the most promising results. It was able to capture fine-grained stroke patterns more effectively than feature-based models. These findings, which extend the prior work of Drotár and Dobeš [1], highlight clustering as the most robust approach for this dataset. Notably, while previous work [1] reported dysgraphia detection accuracy approaching 80% using traditional machine learning methods, our clustering-based approach achieves higher performance.

Overall, such AI-driven methods offer significant potential to improve access to affordable and timely dysgraphia diagnosis.

References

[1] Peter Drotár and Marek Dobeš, *Dysgraphia detection through machine learning*, Scientific Reports, vol. 10, no. 1, 2020, https://doi.org/10.1038/s41598-020-78611-9.

VI. Tables

TABLE I: Performance of SVM Models for Dysgraphia Classification

Kernel	CV Accuracy (All)	CV Accuracy (No SHA)	Test Accuracy (All)	Test Accuracy (No SHA)	CV Diff	Test Diff
Linear	0.6644 ± 0.1457	0.6856 ± 0.1432	0.7333	0.6333	0.0211	-0.1000
Poly	0.6678 ± 0.1559	0.6722 ± 0.1551	0.8000	0.7333	0.0044	-0.0667
RBF	0.5800 ± 0.1339	0.5922 ± 0.1352	0.8333	0.8000	0.0122	-0.0333
Sigmoid	0.5111 ± 0.0544	0.5111 ± 0.0544	0.5667	0.5667	0.0000	0.0000

TABLE II: Performance of Random Forest Models for Dysgraphia Classification

Trees	CV Accuracy (All)	CV Accuracy (No SHA)	Test Accuracy (All)	Test Accuracy (No SHA)	CV Diff	Test Diff
10	0.7322 ± 0.1580	0.7300 ± 0.1550	0.5333	0.6333	-0.0022	0.1000
50	0.7589 ± 0.1516	0.7400 ± 0.1642	0.6333	0.6000	-0.0189	-0.0333
100	0.7544 ± 0.1408	0.7478 ± 0.1514	0.6000	0.6000	-0.0067	0.0000
200	0.7500 ± 0.1417	0.7444 ± 0.1560	0.6000	0.6333	-0.0056	0.0333

TABLE III: Performance of AdaBoost Models for Dysgraphia Classification

Estimators	CV Accuracy (All)	CV Accuracy (No SHA)	Test Accuracy (All)	Test Accuracy (No SHA)	CV Diff	Test Diff
50	0.7444 ± 0.1519	0.7500 ± 0.1590	0.7000	0.7333	0.0056	0.0333
100	0.7456 ± 0.1417	0.7422 ± 0.1594	0.6333	0.7000	-0.0033	0.0667
200	0.7422 ± 0.1439	0.7511 ± 0.1533	0.7000	0.7667	0.0089	0.0667
500	0.7378 ± 0.1453	0.7478 ± 0.1522	0.6667	0.7667	0.0100	0.1000

TABLE IV: Mean CV Accuracies and Differences Across Shallow ML Classifiers

Classifier	CV Accuracy (All)	CV Accuracy (No SHA)	Mean CV Diff	Mean Test Diff
SVM	60.58%	61.53%	0.94%	-5.00%
Random Forest	74.89%	74.06%	-0.83%	2.50%
AdaBoost	74.25%	74.78%	0.53%	6.67%

TABLE V: Performance of Neural Network Models for Dysgraphia Classification

Training Data	Training Accuracy	CV Accuracy
Real Data Only	45.83%	54.42% $\pm$ 17.87%
Fake Data Only	77.88%	74.95% $\pm$ 27.14%
Combined Real + Fake Data	83.65%	75.74% $\pm$ 25.56%