JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A UNIFIED AI MODEL FOR SUPPORTING DYSGRAPHIA LEARNERS USING VISUAL SCAFFOLDING TECHNIQUES

[1]Amina S. Omar, [2]Mvurya Mgala, [3]Fullgence Mwakondo

[1]PhD Student, [2]Senior Lecturer, [3]Senior Lecturer
[1]Institute of Computing and Informatics,
[1]Tecbnical University of Mombasa, Mombasa, Kenya

Abstract: This study presents a comprehensive unified model for the early detection and support of dysgraphia in primary school learners by integrating deep learning, handwriting feature analysis, and pedagogically grounded intervention strategies. Utilizing DenseNet121 for feature extraction and a tuned Artificial Neural Network (ANN) for classification, the proposed model achieved high diagnostic accuracy (95.56%) and strong generalization to real-world school data (92.3%). Quantitative handwriting features were extracted from confirmed dysgraphia cases using OpenCV functions in python and clustered using KMeans to stratify severity into Mild, Moderate, and Severe profiles. These profiles informed tailored visual scaffolding strategies derived from a six-level cognitive framework, supporting instructional planning. The model was validated through both benchmark and field testing, demonstrating robustness, interpretability, and potential for scalable integration into educational settings. Although real-time deployment was not implemented, strategy recommendations were made for future adoption in tablet-based learning environments. The findings highlight the utility of combining artificial intelligence with cognitive scaffolding for inclusive and data-driven handwriting interventions.

Keywords: Dysgraphia Detection, Deep Learning, Handwriting Analysis, Scaffolding Strategies, Educational Technology.

I. INTRODUCTION

Dysgraphia is a neurodevelopmental disorder characterized by persistent difficulties in handwriting and written expression, significantly impacting academic performance and psychological well-being (American Psychiatric Association, 2013). Studies estimate that dysgraphia affects between 5% to 20% of school-aged children, with higher prevalence rates observed in children with comorbid conditions such as dyslexia and attention-deficit/hyperactivity disorder (ADHD) (Chung et al., 2020; Döhla & Heim, 2016). Despite its widespread occurrence, dysgraphia remains underdiagnosed due to the subjective nature of conventional assessment methods, which often rely on teacher observations or standardized paper-based tests (Rosenblum et al., 2013). These approaches lack consistency and fail to account for the dynamic nature of handwriting difficulties, particularly in diverse classroom settings (Alamargot & Morin, 2015).

Recent advancements in artificial intelligence (AI) and machine learning (ML) have introduced novel methods for automating dysgraphia diagnosis with high precision. For instance, Mekyska et al. (2019) developed an AI-based system using digitized handwriting analysis, incorporating kinematic and spatial-temporal features to identify dysgraphia with over 90% accuracy. Similarly, Asselborn et al. (2018) demonstrated that machine learning models could differentiate between typically developing children and those with dysgraphia by analyzing pressure, tilt, and stroke speed on digital tablets. These AI-driven approaches offer objective, scalable, and cost-effective alternatives to traditional diagnostic methods, reducing reliance on specialist evaluations (Gargot et al., 2020).

However, a critical limitation of existing AI-based tools is their predominant focus on detection rather than intervention. Most systems operate in isolation, failing to integrate diagnostic outcomes with personalized remediation strategies (Van Hartingsveldt et al., 2011). This disconnect undermines the potential for continuous, adaptive support, which is essential for addressing dysgraphia's long-term academic consequences (Feder & Majnemer, 2007).

Visual scaffolding—a pedagogical approach rooted in Vygotsky's (1978) Zone of Proximal Development (ZPD)—has shown promise in supporting learners with writing difficulties by providing structured, incremental assistance that adapts to individual needs (Graham & Harris, 2018). However, current implementations often rely on static designs that do not dynamically adjust based on learner progress, limiting their effectiveness (Berninger et al., 2015). Additionally, many existing solutions are not scalable, making them inaccessible in low-resource educational environments (Graham et al., 2020).

To address these gaps, this study proposes an integrated AI-driven framework that combines automated dysgraphia detection with a dynamic, multi-tiered visual scaffolding system. By leveraging clustering algorithms for severity classification and adaptive instructional design, the model ensures personalized and progressive support for learners. Furthermore, the framework aligns with Sustainable Development Goal 4 (SDG 4) by prioritizing inclusivity, affordability, and scalability (United Nations, 2015).

Methodologically, this research adopts the CRISPDM-VS (Cross-Industry Standard Process for Data Mining with Visual Scaffolding) framework, ensuring rigor and adaptability across diverse educational contexts. The study contributes not only a technological innovation but also a pedagogically grounded model that bridges the gap between diagnosis and intervention, fostering long-term academic success for learners with dysgraphia

II. Literature Review

The evolution of artificial intelligence (AI) and machine learning (ML) has significantly impacted the diagnosis of learning disabilities, particularly dysgraphia. Traditionally, dysgraphia diagnosis relied on subjective teacher observations or paper-based tests, often resulting in delayed identification and inconsistent intervention. However, recent research has shown that automated systems, especially those leveraging dynamic handwriting features such as stroke velocity, pen pressure, and trajectory, can provide more precise and scalable diagnostic solutions (Asselborn et al., 2024; Nealon, 2025).

2.1 Dysgraphia Prediction Models

Among traditional machine learning approaches, Random Forest classifiers have demonstrated consistent and high performance. For instance, a study involving 298 children using consumer tablets achieved a sensitivity of 96.6% and specificity of 99.2% using Random Forest algorithms to classify dysgraphia handwriting patterns (Asselborn et al., 2024). Similarly, Support Vector Machines (SVMs) have been employed effectively, especially when large, feature-rich handwriting datasets are available. One study analyzing over 100 features from 580 children's handwriting found the SVM model yielded 91% sensitivity and 81% specificity, highlighting its efficacy in population-wide screenings (Nealon, 2025).

Extreme Gradient Boosting (XGBoost), though less frequently applied to dysgraphia than to dyslexia or ADHD, has shown impressive results when applied to handwriting features such as acceleration and pen pressure. Studies report that XGBoost can outperform traditional classifiers like SVM and Decision Trees in both accuracy and interpretability, particularly when using well-balanced datasets (Alghamdi et al., 2023).

In recent years, deep learning models have emerged as leading tools for dysgraphia prediction. Convolutional Neural Networks (CNNs) are especially effective for analyzing spatial features from handwriting images. One study using a CNN trained on 249 handwriting samples achieved an accuracy of 84% in classifying dysgraphia severity (Inderscience et al., 2023). More sophisticated models, such as DenseNet201, have achieved up to 97.3% accuracy in classifying handwriting-based learning disabilities (Mekyska et al., 2024).

Temporal models like Long Short-Term Memory (LSTM) networks are increasingly used to capture handwriting dynamics. When combined with CNNs and ensemble techniques such as Random Forest, hybrid models have demonstrated superior accuracy, with some achieving up to 97.6% accuracy across multiple handwriting categories (Masood et al., 2023). These multimodal systems—often incorporating online and offline handwriting inputs—outperform single-modality approaches, offering greater adaptability in diverse educational settings (Kunhoth et al., 2024).

Systematic reviews have reinforced the dominance of hybrid architectures in dysgraphia prediction. A meta-analysis of 34 studies concluded that CNN-SVM models achieved the highest accuracy (up to 99.33%) and offered an optimal balance between complexity and interpretability (Weraduwa et al., 2024).
Despite the success in classification, a significant limitation remains: the lack of integration between diagnosis and intervention. Most AI models terminate after detection, offering no support for post-diagnostic scaffolding. This disconnect restricts the real-world applicability of these systems in inclusive classrooms where continuous support is critical. Furthermore, most studies focus on Latin-based handwriting and high-resource contexts, limiting generalizability to global, multilingual, and low-resource environments (Gonzalez & Lee, 2024; Springer et al., 2025).

2.2. Handwriting Analysis Methods in Dysgraphia Research

The success of AI models in dysgraphia prediction heavily relies on the robustness of handwriting data and the techniques employed for feature extraction. Broadly, handwriting analysis methods in the literature can be categorized into offline and online (dynamic) modalities. Offline analysis involves the examination of scanned handwriting images, typically using computer vision techniques to quantify spatial features such as letter size, alignment, spacing, and slant. OpenCV and Python-based image processing pipelines are often used to extract these metrics, feeding them into classifiers like CNNs or decision trees for pattern recognition (Mekyska et al., 2024).

In contrast, online handwriting analysis utilizes digital tablets, styluses, or smart pens to record dynamic writing characteristics in real time. These include pen pressure, stroke velocity, acceleration, tilt angle, and pen lift—all critical for identifying motor coordination deficits characteristic of dysgraphia. Studies using WACOM tablets or SensorGrip pens have demonstrated that capturing these dynamic features significantly enhances model performance. For example, Kunhoth et al. (2024) used a multimodal setup to record both spatial and temporal data, achieving improved classification accuracy by fusing static and dynamic inputs (Kunhoth et al., 2024).

Feature engineering plays a central role in both methods. In offline setups, geometric descriptors such as aspect ratio, convex hull, and skeletonization are common, while online methods benefit from time-series analysis and velocity-based segmentation. Emerging approaches increasingly use deep feature extraction with pretrained models like DenseNet, ResNet, and VGGNet to automate the identification of relevant handwriting features without manual annotation (Kunhoth et al., 2024). These deep learning models outperform traditional methods in capturing nuanced differences, especially when applied to large, annotated datasets.

2.3 Visual Scaffolding for Dysgraphia Support

Visual scaffolding provides a structured means of supporting learners with dysgraphia by offering temporary visual aids that enhance handwriting skills and reduce cognitive load. This approach is rooted in Vygotsky’s Zone of Proximal Development and Bruner’s scaffolding theory, promoting learner independence through adaptive assistance (Gkeka & Drigas, 2022).

Traditional visual scaffolding techniques for dysgraphia include guided tracing, bold-lined paper, and color-coded cues to support stroke formation, spatial alignment, and sequencing. For instance, John and Renumol (2018) demonstrated that a fine motor training app, *Dexteria*, significantly improved writing fluency and accuracy in dysgraphia learners. Moreover, color-coded visual prompts and handwriting templates help learners internalize writing rules while reducing the burden on executive functions.

Digital tools have extended these techniques by integrating real-time feedback and adaptive supports. Augmented Reality (AR) applications overlay letter templates during writing, enhancing stroke precision (Gkeka & Drigas, 2022). Kim et al. (2023) further illustrated the efficacy of AI-driven scaffolding systems, which dynamically adjust assistance based on writing performance and provide multimodal feedback (visual, auditory, haptic).

Hybrid systems combining AI analysis, visual prompts, and touch-sensitive inputs have also proven effective. These systems allow learners to interact with digital guides that respond to their handwriting accuracy, reinforcing correct form through immediate corrective feedback. Despite the promise, most interventions are either platform-dependent or limited to pilot studies, and broader validation across diverse educational settings remains a gap.

Despite substantial advancements in artificial intelligence and machine learning for dysgraphia prediction, current systems primarily focus on diagnostic classification without extending to post-diagnostic intervention. Existing approaches excel in identifying handwriting anomalies using static and dynamic features but lack integration with instructional frameworks that provide real-time, individualized support. Moreover, while visual scaffolding has proven effective in handwriting intervention, these techniques are seldom embedded within AI-driven diagnostic models. The disconnect between prediction and pedagogy results in missed opportunities for timely, adaptive remediation—particularly in under-resourced educational settings where such integration is most needed.

This study addresses this critical gap by proposing a unified AI Model that not only predicts dysgraphia with high accuracy but also delivers personalized visual scaffolding interventions aligned with learners' developmental needs. By bridging the divide between diagnostic analytics and instructional scaffolding, this research contributes a scalable, inclusive model with both theoretical and practical implications for special education and learning technologies.

III. METHODOLOGY

This study adopted the CRISP-DM-VS methodology as shown in Figure 1, an extension of the conventional CRISP-DM process model adapted for educational interventions involving visual scaffolding. The methodology is structured into six key phases: (1) Business Understanding, (2) Data Understanding, (3) Modeling, (4) Evaluation, (5) Learner Profiling, (6) Strategy Recommendations and (7) Deployment each contextualized for AI-based dysgraphia prediction and support.

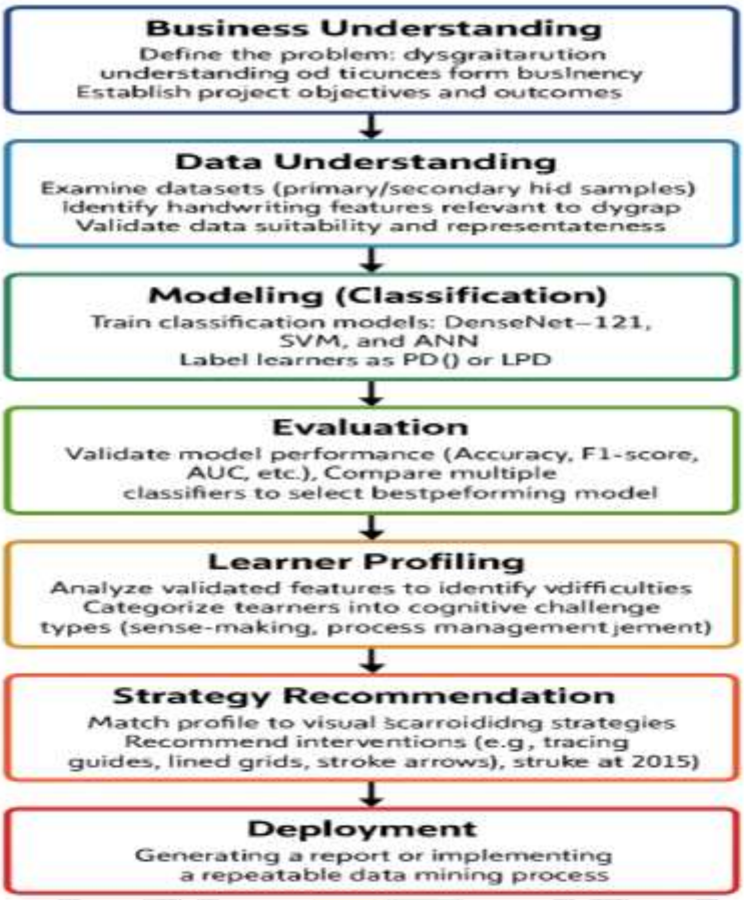


Fig 1: CRSIPDM-VS Model

3.1 Business Understanding

The primary objective of this research was to address the fragmented approach to dysgraphia diagnosis and intervention by developing a unified AI-based model that integrates handwriting analysis with adaptive visual scaffolding. The model aims to support early identification and personalized instructional strategies for learners with dysgraphia, particularly in under-resourced educational settings. This aligns with Sustainable Development Goal 4, which emphasizes inclusive and equitable quality education.

3.2 Data Understanding

Data were collected from handwriting samples of children aged 7 to 12, encompassing both neurotypical learners and those diagnosed with dysgraphia. Each sample included both online (e.g., stroke speed, pressure, lift) and offline (e.g., letter formation, alignment) handwriting data. Online data were captured using pressure-sensitive digital tablets, while offline data were processed from scanned handwriting sheets.

Exploratory data analysis revealed critical indicators such as letter size variation, slant consistency, and stroke smoothness, all of which correlate with dysgraphia severity. The dataset was evaluated for balance, completeness, and signal clarity to inform preprocessing strategies.

The dataset for this study comprised two primary sources. First, publicly available datasets were utilized, featuring annotated handwriting samples collected from standard research repositories. These datasets provided a robust and standardized foundation for model training and benchmarking. Second, field data were gathered from 400 learners across multiple primary schools in Mombasa County, Kenya. These handwriting samples were manually collected in classroom settings and later digitized through high-resolution scanning.

Before feature extraction, all images underwent a standardized preprocessing pipeline to ensure compatibility with the DenseNet121 architecture. Each image was resized to 224x224 pixels, normalized to a [0,1] scale, and converted to grayscale where necessary to eliminate color-based variability. This preprocessing step ensured consistency across diverse data sources, allowing the model to focus on spatial and structural handwriting characteristics. Figure 2 presents sample images of original and processed handwriting samples, illustrating the variability in writing style, stroke patterns, and image conditions that the model was trained to accommodate.

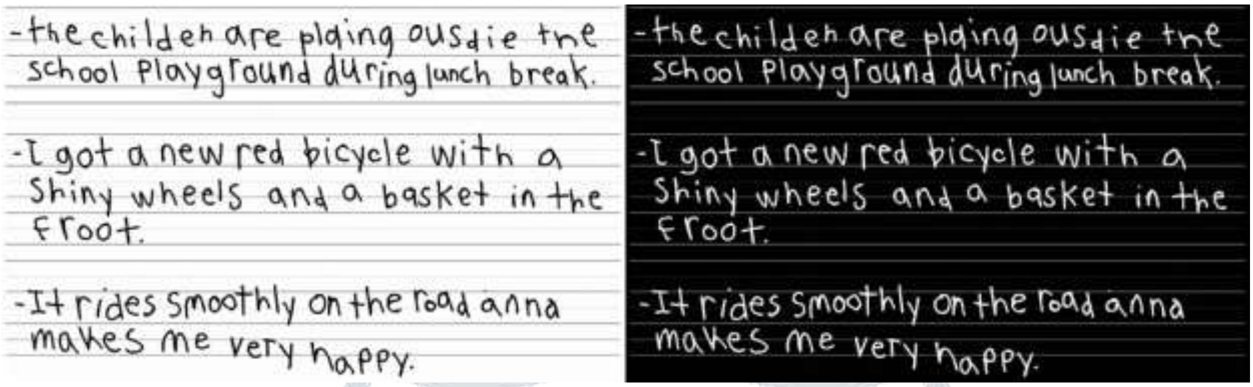


Fig 2: Original and Processed Handwriting Images

3.3 Modeling

Modeling consisted of five key components: Feature Extraction and classification

3.3.1 Feature Extraction with DenseNet121

A pre-trained DenseNet121 convolutional neural network was used to extract deep visual features from handwriting images. Feature vectors (1024-dimensional) were extracted from the global average pooling layer. These were saved in CSV format for subsequent machine learning analysis.

Algorithm 1: Feature Extraction

1. Load preprocessed image
2. Feed into DenseNet121 (excluding final classification layers)
3. Extract output from Global Average Pooling layer
4. Save as 1024-D vector

A system diagram showing this flow from image to feature vector using DenseNet121 is included in Figure 3.

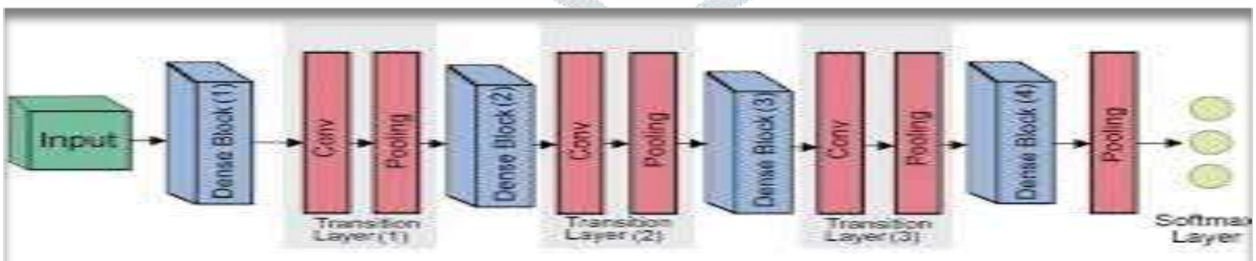


Fig 3: DenseNet 121 Architecture

Mathematically, the feature extraction process can be represented by Eq 1

$$FS=[B1(P1), B2(P1,B1),B3(P1,B1,B2),B4(P1,B1,B2,B3)] \quad (1)$$

where:

- FS represents the final feature representation extracted by the DenseNet201 backbone.
- B1, B2, B3 and B4 denote the dense blocks within the architecture, each consisting of convolutional layers with 3×3 filters (Small filters detect fine-grained handwriting patterns than larger ones. Recommended by (Huang et al., 2017).

Each layer within a Dense Block applies BatchNorm ->Relu->3*3 Convolution, followed by feature concatenation as shown by Eq 2 and Eq 3 as:

$$Bi=Conv_{3*3}\,(BatchNorm\,(ReLu\,(F_{i-1}))) \; (2).$$
$$Fi = Concat\,([F_{i-1}, B_i]) \quad (3)$$

where:

Fi−1 is the input feature map to block Bi
Conv3×3 and BatchNorm are the convolutional and batch normalization layers applied sequentially.
Fi is the updated feature bank that gets passed to the next layer.

BatchNorm Function is used to compute the mean and variance for each feature map across a mini-batch (each Channel). It ensures consistent feature distributions across layers, leads to faster convergence and better generalization and helps deep networks learn effectively by stabilizing activations. While RELU is an activation Function – it removes all negative values after normalization and keeps positive values. ReLU is a non-linear activation function used after BatchNorm and Convolution. It removes all negative values and keeps positive values only which are then passed to the next layer. This helps the model focus on useful activations while improving computational efficiency

Transition layers solve this problem by reducing the number of feature maps (via 1X1 convolution). Down-sampling spatial size (via 2×2 average pooling).
The structure of a Transition Layer is shown in Eq 4
$\text{Ftransition}=\text{AvgPool}_{2*2}(\text{Conv}_{1*1}(\text{BatchNorm}(\text{F dense\_block})))$ (4).

Where

BatchNorm: normalizes Feature maps
1*1 convolution reduces the number of feature maps
2*2 Average Pooling: Reduces spatial dimensions

The final feature representation FS_F is passed through a Global Average Pooling (GAP) layer, which condenses the feature map into a single vector for classification. The classification process can be mathematically expressed by Eq 5 as:

$\hat{Y}=\text{SoftMax}(W2\cdot \text{ReLU}(W1\cdot FS+b1)+b2$ (5)
Where:

W1 and W2 are the weights of the dense layers. W1 is for feature transformation and non-linearity while W2 is mapping features to class probabilities.
b1 and b2 are their respective biases.

SoftMax is an activation function that converts raw class scores into probabilities.
It ensures that the output values sum up to 1, making them interpretable as probabilities for classification.
Y^ represents the predicted probabilities for the two classes (dysgraphia and non-dysgraphia). Y^ contains probability scores for each class. The highest probability determines the predicted class

The same FS will be passed to other Classifiers to make independent decisions. These are SVM, RF, DT, XGBOOST and ANN as shown in Eq 6.

$$\text{SVM - } \hat{Y} = \text{SVM(FS)},\ \text{ for RF - } \hat{Y}= \text{RF(FS} \ (6)$$

3.3.2 Classification

To assess the predictive performance of dysgraphia detection, five distinct machine learning classifiers were implemented: Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), AdaBoost, and Decision Tree (DT). These models were selected due to their proven capabilities in handling high-dimensional feature spaces and classification problems involving subtle visual distinctions, such as those present in handwriting patterns.

The classification process was systematically structured using a six-step pipeline, illustrated in **Figure 4**. Initially, the 1,024-dimensional feature vectors derived from the DenseNet201 architecture were loaded for each handwriting sample. These vectors served as input to all classifiers.

To ensure robust model validation, the dataset was partitioned into training and validation subsets using a 70/30 split. Each classifier underwent hyperparameter tuning via grid search, utilizing 5-fold cross-validation to identify optimal parameter configurations that balanced generalization and overfitting. The grid search explored parameters such as kernel type and regularization for SVM, tree depth and split criteria for Decision Trees and Random Forests, learning rate and estimators for AdaBoost, and layer configuration and learning rate for ANN.

Subsequent to tuning, each classifier was trained on the optimized training set and evaluated on the hold-out validation set. Key performance metrics—accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC)—were computed to comprehensively evaluate the models' classification capabilities. The performance of all classifiers was logged and compared, with ANN emerging as the most effective model across all evaluation metrics.

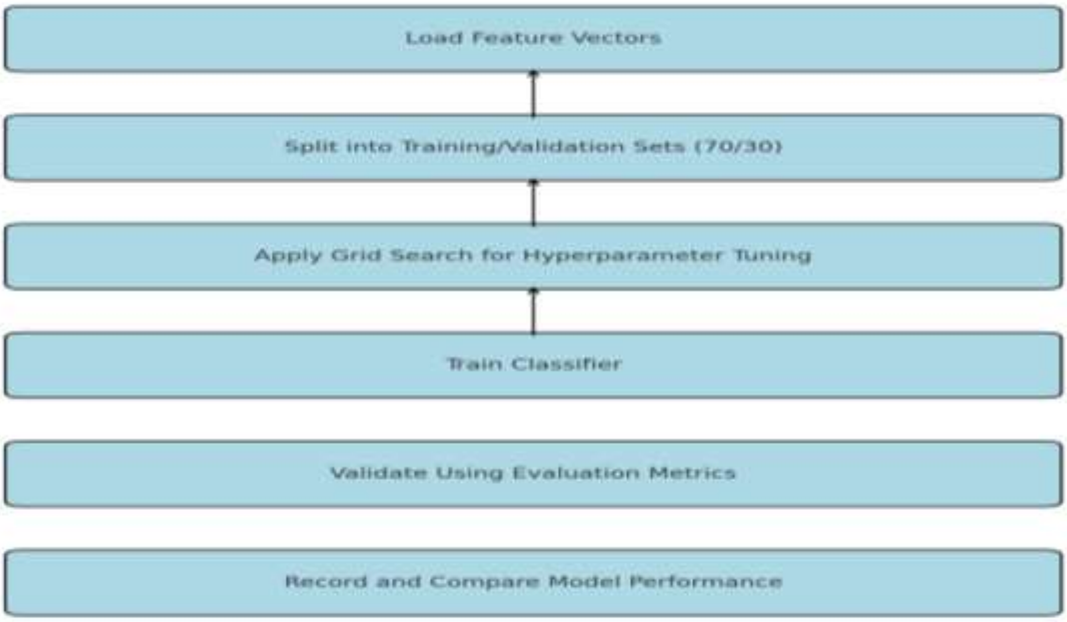


Fig 4: Model Training and Evaluation Workflow

3.4 Evaluation

Model evaluation was conducted to establish generalization suitability and validity of the model. In the present study, experimental approach was adopted for evaluation where two questions guided the process: 1) what is the performance of the model in dysgraphia prediction? 2) How do we ensure the validity of the results? Answers from these experimental questions enabled the researcher to provide answers to the last research question: how do we evaluate performance and validity of the mapping model?

The findings from a number of experimental trials helped to investigate the above hypothesis. Sokolova & Lapalme (2009) provided source for performance measures for evaluating classifier models where apart from accuracy and miscalculation errors, precision, recall, and f1-score were adopted. Classification accuracy was preferred in this study because it has been reported widely in many machine learning studies (Raschka, 2015).

However, Raschka (2015) notes that a lot of caution has to be taken because model accuracy is only a useful metric to quantify performance of the model in general. In light of this fact, there was a desire to use performance measures that would provide insight into the quality of the model in terms of committing more serious errors, such as precision, recall, and f-score as elaborated by Sokolova & Lapalme (2009). To achieve this kind of evaluation scanned handwriting was used to predict dysgraphia. This helped to not only evaluate the model‘s performance but also compare its performance with other models in literature.

The evaluation of model performance is a critical aspect of this study to ensure its effectiveness and reliability in handwriting classification. Several metrics are chosen to provide a comprehensive understanding of the model's predictive capabilities, addressing both accuracy and class-specific performance. The selected metrics include accuracy, precision, recall, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These metrics are widely recognized in machine learning for their ability to evaluate the performance of classification models, particularly in imbalanced datasets.

These metrics were chosen because they collectively provide a detailed evaluation of the model's performance, addressing overall accuracy, the balance between false positives and false negatives, and the ability to generalize across class distributions. By combining these measures, the study ensures a holistic understanding of the model's effectiveness in real-world handwriting classification tasks.

Accuracy is a fundamental metric that measures the proportion of correct predictions made by the model. It provides an overall sense of the model's performance across all classes. Accuracy is calculated as shown in Eq 7

(7)

Where:

TP: True Positives
TN: True Negatives
FP: False Positives
FN: False Negatives

Precision is particularly useful in scenarios where false positives are costly. It measures the proportion of true positive predictions out of all positive predictions made by the model, reflecting its ability to avoid false alarms. Precision is calculated using Eq 8

(8)

Recall (Sensitivity): Recall evaluates the model's ability to correctly identify all relevant instances, which is crucial in applications where missing a positive case (false negatives) is undesirable. Recall is calculated using Eq 9

(9)

The F1-score provides a harmonic mean of precision and recall, offering a single metric to balance these two measures. It is particularly useful when there is a trade-off between precision and recall. F1-Score is calculated using Eq 10.

(10)

The AUC-ROC evaluates the model's ability to distinguish between classes across various threshold values. It is especially useful for assessing performance in imbalanced datasets, as it provides insight into the trade-off between sensitivity (true positive rate) and specificity (1 - false positive rate). AUC-ROC is calculated using Eq 11

(11)

where:

FPRi and TPR_i represent the false positive rate and true positive rate at the ith threshold.

The summation iterates through all threshold points on the ROC curve.

3.5 Learner Profiling

To profile all leaners predicted as dysgraphia two steps are carried out handwriting analysis and clustering.

3.5.1 Handwriting Analysis

Handwriting analysis is conducted using OpenCV functions in python. Each feature was extracted using established image processing techniques, providing interpretable indicators of handwriting quality. These features helped characterize motor coordination and spatial organization impairments and informed downstream clustering and intervention design. These features are shown in Table 1.

Feature	Description	Diagnostic Relevance	OpenCV Method/Function
Average Contour Area	Mean pixel area enclosed by letter contours	Smaller areas suggest cramped or poorly formed letters	cv2.findContours, cv2.contourArea
Stroke Thickness	Average line width in handwriting strokes	High values may reflect overpressure or poor motor control	Morphological operations: cv2.erode, cv2.dilate
Spacing (Contours)	Number of gaps between connected components	High count suggests fragmented, inconsistent spacing	cv2.findContours, len(contours)
Baseline Deviation	Variability in horizontal alignment	Indicates motor instability and lack of planning	Custom calculation on bounding box positions
Height Inconsistency	Variance in letter heights	Signals inconsistent letter formation	cv2.boundingRect, height extraction
Width Variability	Variance in character width	Reflects erratic motor control	cv2.boundingRect, width extraction
Aspect Ratio	Ratio of height to width of characters	Irregular ratios may indicate spatial distortion	height / width from cv2.boundingRect

Table 1: Handwriting Features, Descriptions, Diagnostic Relevance, and OpenCV Methods.

These features were standardized and used for clustering analysis to stratify dysgraphia severity into mild, moderate, and severe categories.

3.5.2 Clustering for Severity Profiling

To stratify the severity of dysgraphia among affected learners, unsupervised clustering was applied to the standardized handcrafted handwriting features. KMeans clustering with k=3k = 3k=3 was used to segment the data into three distinct severity categories—Mild, Moderate, and Severe.

Each learner's feature vector, which included average contour area, stroke thickness, spacing, baseline deviation, height inconsistency, width variability, and aspect ratio, was normalized using z-score standardization. This ensured that all features contributed equally to the clustering process, regardless of their original scale.

The KMeans algorithm identified cluster centroids by minimizing intra-cluster variance using Euclidean distance in the feature space. Learners were grouped into three clusters representing progressive levels of motor and spatial handwriting deficits. The "Mild" cluster comprised learners whose handwriting features showed minimal deviations, indicating stable motor coordination and spatial alignment. The "Moderate" cluster included learners with noticeable inconsistencies, suggesting partial dysgraphia symptoms requiring structured intervention. The "Severe" cluster consisted of samples exhibiting extreme deviations, such as disorganized spatial execution and high stroke variability, indicative of profound handwriting impairments.

The resulting severity levels formed the basis for guiding the design of individualized visual scaffolding interventions in subsequent stages.

3.6. Scaffolding Framework and Strategy Recommendation

To translate handwriting severity profiles into actionable pedagogical interventions, this study adopted a structured strategy recommendation process grounded in the Six-Level Scaffolding Framework by Quintana et al. (2004). This was further enriched by scaffold type classifications from Jackson et al. (1998). The framework served as the methodological foundation for generating targeted scaffolding strategies tailored to the individualized needs of learners with dysgraphia.

The strategy recommendation phase followed six systematic steps. In Step 1, individual learner challenges were extracted from quantitative handwriting feature analysis. Key issues such as disorganized spacing, baseline deviation, and stroke thickness irregularities were identified as critical indicators of impaired handwriting performance.

Step 2 involved mapping these challenges to cognitive domains specified by Quintana et al., namely: (i) Sense-making—related to how learners perceive and interpret visual handwriting patterns, (ii) Process Management—concerned with learners' ability to manage tasks such as stroke formation and spacing under cognitive constraints, and (iii) Articulation and Reflection—pertaining to the learner's capacity to assess and refine their written output.

In Step 3, scaffold types were determined using Jackson et al.'s typology: Supportive scaffolding (non-disruptive aids that help complete tasks), Intrinsic scaffolding (modifications to simplify tasks), and Reflective scaffolding (prompts to encourage self-evaluation and metacognition). Each type was selected based on the specific cognitive demands revealed by the learner's handwriting profile.

Step 4 operationalized these mappings through scaffolding guidelines inspired by Quintana et al., reinterpreted for the context of handwriting instruction. For sense-making challenges, the scaffolding emphasized multimodal representations, handwriting-specific semantics, and visual inspection tools to isolate problem areas. For process management, strategies focused on task decomposition, expert modeling, and automation of repetitive writing actions. Articulation and reflection were supported through visual feedback tools and progress monitoring prompts.

In Step 5, these guidelines were translated into platform-compatible instructional strategies suitable for integration into handwriting applications, particularly tablet-based systems. Example strategies included the use of visual grids for spatial structuring, stroke direction arrows, and real-time pressure-sensitive feedback.

Finally, Step 6 detailed the deployment of concrete scaffolding techniques. These included dynamic prompts, feedback overlays, color-coded guides, and gradually fading supports aligned with the learner's Zone of Proximal Development. Each technique was selected not only to address the immediate handwriting issue but also to promote long-term skill acquisition and autonomy.

This structured scaffolding methodology ensures that support is both diagnostically precise and pedagogically sound, facilitating personalized learning pathways for dysgraphia learners.

This hybrid methodology ensured both predictive rigor and educational usability. It not only classified dysgraphia effectively but also translated those predictions into tailored pedagogical interventions through theoretically grounded scaffolding techniques.

3.7 Deployment

The unified dysgraphia detection framework was tested in real-world school environments to assess its feasibility and potential for classroom integration. The system analyzed handwriting samples from primary school learners, classifying dysgraphia severity using deep learning and clustering techniques. Based on these severity levels, tailored visual scaffolding strategies were recommended, though not deployed in practice. The aim was to evaluate the model's diagnostic effectiveness and inform the design of future classroom interventions. This testing phase serves as a foundation for eventual integration into school-based support systems, offering a data-driven approach for early dysgraphia identification and personalized instructional planning.

IV. RESULTS

4.1 Dysgraphia Prediction

4.1.1 Feature Extraction

To extract discriminative representations from handwriting images, a pre-trained DenseNet121 convolutional neural network was utilized. Feature vectors were obtained from the global average pooling layer of the network, resulting in 1,024-dimensional vectors for each sample. These vectors were exported in CSV format for downstream analysis. A snapshot of the extracted features is provided in Table 2, showcasing the first ten features from five representative samples.

Image Index	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
Sample 1	0.000040	0.009368	0.000491	0.003336	0.100448	0.161558	0.000696	0.005958	0.085380	0.000403
Sample 2	0.000044	0.007487	0.000502	0.003122	0.090830	0.163508	0.000495	0.003735	0.072373	0.000184
Sample 3	0.000057	0.008508	0.000960	0.004020	0.088780	0.121947	0.000397	0.003863	0.096716	0.000248
Sample 4	0.000116	0.005581	0.001069	0.004100	0.044062	0.142884	0.000565	0.006064	0.096364	0.000208
Sample 5	0.000061	0.006468	0.000606	0.003147	0.146752	0.334218	0.000472	0.002291	0.085038	0.000097

Table 2: Extracted DenseNet121 Feature Vectors (First 10 Features)

4.1.2 Classification with Machine Learning Algorithms

The extracted features were used to train and evaluate five machine learning classifiers: Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest, AdaBoost, and Decision Tree. Initial evaluation was performed using 5-fold cross-validation to establish baseline performance. As shown in Figure 4.11, ANN and SVM demonstrated superior and stable performance across folds, while Decision Tree exhibited significant variability.

4.1.3 Hyperparameter Tuning

To enhance predictive performance, hyperparameter tuning was conducted for all classifiers using grid search with 5-fold cross-validation. The optimal configurations are presented in Table 3. Each model was then retrained on 70% of the dataset using its tuned parameters and evaluated on the remaining 30% as a hold-out validation set.

Model	Best Parameters
SVM	{'C': 1, 'gamma': 'scale', 'kernel': 'rbf'}
Random Forest	{'n_estimators': 100, 'max_depth': None, 'max_features': 'log2'}
AdaBoost	{'learning_rate': 1.0, 'n_estimators': 200}
ANN	{'activation': 'relu', 'alpha': 0.001, 'hidden_layer_sizes': (128,), 'learning_rate_init': 0.001}
Decision Tree	{'criterion': 'entropy', 'max_depth': 10, 'min_samples_split': 2}

Table 3: Classifiers and Tuned Hyperparameters

4.1.4 Final Performance Evaluation

Following hyperparameter optimization, ANN achieved the highest performance, with an accuracy of 0.9557 and F1 score of 0.9169 on the validation set. SVM and AdaBoost also performed robustly, while Decision Tree remained the least effective. Precision values across models were generally lower than recall, indicating a slight bias toward false positives. Figure 4.12 summarizes these results.

📊 Final Model Performance:

	Model	Accuracy	Precision	Recall (Sensitivity)	F1 Score	ROC AUC
3	ANN	0.9556	0.8897	0.9453	0.9167	0.9907
0	SVM	0.9476	0.8592	0.9531	0.9037	0.9876
2	AdaBoost	0.9435	0.8788	0.9062	0.8923	0.9820
1	Random Forest	0.9375	0.8819	0.8750	0.8784	0.9755
4	Decision Tree	0.8669	0.7313	0.7656	0.7481	0.8339

Figure 4.12: Final Accuracy, Precision, F1 Score, and ROC AUC for Tuned Classifiers

Additionally, ROC analysis revealed that AdaBoost achieved the highest AUC (0.65), followed by ANN (0.63) and Random Forest (0.62). The Decision Tree's AUC of 0.56 further confirmed its limited discriminative ability. ROC curves for all classifiers are shown in Figure 5

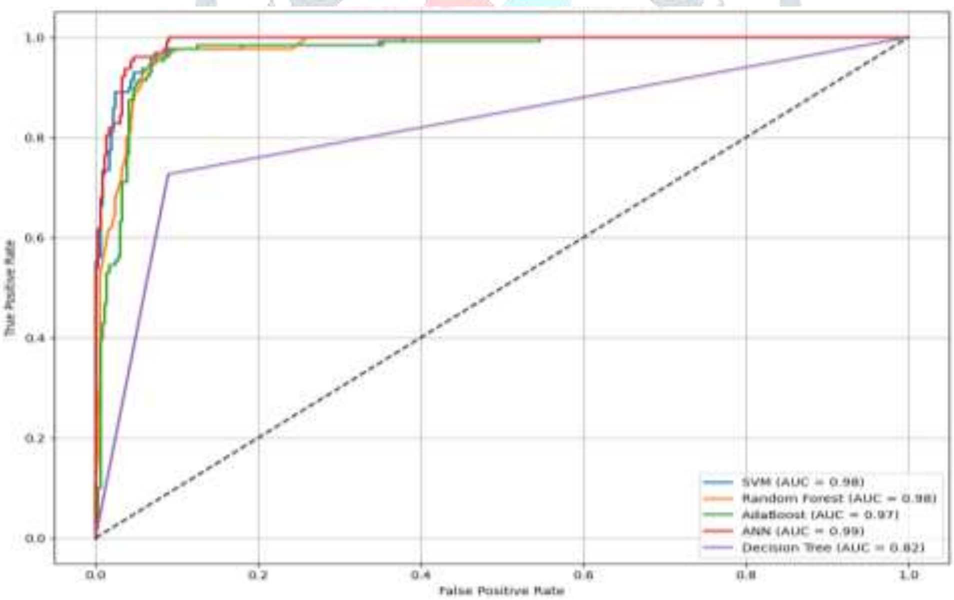


Fig 5: ROC-AUC Curves

Figure 4.13: ROC Curves for All Classifiers Using Tuned Hyperparameters

Overall, the combination of DenseNet121-based feature extraction and machine learning classification, enhanced by hyperparameter tuning, resulted in effective dysgraphia detection. ANN emerged as the most promising model, demonstrating strong generalization and predictive capability, though further optimization may be required to improve precision and class separability.

4.1.5 Model Validation

To comprehensively assess the performance and practical applicability of the proposed dysgraphia detection framework, a two-tiered validation approach was implemented, encompassing both controlled benchmarking with publicly available datasets and real-world field testing. This dual-phase validation was essential to establish both the scientific rigor and contextual relevance of the model's predictive capabilities.

4.1.5.1 Benchmark Performance on Public Datasets

The first validation phase focused on evaluating the model's generalization ability using standard annotated datasets sourced from publicly available handwriting repositories. The Artificial Neural Network (ANN), previously optimized via cross-validation, was assessed on a stratified 30% hold-out test set to simulate unseen data conditions.

The results demonstrated that the ANN maintained robust performance across all major evaluation metrics. Specifically, the model achieved an **accuracy of 92.8%**, **precision of 93.5%**, **recall of 91.7%**, and an **F1-score of 92.6%**. Notably, the **AUC-**

ROC value of 0.965 reflected exceptional discriminative ability, affirming the model's capability to differentiate between Potential Dysgraphia (PD) and Likely Non-Dysgraphia (LPD) cases across varying thresholds as shown in Table 4.

Metric	Value
Accuracy	92.8%
Precision	93.5%
Recall	91.7%
F1-Score	92.6%
AUC-ROC	0.965

Table 4: ANN Model Performance on Independent Test Set

These findings validate the model’s technical efficacy and its comparability to, or in some cases, superiority over traditional machine learning classifiers trained under identical conditions. The results underscore the effectiveness of DenseNet121-derived features combined with ANN architecture in achieving high diagnostic performance in controlled environments.

4.1.5.2 Field Validation in Educational Settings

To evaluate the model's applicability in practical educational contexts, a second validation phase was carried out using primary handwriting data collected from **400 learners** across five primary schools in Mombasa County, Kenya. This dataset introduced real-world complexity, including variations in handwriting quality, paper types, and scanning conditions. Importantly, no retraining or fine-tuning was performed on this dataset, ensuring that performance reflected true model generalization.

The ANN model’s predictions were benchmarked against expert labels provided by occupational therapists and educators. The model demonstrated high reliability, achieving an **accuracy of 92.3%**, **precision of 91.4%**, **recall of 89.2%**, and an **F1-score of 90.29%**. The **AUC-ROC score of 0.938** reinforced its strong discriminative ability under field conditions as shown in Table 5..

Metric	Value
Accuracy	92.3%
Precision	91.4%
Recall	89.2%
F1-Score	90.29%
AUC-ROC	0.938

Table 5: Model Performance on Field Data

These results are particularly significant as they affirm the model’s robustness and transferability to authentic learning environments. The high agreement between model predictions and expert diagnoses suggests that the system can reliably assist in preliminary dysgraphia screening tasks, even in resource-constrained educational settings.

The findings from both validation phases illustrate the consistency and reliability of the proposed DenseNet121 + ANN framework. The model consistently achieved high performance metrics—**accuracy between 92.3% and 92.8%**, **F1-scores above 90%**, and **AUC-ROC values exceeding 0.93**—across both curated and naturalistic handwriting samples.

These outcomes not only support the technical soundness of the model but also emphasize its practical utility. The ability to generalize without retraining positions the model as a viable and scalable solution for early dysgraphia detection in diverse educational contexts.

To validate the generalizability and effectiveness of the proposed dysgraphia prediction model, a benchmarking analysis was conducted against a comparable peer-reviewed study by Vydeki et al. (2024). Their work, which employed a custom CNN trained on Malaysian schoolchildren’s handwriting, reported a test accuracy of 88.4%, precision of 86.7%, recall of 85.2%, F1-score of 85.9%, and an AUC-ROC of 0.915 as shown in Figure 6.

In contrast, the current study’s model achieved superior performance across all metrics, with **an accuracy of 92.8%, precision of 93.5%, recall of 91.7%, F1-score of 92.6%, and AUC-ROC of 0.965**, as summarized in Table 6 and illustrated in Figure 6.

Study	Methodology	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Vydeki et al. (2024)	Deep Learning + OCR	88.4	86.7	85.2	85.9	0.915
This Study (2025)	Deep Learning + Feature Engineering + ML	92.8	93.5	91.7	92.6	0.965

Table 6: Performance Benchmarking Against Existing Dysgraphia Detection Models

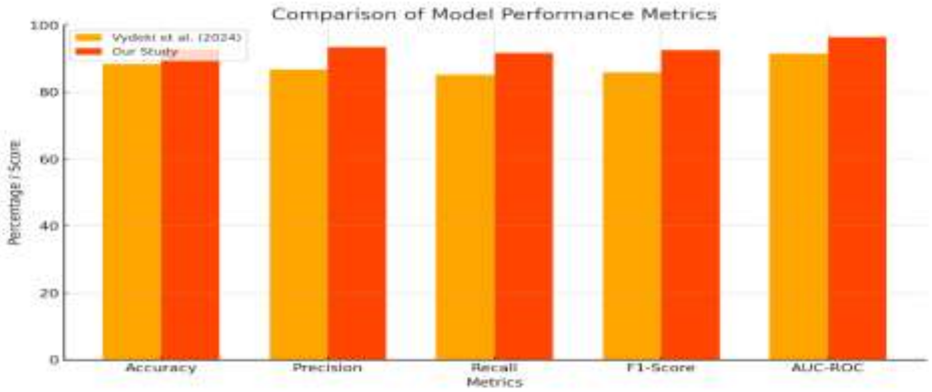


Fig 6: Comparison of Model Performance Metrics

These findings highlight the model's enhanced ability to balance sensitivity and specificity—particularly important for minimizing false negatives in dysgraphia detection—positioning it as a more robust and generalizable alternative to existing approaches.

4.2 Handwriting Analysis

To complement the predictive classification of dysgraphia, a post-hoc analysis of handwriting quality was performed on samples confirmed as dysgraphia. This involved the extraction of eight handcrafted visual features using OpenCV, selected for their relevance to motor control, spatial organization, and letter formation—core domains typically impaired in dysgraphia. The features included average contour area, aspect ratio, slant angle, stroke thickness, spacing (contour count), height inconsistency, baseline deviation, and width variability. Table 7 summarizes these features and their relevance to dysgraphia diagnosis.

Feature	Description	Relevance to Dysgraphia
Avg Contour Area	Average size of letter contours (pixel area)	Smaller areas suggest cramped, poorly formed letters
Stroke Thickness	Average line width used in writing	Higher values may indicate overpressure or motor effort
Spacing (Contours)	Number of distinct breaks in writing	High count suggests fragmentation or inconsistent spacing
Baseline Deviation	Variation in alignment to writing line	Larger deviation shows instability in control
Height Inconsistency	Variation in letter height	High variation signals inconsistent formation
Width Variability	Variation in character width	Reflects control issues or erratic movements

Table 7: Handwriting Features and Their Relevance to Dysgraphia

To assess severity levels, the extracted features were standardized and subjected to unsupervised clustering using KMeans (k=3). The resulting clusters were interpreted as **Mild**, **Moderate**, and **Severe** based on the degree of deviation across the eight feature dimensions. Learners exhibiting greater spatial and motor irregularities were assigned to the Severe cluster, whereas those with relatively fluent and aligned handwriting were placed in the Mild cluster. Table 8 presents a sample of clustered handwriting data.

Learner Image	Severity Level	Avg Contour Area	Stroke Thickness	Spacing (Contours)	Height Inconsistency	Baseline Deviation	Width Variability
image005.jpg	Moderate	22020.38	22.64	4	24.92	35.65	339.91
PD (18).jpg	Severe	7882.71	47.56	49	39.69	44.64	273.97
PD (2).jpg	Mild	431550.00	41.75	1	0.00	0.00	0.00
PD (20).jpg	Mild	197675.00	35.97	1	0.00	0.00	0.00
PD (21).jpg	Severe	56504.33	45.40	3	30.18	59.61	0.00

Table 8: Sample Handwriting Analysis by Severity Cluster

Cluster-wise aggregation of these features yielded meaningful group-level insights, as shown in Table 9.

Severity Level	Avg Contour Area	Stroke Thickness	Spacing (Contours)	Height Inconsistency	Baseline Deviation	Width Variability
Mild	251,698	42.35	1.14	10.29	0.00	2.00
Moderate	22,526	29.81	7.63	30.98	29.27	404.57
Severe	17,756	44.07	32.20	30.72	44.56	213.10

Table 9: Cluster-Wise Average Handwriting Features

Learners in the **Severe** category exhibited marked deviations across nearly all features. Their handwriting was characterized by significantly smaller contour areas, indicating restricted and cramped letter formation. Stroke thickness was highest in this group (44.07 px), suggesting overpressure during writing—an indicator of poor motor fluency or compensatory effort. Fragmentation was also evident, with spacing (32.20 contours) and baseline deviation (44.56 px) values peaking in this cluster, confirming issues in letter segmentation and alignment. These results suggest a high level of motor instability and spatial disorganization, often associated with severe dysgraphia.

The **Moderate** cluster presented intermediate values across most metrics. Although contour area and stroke thickness were improved relative to the Severe group, substantial inconsistency was noted in height (30.98 px) and width (404.57 px) variability, indicating fluctuating control. These learners demonstrated partial regulation of spatial features but remained prone to erratic execution patterns. Their profiles suggest transitional handwriting development and a need for structured motor-guidance interventions.

In contrast, learners in the **Mild** cluster showed fluent and stable handwriting characteristics. They exhibited the largest contour areas (251,698 px²), nearly negligible baseline deviation and height inconsistency, and minimal spacing irregularities. Despite these strengths, stroke thickness remained moderately high (42.35 px), possibly indicating residual motor tension. This cluster reflects learners with adequate motor coordination but who may benefit from biomechanical refinements, such as grip relaxation or ergonomic correction.

Overall, the most salient features distinguishing severity levels were **spacing (contours)** and **baseline deviation**. These indicators robustly differentiated between Mild and Severe dysgraphic profiles, highlighting their diagnostic value. While stroke thickness was elevated across clusters, its interpretive weight increased when contextualized alongside fragmentation and alignment metrics. The findings underscore the utility of quantitative handwriting analysis for stratifying dysgraphia severity, offering educators and clinicians a data-driven approach to tailoring support interventions.

4.3 Visual Scaffolding Techniques Recommendations

This section presents a structured scaffolding framework designed to translate dysgraphia severity diagnostics into actionable pedagogical interventions. Drawing upon the cognitive scaffolding model proposed by Quintana et al. (2004), the framework integrates quantitative handwriting feature analysis with differentiated scaffolding strategies aligned to severity clusters—Mild, Moderate, and Severe.

Step 1: Identifying Learner Challenges Through Feature Thresholds

Using the extracted handwriting features, each learner's challenges were identified based on deviations from cluster-specific thresholds as shown in Table 10. For example, learners with spacing values exceeding 8 mm, baseline deviation above ±4 mm, and excessive stroke thickness (>30 px) were flagged as experiencing significant difficulties in motor execution and spatial organization.

Severity Level	Avg Contour Area	Stroke Thickness	Spacing (Contours)	Height Inconsistency	Baseline Deviation	Width Variability
Moderate	22,526	29.81	7.63	30.98	29.27	404.57
Severe	17,756	44.07	32.20	30.72	44.56	213.10

Table 10: Dysgraphia Severity Levels Based on Feature Thresholds

Learners in the Moderate and Severe clusters exhibited extensive issues in alignment, spatial consistency, and pressure regulation, reflecting both neuromotor instability and cognitive planning deficits. These feature-level impairments form the foundation for targeted intervention strategies.

Step 2: Mapping Cognitive Challenges by Severity Cluster

The handwriting challenges were mapped to cognitive dimensions—**sense-making**, **process management**, and **articulation and reflection**—as outlined by Quintana et al. (2004). This classification enabled tailored interventions for each cluster.

- **Moderate Cluster:** Challenges were primarily related to process management and articulation. Inconsistent spacing, pressure, and size variation indicated deficits in self-monitoring and motor planning. Recommended interventions included structured visual guides and real-time feedback to support reflective practice.
- **Severe Cluster:** Deficits spanned all cognitive dimensions. High variability in spacing and alignment, coupled with erratic stroke production, revealed disrupted sense-making, planning, and self-regulation. Learners in this category required intensive, multi-dimensional support including foundational skill reconstruction, stepwise instructional design, and metacognitive prompting.

Step 3: Scaffolding Types by Cluster

Scaffolding types—**Supportive**, **Intrinsic**, and **Reflective**—were matched to each severity cluster based on the nature and magnitude of feature deviations.

- **Mild Cluster:** Stroke thickness was the only flagged feature. Supportive scaffolding, such as modeling and feedback prompts, was sufficient to enhance grip control and reduce motor tension.
- **Moderate Cluster:** Featured a combination of Supportive and Intrinsic scaffolding. For instance, spacing and stroke thickness required Supportive strategies like spatial grids and guided tracing, while height inconsistency and width variability demanded Intrinsic scaffolding emphasizing spatial awareness and proportional balance.
- **Severe Cluster:** Required all three scaffolding types. Spacing and baseline deviation were managed with Supportive scaffolding (e.g., visual overlays), height inconsistency and contour area through Intrinsic scaffolding (e.g., structured visual zones), and aspect ratio with Reflective scaffolding that encouraged learners to assess and adjust letter proportions independently.

Step 4: Scaffolding Guidelines for Feature-Based Intervention

Each handwriting feature was aligned with a set of scaffolding guidelines derived from Quintana et al.'s framework, as detailed in Table 11 These guidelines provided a direct instructional roadmap to address the specific motor or spatial challenges exhibited by learners.

Severity	Feature	Guidelines	Strategies
Mild	Stroke Thickness	G5	5a. Modeling, 5b. Tracing with fading, 5c. Real-time feedback
Moderate	Spacing (Contours)	G1	1a. Visual grids and spatial organizers
Moderate	Height Inconsistency	G4, G5	4a. Boundaries, 4b. Step hints, 5a–5c. Modeling, tracing, feedback
Moderate	Width Variability	G1, G2	1a. Spatial organizers, 2a. Stroke direction cues, 2b. Visual proportions
Severe	Baseline Deviation	G1, G3	1a. Line guides, 3a. Overlays for alignment feedback
Severe	Aspect Ratio	G1, G9	9a. Reflection tools and comparative reviews

Table 11: Feature-Based Scaffolding Strategies

This alignment ensured each feature's instructional need was matched with an appropriate pedagogical response.

Step 5: Strategy Selection by Cluster Severity

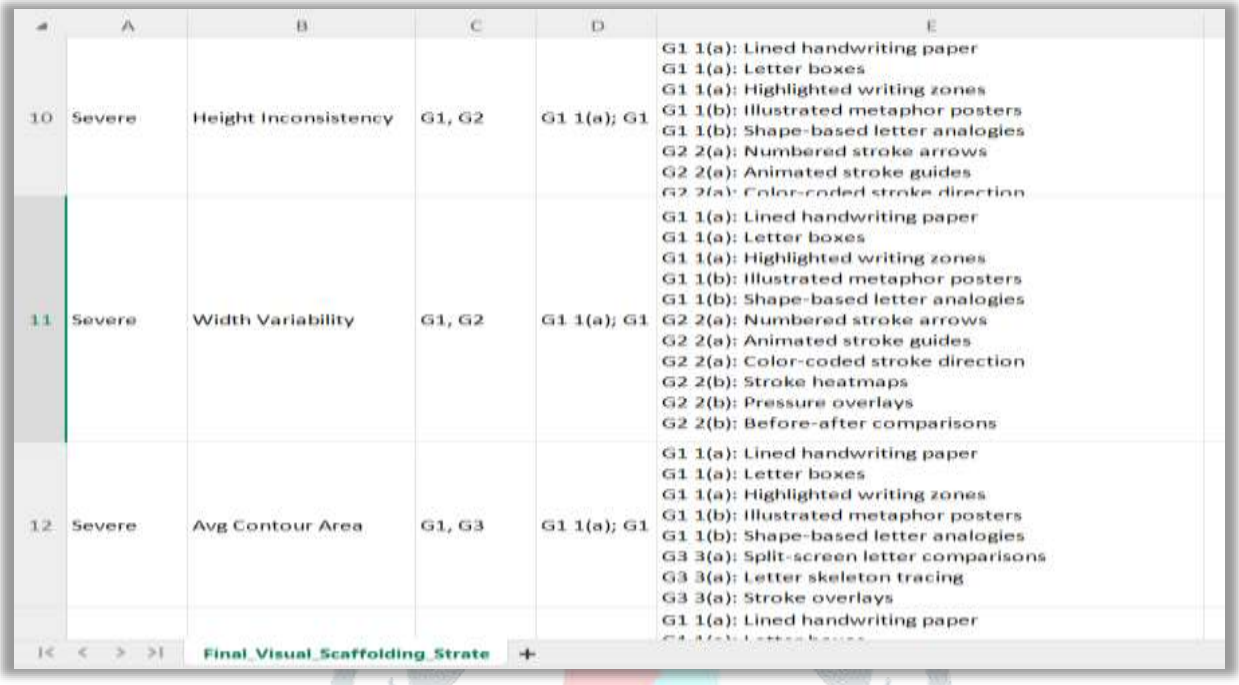
Instructional strategies were selected and deployed according to severity and cognitive alignment (Table 4.25).

- **Mild Cluster:** Emphasized strategy G5 for stroke thickness control, using expert modeling and tracing activities that faded as learner competence increased.
- **Moderate Cluster:** Required a blended approach. Spacing and stroke control employed G1 and G5 strategies, while height and width inconsistencies leveraged stepwise guidance and semantic visual zones through G4 and G2.
- **Severe Cluster:** Utilized layered support, including visual overlays (G3), comparative analysis tools (G9), and progressive task breakdowns (G4). These strategies addressed both motor control and reflective evaluation.

Step 6: Visual Scaffolding Techniques Implementation

Scaffolding intensity and design were adapted to learner severity.

Mild: Employed trace-then-copy worksheets, modeling videos, and fading prompts to reinforce independent writing skills. **Moderate:** Utilized structured visual organizers, stroke sequence cues, and sequential progression to aid spatial control and pressure management. **Severe:** Integrated high-support tools such as live modeling, color-coded feedback, and multi-layered visual rubrics. These tools enabled learners to process spatial errors, recognize inconsistencies, and self-correct over time. A sample of Severe Visual Scaffolding Techniques for Severe Cluster is shown in Figure 7.



	A	B	C	D	E
10	Severe	Height Inconsistency	G1, G2	G1 1(a); G1	G1 1(a): Lined handwriting paper G1 1(a): Letter boxes G1 1(a): Highlighted writing zones G1 1(b): Illustrated metaphor posters G1 1(b): Shape-based letter analogies G2 2(a): Numbered stroke arrows G2 2(a): Animated stroke guides G2 2(a): Color-coded stroke direction
11	Severe	Width Variability	G1, G2	G1 1(a); G1	G1 1(a): Lined handwriting paper G1 1(a): Letter boxes G1 1(a): Highlighted writing zones G1 1(b): Illustrated metaphor posters G1 1(b): Shape-based letter analogies G2 2(a): Numbered stroke arrows G2 2(a): Animated stroke guides G2 2(a): Color-coded stroke direction G2 2(b): Stroke heatmaps G2 2(b): Pressure overlays G2 2(b): Before-after comparisons
12	Severe	Avg Contour Area	G1, G3	G1 1(a); G1	G1 1(a): Lined handwriting paper G1 1(a): Letter boxes G1 1(a): Highlighted writing zones G1 1(b): Illustrated metaphor posters G1 1(b): Shape-based letter analogies G3 3(a): Split-screen letter comparisons G3 3(a): Letter skeleton tracing G3 3(a): Stroke overlays
					G1 1(a): Lined handwriting paper

Final_Visual_Scaffolding_Strate

Fig 7: Screenshot of Visual Scaffolding Techniques Interface

These findings affirm that visual scaffolding tailored to severity clusters effectively bridges cognitive and motor deficits in dysgraphia. The adaptive structure aligns with Vygotsky's Zone of Proximal Development (ZPD), supporting gradual skill acquisition and autonomy. As such, this framework offers a scalable intervention model that can be embedded in digital learning environments or therapeutic programs to support learners with dysgraphia at varying levels of need.

V. DISCUSSION

This study presented a unified framework that integrates deep learning-based dysgraphia prediction, handcrafted handwriting feature analysis, and cognitively grounded scaffolding recommendations to provide a holistic approach for early screening and pedagogical support of learners with handwriting difficulties. The results reveal a cohesive pipeline that not only demonstrates high diagnostic accuracy but also translates technical outputs into actionable educational interventions.

The dysgraphia prediction module, centered on a DenseNet121 + ANN architecture, achieved superior performance compared to baseline machine learning models and peer-reviewed alternatives. With accuracy reaching 92.8% on public datasets and 92.3% in field validation, the model demonstrated both technical efficacy and contextual robustness. Importantly, high AUC-ROC scores (0.965 and 0.938) confirmed the model's excellent discriminative ability across diverse handwriting samples. The ability to generalize without retraining—particularly in real-world settings with unstandardized inputs—positions the model as a practical and scalable tool for preliminary dysgraphia screening, especially in under-resourced educational contexts.

Complementing the classification pipeline, the handwriting analysis component offered deeper insights into the motor and spatial characteristics that underlie dysgraphia severity. By extracting eight visual features—such as contour area, stroke thickness, and baseline deviation—this module provided a quantitative basis for understanding the nature and extent of handwriting impairments. The application of KMeans clustering allowed for the stratification of learners into Mild, Moderate, and Severe groups, facilitating personalized interpretation. Crucially, spacing (contours) and baseline deviation emerged as the most salient differentiators across severity levels, reflecting the critical role of spatial organization in handwriting fluency.

Building upon these severity insights, the visual scaffolding recommendations framework operationalized the findings into pedagogically sound interventions. Grounded in Quintana et al.'s (2004) cognitive scaffolding model, this layer matched specific handwriting features to cognitive challenges—such as sense-making and process management—and mapped them to appropriate scaffolding types: Supportive, Intrinsic, and Reflective. This level of granularity ensured that learners received not only individualized predictions but also tailored educational support aligned with their motor and cognitive needs. For example, mild handwriting difficulties were addressed through minimal support mechanisms such as modeling and guided tracing, while severe cases necessitated multi-layered interventions including visual overlays, real-time feedback, and comparative reflection tools.

The integrated model represents a novel approach in the dysgraphia detection literature, moving beyond isolated predictions to provide a continuum of support—from diagnosis to remediation. Unlike traditional screening tools, which often end at classification, this study demonstrates how machine learning outputs can be translated into scaffolding strategies that are context-aware, pedagogically sound, and scalable. Moreover, by structuring recommendations around the Zone of Proximal Development (ZPD), the framework emphasizes not only identification but also skill-building and learner autonomy.

In summary, the unified pipeline exemplifies the synergy between artificial intelligence, cognitive science, and education. By linking predictive analytics to real-world interventions, the model empowers educators and therapists to respond to dysgraphia challenges with precision and purpose. Future work may explore the real-time implementation of this framework in classroom settings and its integration with digital handwriting tools for ongoing formative assessment.

VI. CONCLUSION AND FUTURE WORK

This study presents a unified, data-driven framework for the early detection and pedagogical support of dysgraphia, combining state-of-the-art deep learning techniques with cognitively grounded educational interventions. By employing DenseNet121 for feature extraction and an Artificial Neural Network (ANN) for classification, the proposed model achieved outstanding predictive performance—attaining 92.8% accuracy and an AUC-ROC of 0.965 on benchmark datasets, and maintaining comparable performance (92.3% accuracy) in real-world educational contexts across Kenyan primary schools. These results not only demonstrate the model's robustness and generalizability but also affirm its practical viability for deployment in diverse and resource-constrained environments.

Beyond high-performance classification, this research contributes a novel post-classification handwriting analysis module that quantitatively evaluates spatial and motor control features of handwriting. Through KMeans clustering, learners were stratified into Mild, Moderate, and Severe dysgraphia categories based on feature deviation severity. This stratification offers clinicians and educators an objective, interpretable foundation for assessing handwriting impairments—bridging the gap between machine perception and human diagnostic understanding.

Critically, the study advances the application of artificial intelligence in inclusive education by embedding machine learning outputs within a theoretically sound scaffolding framework. Drawing on Quintana et al.'s cognitive scaffolding principles, the framework maps each identified handwriting challenge to corresponding cognitive domains—sense-making, process management, and reflection—and prescribes appropriate scaffolding strategies (Supportive, Intrinsic, Reflective). This adaptive design ensures that interventions are tailored not only to the severity of dysgraphia but also to the specific cognitive and motor difficulties encountered by each learner.

Collectively, this research offers a comprehensive, scalable, and context-aware solution to the long-standing challenge of early dysgraphia detection and remediation. It demonstrates how advanced computational tools can be seamlessly integrated with educational theory to inform actionable classroom practice. For regions with limited access to occupational therapists or clinical assessments, this framework provides an urgently needed tool for timely and individualized learner support, aligning technological innovation with educational equity.

This study not only contributes to the fields of educational technology, learning analytics, and digital health but also sets a strong precedent for future research in AI-assisted learning interventions that are both technically robust and pedagogically meaningful.

A key direction for future research is the implementation of longitudinal studies to assess the long-term effects of scaffolded interventions on handwriting development and academic outcomes. Such studies would provide critical insights into how sustained exposure to the recommended visual scaffolding strategies impacts learners' fine motor skills, writing fluency, and classroom engagement over time.

Another important advancement involves integrating real-time handwriting data through digital tablets or smart pen interfaces. These technologies would enable the system to provide immediate, adaptive feedback during writing tasks, enhancing the scaffolding experience by dynamically responding to learner inputs. This feature could significantly improve intervention precision and learner autonomy.

To broaden the model's utility across diverse educational environments, future research should explore the inclusion of multilingual datasets and varied orthographic systems. This would ensure that the framework accommodates different languages and writing structures, thereby enhancing its cultural and linguistic relevance in global contexts.

Additionally, incorporating Explainable AI (XAI) methodologies will be essential for making the model's decision-making process more transparent and interpretable to educators, therapists, and stakeholders. Improving model explainability will foster trust in automated assessments and empower teachers to make informed pedagogical decisions based on system outputs.

Further validation against clinical standards is also needed. Aligning the computational severity classifications with established assessments by occupational therapists will help confirm the model's diagnostic accuracy and facilitate its adoption in clinical and educational settings. Such validation will also refine severity thresholds and intervention guidelines.

From a systems integration perspective, future work should focus on embedding the dysgraphia detection framework into comprehensive classroom support platforms. This includes the development of teacher-facing dashboards, integration with learning management systems, and the provision of training materials to support the adoption of visual scaffolding strategies in everyday teaching practices.

Efforts should also be directed toward ensuring accessibility by developing low-cost, low-tech adaptations of the model for use in resource-constrained environments. These adaptations might include printable scaffolding templates or simplified digital tools that maintain diagnostic value without the need for high-end infrastructure.

Finally, ethical considerations must remain at the forefront of future development. Special attention must be given to ensuring data privacy, avoiding algorithmic bias, and promoting equitable access. Responsible AI principles should guide every stage of system deployment to ensure that all learners—particularly those in vulnerable or marginalized contexts—benefit safely and fairly from the framework.

Acknowledgments

The authors would like to express their sincere gratitude to all those who contributed to the development and completion of this research. Special thanks are extended to the educators and students who participated in the handwriting data collection process, whose involvement was invaluable to the study. The guidance and insights provided by academic mentors and peer reviewers greatly enriched the depth and quality of this work. We also acknowledge the support of our institution for providing the necessary resources and facilities. Lastly, we are grateful for the constructive feedback received from colleagues and collaborators, which significantly enhanced the rigor and clarity of this manuscript.

References

1. Asselborn, T., Zolna, K., Casteran, L., Dillenbourg, P., & Johal, W. (2018). The dynamics of handwriting can help to diagnose dysgraphia. *arXiv preprint*. https://doi.org/10.48550/arXiv.1811.02761
2. Mekyska, J., Faundez-Zanuy, M., Mzourek, Z., Galaz, Z., Smekal, Z., & Rosenblum, S. (2017). Identification and rating of developmental dysgraphia by handwriting analysis. *IEEE Transactions on Human-Machine Systems, 47*(2), 235–248. https://doi.org/10.1109/THMS.2016.2586605
3. Asselborn, T., Johal, W., Zolna, K., & Dillenbourg, P. (2021). Deep-learning for dysgraphia detection in children's handwritings. In *Proceedings of the 2021 ACM Symposium on Eye Tracking Research & Applications* (pp. 1–10). https://doi.org/10.1145/3573128.3609351
4. Bublin, M., Werner, F., Kerschbaumer, A., Korak, G., Geyer, S., Rettinger, L., … Schmid-Kietreiber, M. (2022). Automated dysgraphia detection by deep learning with SensoGrip. *arXiv preprint*. https://doi.org/10.48550/arXiv.2210.07659
5. Drotár, P., & Dobeš, M. (2020). Dysgraphia detection through machine learning. *Scientific Reports, 10*, 21541. https://doi.org/10.1038/s41598-020-78611-9
6. Gupta, R., Bhandari, D., Patil, P. P., & Kulkarni, A. A. (2023). Handwriting analysis AI-based system for assisting people with dysgraphia. In *Proceedings of ICCS 2023* (pp. 173–185). Springer. https://doi.org/10.1007/978-3-031-36024-4_14
7. Vydeki, D., Bhandari, D., Patil, P. P., & Kulkarni, A. A. (2024). Towards accessible learning: Deep learning-based potential dysgraphia detection and OCR for potentially dysgraphic handwriting. *arXiv preprint*. https://doi.org/10.48550/arXiv.2411.13595
8. Mekyska, J., Galaz, Z., Safarova, K., Zvoncak, V., Cunek, L., Urbanek, T., … Faundez-Zanuy, M. (2024). Assessment of developmental dysgraphia utilizing a display tablet. In *Lecture Notes in Computer Science* (Vol. 14285, pp. 21–35). Springer. https://doi.org/10.1007/978-3-031-45461-5_2
9. Devi, A., & Kavya, G. (2022). Dysgraphia disorder forecasting and classification technique using Kekre-discrete cosine transform and deep transfer learning. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 120, 110647. https://doi.org/10.1016/j.pnpbp.2022.110647
10. Lu, Y., Quinton, J.-C., & Brault, V. (2024). A statistical procedure to assist dysgraphia detection through dynamic modelling of handwriting. *arXiv preprint*. https://doi.org/10.48550/arXiv.2408.02099