Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 3 Issue 1 June, 2026, pp. 33-40
https://idjams.com/index.php/jms/index

Early Dyslexia Detection Using Deep Learning: Classifying Children's Handwriting with Convolutional Neural Networks

Muhammad Imamul Caesar[1,*], Aditya Prihandhika[2], Jasem Al Tamar[3]

[1] *Universitas Padjadjaran, Jl. Soekarno Km 21 Jatinangor, Kab. Sumedang, West Java 45363, Indonesia*
[2] *Universitas Singaperbangsa Karawang, Jl. H. S. Ronggo Waluyo, Karawang, West Java 41361, Indonesia.*
[3] *College of Education, Kuwait University. Sabah Al Salem University City, P.O. Box 5969, Safat – 13060, Kuwait*
[*] *Corresponding author: imamulcaesar123@gmail.com*

Copyright: ©2026 by Indonesian Journal of Applied Mathematics and Statistics. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License-(CC-BY-SA) (https://creativecommons.org/licenses/by-sa/4.0/

Article Info
Article history:
Received: December 1, 2025
Revised: December 16, 2025
Accepted: December 17, 2025
Available online: June 1, 2026

DOI:
https://doi.org/10.71385/idjams.v3i1.34

Abstract

This study investigates the development of an automated handwriting-based dyslexia classification model using a Convolutional Neural Network (CNN). The dataset comprises scanned handwriting samples from children diagnosed with dyslexia and those without learning difficulties. Prior to model training, the images were resized to a uniform dimension, converted to grayscale, and normalized to standardize pixel intensity values. Data augmentation techniques, including rotation, scaling, and horizontal shifting, were applied to increase data diversity and reduce overfitting. A lightweight CNN architecture was then employed to perform binary classification between dyslexic and non-dyslexic handwriting samples. Experimental results indicate that the proposed model achieved an accuracy of 51%, with a precision of 0.51 and a recall of 0.42, suggesting that its current predictive performance remains limited. These findings highlight the challenges of dyslexia classification using handwriting features alone, particularly when constrained by model simplicity and data resolution. Nevertheless, this study serves as an exploratory step toward automated dyslexia screening and provides insights for future work, where performance may be improved through the use of deeper network architectures, such as ResNet-18, and higher-resolution handwriting representations.

Keywords: Dyslexia Detection; Handwriting Classification; Deep Learning; Convolutional Neural Networks

I. Introduction

Dyslexia is a common learning disorder experienced by children, primarily affecting their ability to read, recognize letters, and interpret written symbols. Importantly, dyslexia is not associated with low intellectual ability, but rather with differences in how the brain processes phonological and visual information [1]. This distinction underscores that children with dyslexia are capable learners who require instructional approaches tailored to their neurological profiles. In Indonesia, public and academic discourse surrounding dyslexia has gradually gained attention; however, awareness among educators and parents regarding its early indicators remains limited. Several reports estimate that the prevalence of dyslexia among Indonesian school-aged children ranges between 5–10%, which is consistent with global estimates [2]. Despite this prevalence, comprehensive national data are still scarce, suggesting that early identification has not yet become a systematic priority within the education system. As a result, many children are diagnosed only after experiencing prolonged academic difficulties.

Early detection plays a critical role in minimizing the negative consequences of delayed diagnosis, including reduced academic confidence, learning disengagement, and psychological distress. This priority aligns closely with Sustainable Development Goal (SDG) 4, which emphasizes inclusive and equitable quality education for all learners [3]. Accurate and timely screening enables earlier intervention, thereby improving long-term educational outcomes. Consequently, there is a growing need for technological innovations that can support early dyslexia screening in a scalable and accessible manner. Handwriting characteristics have been widely recognized as potential early indicators in dyslexia assessment. Previous studies report that children with dyslexia often exhibit inconsistent handwriting patterns, such as unstable letter sizing, irregular spacing, and difficulties in maintaining letter shape consistency [4]. Since handwriting is a routine activity in school environments, it provides a practical and non-invasive data source without requiring clinical or medical diagnostic procedures.

Several studies have attempted to automate dyslexia detection using handwriting-related data. Zargar et al. [5]developed a classification framework based on geometric feature extraction combined with traditional machine learning algorithms. Other studies have explored dynamic handwriting attributes, including pen trajectory, writing speed, or stylus pressure captured through digital writing devices [6]. While these approaches demonstrate promising directions, many rely on handcrafted features or specialized hardware, which may limit their adaptability and scalability in real-world educational settings. In recent years, deep learning particularly Convolutional Neural Networks (CNNs) has emerged as a dominant approach in image analysis. CNNs are capable of autonomously learning hierarchical visual representations and have demonstrated strong performance in handwriting recognition tasks, ranging from benchmark datasets such as MNIST to more complex handwriting analysis applications [7]. Arora et al. [8] further showed that CNN-based models can identify dyslexia-related handwriting traits with encouraging classification results. CNNs are particularly well-suited to this domain, as handwriting differences associated with dyslexia are often subtle and require robust representational learning.

Despite these advances, a clear research gap remains. Existing studies on handwriting-based dyslexia detection frequently focus on achieving high classification accuracy using complex architectures or specialized data acquisition tools, while comparatively less attention is given to exploratory evaluations of lightweight CNN models trained on standard image-based handwriting data. Moreover, many studies report strong performance without sufficiently discussing the limitations encountered during early-stage model development, particularly in contexts with constrained data quality or resolution [7]. In the Indonesian context, research on automated dyslexia screening using handwriting images is still limited, leaving open questions regarding the feasibility, challenges, and practical implications of deploying such models in real educational environments. Addressing this gap is important to better understand both the potential and the limitations of CNN-based approaches for early dyslexia screening.

Motivated by these considerations, the present study proposes a CNN-based framework for classifying handwriting samples from dyslexic and non-dyslexic children using image-based inputs. Rather than focusing solely on performance optimization, this study emphasizes an exploratory evaluation of CNN capability in capturing visually meaningful handwriting patterns related to dyslexia. This perspective is particularly relevant for early screening initiatives, as image-based models can be more easily implemented in schools or educational psychology services without requiring specialized equipment. By situating the findings within their methodological constraints, this study aims to contribute practical insights that may inform future improvements in automated dyslexia detection systems and support inclusive education initiatives in Indonesia.

The primary aim of this study is to explore the feasibility of using a Convolutional Neural Network (CNN) to classify handwriting samples from children with and without dyslexia based on image-based features. Specifically, the objectives of this study are as follows:

1. To develop a lightweight CNN architecture for binary classification of dyslexic and non-dyslexic handwriting images.
2. To evaluate the classification performance of the proposed model using standard metrics, including accuracy, precision, and recall.

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 3 Issue 1 June, 2026, pp. 33-40
https://idjams.com/index.php/jms/index

3. To examine the extent to which CNNs can capture visual handwriting patterns that may serve as early indicators of dyslexia.
4. To identify methodological limitations and challenges that arise in early-stage handwriting-based dyslexia classification, thereby providing direction for future research improvements.

II. Method

This study uses a publicly available dataset consisting of 4,719 handwriting images from children with dyslexia and 3,883 handwriting images from non-dyslexic children, resulting in a total of 8,602 images. The size of the dataset allows the model to learn visual writing patterns, including orientation errors and character distortions that are commonly observed in individuals with dyslexia. Given the scale of the dataset, a stratified split was applied to divide the data into training, validation, and testing subsets, ensuring balanced class proportions across all phases of model evaluation. By utilizing a verified and annotated dataset, this study ensures that the CNN model is trained using handwriting samples that exhibit representative variations in letter shape and orientation. This approach allows the classification output to better reflect the model's ability to identify handwriting characteristics typically associated with dyslexia.

The preprocessing stage was carried out to ensure that the images were in a format compatible with model training. Each handwriting sample was converted into grayscale so that the extracted information focused solely on shape, contour, and writing patterns rather than color variations. This step aligns with standard practices in handwriting recognition research, such as studies using the MNIST or IAM datasets [9]. Next, the images were resized to 240 × 240 pixels to maintain uniform dimensions across samples. Pixel values were then normalized to a range of [0, 1] using the following equation:

$$x' = \frac{x}{255} \tag{1}$$

Normalization helps accelerate convergence during CNN training by ensuring that the input values operate within a controlled numerical range [3].

To increase data variability, data augmentation techniques were applied, including small rotation (±10°), shearing, minor brightness adjustments, and horizontal shifting. These augmentation strategies are considered essential, as handwriting produced by children naturally varies widely in direction, spacing, and form.

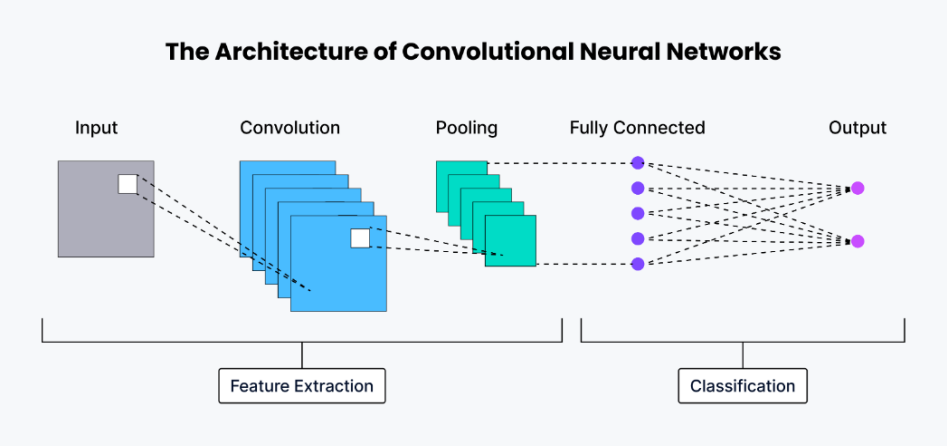


Figure 1. CNN Model Architecture

Figure 2.1 illustrates the overall workflow of the Convolutional Neural Network (CNN) model from input to final classification. Conceptually, the architecture consists of two major phases: feature extraction and classification. The model first receives the handwriting image as input, which is then processed through a convolution layer equipped with small filters designed to scan the image [10].

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 3 Issue 1 June, 2026, pp. 33-40
https://idjams.com/index.php/jms/index

During this stage, the CNN begins capturing basic structural patterns such as edges, curves, and texture details. The output of the convolution layer is fed into a pooling layer, which reduces spatial size while retaining essential information. This pooling step helps improve computational efficiency and increases model robustness against minor variations such as slanted writing or inconsistent spacing between letters.

After the extracted feature maps are compressed, they are flattened and passed into the fully connected layer. The fully connected layer, represented by dense units, is responsible for synthesizing previously extracted feature information to determine the final class label. Overall, the CNN functions as a multi-stage filtering process, beginning with recognition of low-level visual cues and progressing to higher-level writing structures. This layered mechanism makes CNNs well-suited for image-based handwriting analysis, especially when identifying subtle stroke patterns associated with dyslexia [11]. The model architecture developed includes the following components:

1. A convolution layer with a 3×3 kernel for extracting visual features such as letter edges and stroke width differences.
   ReLU activation function, defined as:

$$f(x) = max(0, x) \tag{2}$$

2. A 2×2 max-pooling layer for reducing feature dimensions without eliminating key information.
   Dropout (0.3–0.5) to prevent overfitting by randomly disabling neurons during training.
3. A flattening layer followed by a dense layer for final classification, using a sigmoid activation function due to the binary nature of the task (dyslexia = 1, non-dyslexia = 0).

Model training was performed using the Adam optimizer, which is widely recognized for its fast convergence and stable adaptive weight updates [12]. The initial learning rate was set to 0.001, with a batch size of 32. The training process was conducted for 20–30 epochs, while validation loss was closely monitored. If validation loss increased while training loss continued to decrease, early stopping was applied to prevent overfitting. The loss function used was Binary Cross-Entropy, defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \, log \, log \, (\hat{y}_i) + (1 - y_i) \, log(1 - \hat{y}_i)] \tag{3}$$

This loss function is appropriate for binary classification tasks requiring accurate and stable probability outputs.

Model performance was evaluated using several standard binary classification metrics, including:

1. Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

   Accuracy reflects the proportion of correctly predicted instances across the testing dataset. While intuitive, accuracy alone may be misleading when class distributions are imbalanced, as a model may appear effective even if it predicts only the majority class.
2. Precision

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

   Precision assesses the correctness of positive predictions (dyslexia predictions). This metric is important when minimizing false positives is necessary, for example to avoid unnecessary concern among parents.
3. Recall

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 3 Issue 1 June, 2026, pp. 33-40
https://idjams.com/index.php/jms/index

Recall measures the model's success in identifying all actual dyslexia cases. In early screening, missing a dyslexic child (false negative) is more detrimental than mistakenly flagging a non-dyslexic sample.

4. F1-Score

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{7}$$

The F1-score provides a balanced evaluation when both precision and recall are critical indicators.

In addition to these metrics, a confusion matrix was used to analyze misclassification patterns between classes. This evaluation framework follows best practices commonly used in modern image-based classification research [13].

III. Results and Discussion

The dataset used in this study consists of 8,602 handwriting images categorized into dyslexia and non-dyslexia classes. More specifically, 3,883 images belong to the non-dyslexia category, while 4,719 images belong to the dyslexia category. The relatively balanced class distribution gives the model a fair opportunity to learn the structural differences between both handwriting types, although handwriting produced by children with dyslexia generally exhibits greater variability in structure and form. The dataset was divided into training (70%), validation (15%), and testing (15%), following common practices in CNN-based classification research [14]. This split supports a stable model training process while enabling objective model evaluation on unseen data. Based on recommendations from Goodfellow et al. [14], such a partitioning strategy helps prevent the model from memorizing training patterns and promotes its ability to generalize across new handwriting samples. In this study, the data split resulted in a proportional structure, although the inherent complexity of children's handwriting indicates that the model may require a deeper feature extraction strategy to achieve improved classification performance.

To address potential overfitting and enhance variation in visual input, data augmentation was applied using the *ImageDataGenerator* preprocessing tool. The augmentation configuration included rescaling (1/255), small rotation range, zoom range, shear transformation, width shift, and height shift. These techniques are widely recognized as standard preprocessing procedures in image-based deep learning and have proven beneficial in increasing model generalization capability [15]. However, despite appropriate augmentation, the input size used was 240 × 240 pixels, which may still be insufficient for handwriting recognition tasks. Research indicates that successful analysis of handwriting-based patterns especially those involving micro visual characteristics such as subtle strokes, texture variations, and writing pressure representation often benefits from higher resolution inputs [16]. In this study, the use of relatively low-resolution images likely resulted in the loss of important visual cues, particularly in fine motor irregularities often reflected in dyslexic handwriting. This limitation may explain why the model struggled to achieve stable validation performance.

The implemented CNN model follows a relatively simple architecture consisting of sequential layers: Conv → Conv → MaxPooling → Dropout → Dense → Sigmoid. Theoretically, such a baseline architecture may be sufficient for learning simple spatial patterns, such as numeric characters in MNIST or objects with clear structural boundaries [17]. However, handwriting produced by children especially those with dyslexia is characterized by complex and irregular structural patterns. These include variations in stroke pressure, disrupted letter boundaries, varying proportions, and inconsistent spatial alignment. Previous studies emphasize that as visual pattern complexity increases, model depth and representational capacity must also increase to capture hierarchical variations in feature space [18]. Accordingly, the relatively shallow architecture used in this study likely contributed to limited model performance. Although dropout reduced overfitting risk, the underlying architecture still lacked the capacity to learn robust representations. This limitation corresponds with the observation that the model did not demonstrate a meaningful improvement between training and validation accuracy, suggesting that the architecture's learning capability was constrained.

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 3 Issue 1 June, 2026, pp. 33-40
https://idjams.com/index.php/jms/index

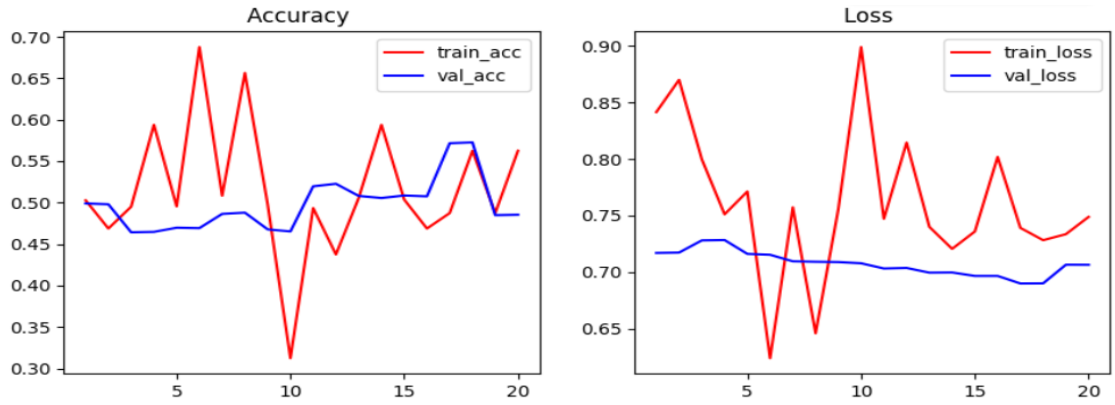


Figure 2. First-Stage Model Training Result

During the first training phase (20 epochs), the model was trained using MobileNetV2 in a frozen state, meaning only the classifier layers were updated. Throughout this stage, model performance showed variability, with accuracy fluctuating between approximately 0.30 and 0.65. This instability suggests that the model struggled to identify consistent visual patterns that distinguish dyslexic handwriting. Meanwhile, validation accuracy remained relatively flat, between 0.48 and 0.52, showing no improvement despite fluctuations in training accuracy. When training accuracy changes significantly while validation accuracy remains stagnant, the model is typically described as underfitting, meaning it has not yet captured meaningful representations required for accurate prediction. A similar pattern was observed in the loss curve, with training loss showing steep movement while validation loss remained relatively stable between 0.69 and 0.73. The mismatch between the two curves reinforces that initial training did not sufficiently improve feature representation.

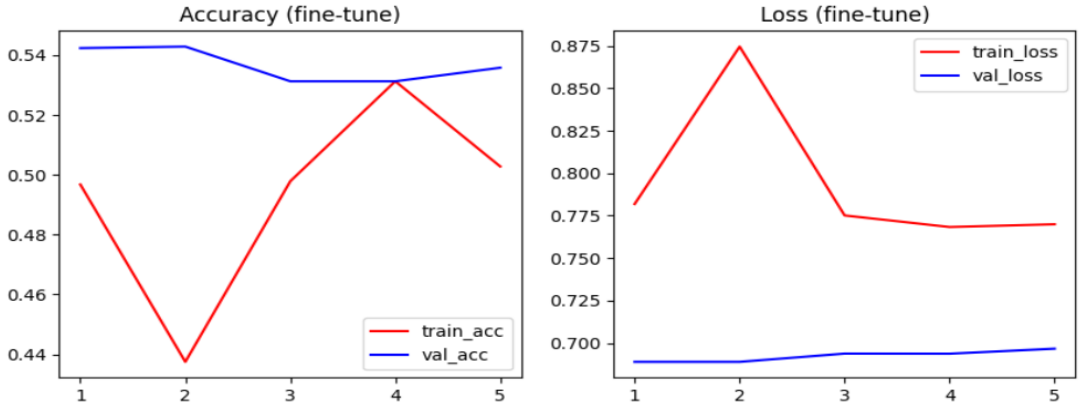


Figure 3. Second-Stage Model Training Result

In the second stage, fine-tuning was performed by unfreezing MobileNetV2 from layer 100 onward. While fine-tuning generally increases specialization to domain-specific patterns, in this case validation accuracy plateaued around 0.53, showing only minimal improvement. Training accuracy briefly decreased before stabilizing, indicating learning difficulty rather than overfitting. Validation loss remained higher than training loss, yet the gap did not meaningfully change, confirming that model representational learning remained limited. These results suggest that deeper network adaptation alone was insufficient. The model likely requires enhanced feature extraction strategies, higher-resolution samples, or a more specialized architecture designed for handwriting variability.

Testing was conducted using an equal number of samples from both classes: 457 non-dyslexia images and 457 dyslexia images. The confusion matrix reveals that the model correctly classified 274

non-dyslexia samples, while misclassifying 183 samples. Conversely, only 190 dyslexia samples were correctly detected, with 267 misclassified.

Table 1. Classification Report

Category	Precision	Recall	F1-Score
Disleksia	0,51	0,42	0,46
Non Disleksia	0,51	0,60	0,55

As shown in Table 1, the proposed CNN model achieved an overall classification accuracy of 0.51, which is only marginally higher than the baseline performance expected from random guessing in a balanced binary classification problem. This comparison indicates that the model, in its current form, has limited capability to effectively distinguish between dyslexic and non-dyslexic handwriting samples. Precision values of 0.51 for both classes further suggest that approximately half of the predicted labels are correct, reflecting insufficient discriminative power in the learned visual representations.

A more detailed inspection of class wise recall highlights a critical imbalance relevant to early screening applications. The recall for the dyslexia class is 0.42, indicating that a substantial proportion of true dyslexia cases were not detected, whereas the non-dyslexia class achieved a higher recall of 0.60. This discrepancy implies a bias toward correctly identifying non-dyslexic handwriting while under-detecting dyslexic patterns. The resulting F1-score of 0.46 for the dyslexia class summarizes this limitation and underscores that, when considered alongside the near baseline accuracy of 51%, the model currently functions as an exploratory benchmark rather than a reliable screening tool, emphasizing the need for further methodological refinement.

Low performance may be attributed to the architectural simplicity and resolution constraints. As highlighted in prior research, handwriting based dyslexia detection requires models capable of learning high resolution micro patterns that represent subtle motor based writing irregularities [19] [20]. Despite these limitations, the findings of this study are still valuable. First, the results demonstrate baseline performance for CNN based handwriting classification within this dataset context. Second, the outcomes highlight the need for deeper architectures such as ResNet-18, transformer based vision models, or more sophisticated preprocessing workflows to improve handwriting based dyslexia detection.

IV. Conclusion

This study developed a CNN-based model to classify handwriting from dyslexic and non-dyslexic This study examined the feasibility of using a Convolutional Neural Network (CNN) to classify handwriting images from dyslexic and non-dyslexic children. The findings explicitly show that, under the current experimental setup, the proposed model achieved an accuracy of approximately 51%, with low recall (0.42) and F1-score (0.46) for the dyslexia class. These results indicate that the implemented lightweight CNN architecture was not yet capable of reliably capturing the complex and highly variable handwriting characteristics associated with dyslexia. Consequently, the research question regarding the effectiveness of CNN-based handwriting classification for early dyslexia screening is answered with important limitations: while CNNs are conceptually suitable for this task, the present model configuration does not yet provide sufficient predictive reliability, particularly in minimizing false-negative cases.

Based on these findings, several actionable directions for future work can be identified. First, deeper and more expressive architectures, such as ResNet-18 or other residual based networks, should be explored to enhance representational capacity. Second, the use of higher resolution handwriting images is recommended to preserve fine grained visual cues related to stroke consistency and spatial irregularities. Third, preprocessing strategies should move beyond generic augmentation by incorporating handwriting specific enhancements, such as stroke or contour emphasis. Together, these improvements may enable image-based deep learning models to more effectively support early and practical dyslexia screening, particularly within inclusive educational settings.

References

[1] S. E. Shaywitz and B. A. Shaywitz, "Dyslexia (Reading Disability)," *N. Engl. J. Med.*, vol. 383, no. 6, pp. 522–532, 2020, DOI: 10.1056/NEJMra1910319.
[2] R. Sugiarti, "Early Identification of Dyslexia in Elementary School Children," *J. Psikologi Pendidik. Perkembangan*, vol. 7, no. 2, 95–103, 2018.
[3] United Nations, "Transforming our world: The 2030 Agenda for Sustainable Development," Department of Economic and Social Affairs Sustainable Development, 2015. [Online]. Available: https://sdgs.un.org/goals [Accesed: November 17, 2025]
[4] T. Asselborn, C. Zablit, M. Chapatte, and P. Dillenbourg, "Automated Dysgraphia Detection using Online Handwriting Signals," *Nature Sci. Rep.*, vol. 8, pp. 1–11, 2018, DOI: 10.1038/s41598-018-25506-1.
[5] S. T. Zargar, A. Rasooli, and H. Ghassemian, "Early Detection of Dyslexia using Handwriting Features and Machine Learning Techniques," *Biomed. Eng.: Appl. Basis Commun.*, vol. 32, no. 6, p. 2050030, 2020, DOI:10.4015/S1016237220500303
[6] P. Drotár, J. Mekyska, and Z. Smékal, "Automatic Evaluation of Handwriting Difficulties in Children with Dysgraphia," *IEEE Trans. Hum.-Mach. Syst.*, vol. 46, no. 1, pp. 123–129, 2016.
[7] Y. LeCun, Y. Bengio, and G. Hinton, (2015). "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, DOI: 10.1038/nature14539.
[8] A. Arora, S. Sharma, and A. Jaiswal, "Dyslexia Detection using Convolutional Neural Networks on Handwriting Images," *Int. J. Ad. Comput. Sci. Appl.*, vol. 13, no. 5, pp. 155–162, 2022.
[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based Learning Applied to Document Recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 2018.
[10] C. M. Bishop, "*Pattern Recognition and Machine Learning*." New York: Springer, 2006.
[11] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.
[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning, *Nature*, vol. 521, pp. 436–444, 2015.
[13] M. Sokolova and G. Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.
[14] H. He and E. Garcia, "Learning from Imbalanced Data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.
[15] C. Shorten and T. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no.60, pp. 1-48, 2019, DOI: 10.1186/s40537-019-0197-0.
[16] R. Plamondon and S.N. Srihari, "On-line and Off-line Handwriting Recognition: A Comprehensive Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 63-84, 2000.
[17] I. Goodfellow, Y. Bengio, and A. Courville, "*Deep Learning*." Cambridge: MIT Press, 2016.
[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE Explore,* pp. 770-778, 2016.
[19] M. Sokolova and G. Lapalme, "Performance Measures for Classification Tasks," *Inf. Process. & Manag.*, vol. 45, no. 4, pp. 427–437, 2009.
[20] Howard, A., et al. "MobileNet and ResNet Architectures for Efficient Vision Learning," *arXiv preprint arXiv: 1704.04861*, 2017, DOI: 10.48550/arXiv.1704.04861.