2024 19th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP) | 979-8-3315-0991-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/iSAI-NLP64410.2024.10799245

Handwriting Image Classification for Automated Diagnosis of Learning Disabilities: A Review on Deep Learning Models and Future Directions

Safura Adeela Sukiman
College of Computing, Informatics, and Mathematics
Universiti Teknologi MARA (UiTM) Johor Branch,
Segamat Campus
Johor, Malaysia
safur185@uitm.edu.my

Nor Azura Husin
Department of Computer Science
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia (UPM)
Selangor, Malaysia
n_azura@upm.edu.my

Hazlina Hamdan
Department of Computer Science
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia (UPM)
Selangor, Malaysia
hazlina@upm.edu.my

Masrah Azrifah Azmi Murad
Department of Computer Science
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia (UPM)
Selangor, Malaysia
masrah@upm.edu.my

***Abstract*—This study reviews deep learning models used in handwriting image classification for the automated diagnosis of learning disabilities. By addressing handwriting diversity and misclassification challenges, two models were highlighted: Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). Literature was retrieved from major databases including IEEE Xplore, Scopus, Web of Science (WoS), and Google Scholar, with studies on Parkinson's disease, tremor patients, and machine learning excluded. CNNs represent a more mature architecture focusing on convolutions, pooling, and activation function. Meanwhile, ViTs emerges as a promising alternative via its multi-head attention architecture. This review also compares the accuracy of both models, specifying the sources of handwriting images, as well as providing future directions relevant to the research field.**

Keywords—automated diagnosis, deep learning, handwriting classification, learning disabilities

I. INTRODUCTION

The term "learning disabilities" encompasses a spectrum of neurologically-based disorders that affect learning, with varying degrees of severity, including mild, moderate, and severe. The manifestation of learning disabilities, particularly among dyslexic and dysgraphia students, has been observed in the form of clumsiness or difficulties with handwriting. The Malaysia Ministry of Education's Dyslexia Checklist Instrument has identified a number of specific handwriting difficulty characteristics that are commonly observed among dyslexic students. We further segmented each of the handwriting difficulty characteristics based on the sentence level as shown in Table I.

TABLE I. HANDWRITING DIFFICULTY CHARACTERISTICS AMONG DYSLEXIC STUDENTS SEGMENTED BASED ON THE SENTENCE LEVEL

Sentence Level	Handwriting Difficulty Characteristics
Line level	• Students write without following the lines. • Insufficient / over-sufficient / no space at all from one word to another. • Non-aligned left margin.
Word level	• Students mix uppercase and lowercase letters. • Broken links between letters in a word. • Students write backwards some of the letters leading to incorrect spelling of the words.
Character / Letter level	• The shape of the letters is not clear. • The size of the letters written are irregular.
Writing speed	• Students are unable to completely copy the information written on the blackboard/whiteboard. • Students are unable to write down the information heard or mentioned by the teacher. • Students cannot compete with their peers in writing ability.

On the other hand, handwriting speed and legibility, inconsistency between spelling ability and verbal intelligence quotient, and pencil grasp have been identified as handwriting difficulty contributing characteristics among dysgraphia students [1]. Also included in Reference [1] are two (2) handwriting samples of students with dysgraphia as shown in Fig. 1.

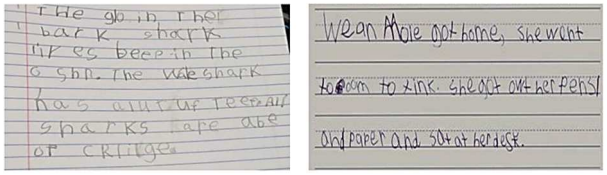

Fig. 1. Handwriting Samples of Dysgraphia Students Specified in Reference [1]

With the advancement of deep learning, traditional score-based assessments (by examining the disparity between IQ scores and standardised achievement tests, i.e., reading, writing, and arithmetic) are being replaced by an automated diagnosis of learning disabilities based on handwriting features. The deep learning enables the deployment of automated diagnostics because of its ability to learn from data and conduct computations using multi-layer neural networks and processing. The term "deep" alludes to the concept of multiple levels or stages through which data is processed during the construction of a data-driven model [2].

The core challenges in the classification of handwriting images among learning disabilities pertains to the diversity of patterns and inaccuracies in classification. As a result, the

979-8-3315-0991-0/24/$31.00 ©2024 IEEE

deployment of a deep learning model is of the utmost importance to researchers. The overall contribution of this article is summarized as follows:

- We investigate how well deep learning models involving Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) have accommodated handwriting image classification for automated learning disability diagnosis. In addition to examining the architectural enhancements, benefits, and limitations of each model, we also examine its accuracy performance.
- We point out and discuss three (3) potential aspects with research directions for future improvement of the current deep learning-based models.

The remaining article is organized as follows: Section II. Methods, Section III. Results, Section IV Discussions, and Section V. Future Directions and Concluding Remark. The Discussions section explores how the CNNs and ViTs model are used in handwriting image classification, along with a succinct analysis of their respective performance. Meanwhile, the Future Directions and Concluding Remark section outlines promising areas of exploration within the scope of our review.

II. METHODS

A. Inclusion Criteria

This review utilized three distinct online databases: IEEE Xplore, Scopus, and Web of Science (WoS), to retrieve articles from prior research endeavors. Additionally, Google Scholar was used to capture relevant publications not indexed by these databases, ensuring a comprehensive review. The search was conducted using keywords related to population participants, intervention, and comparison controls. All retrieved articles were initially screened based on title and abstract, and relevant studies were included for full-text review.

B. Exclusion Criteria

Exclusion criteria were strictly defined to focus on learning disabilities among school students and exclude other conditions that may affect handwriting. Studies involving patients with Parkinson's disease or tremors were excluded, as these conditions typically occur in older populations (ages 40-60), beyond the scope of learning disabilities in educational settings. Furthermore, this review excluded literature on the adoption of machine learning models, as the study focused solely on the use of deep learning models, which offer enhanced precision and reduced dependence on manual feature engineering. Only studies utilizing deep learning for handwriting image classification were considered.

III. RESULTS

A total of 113 search results were initially retrieved. After applying the inclusion and exclusion criteria, 17 peer-reviewed publications were selected for final review. These studies were classified based on their target learning disabilities and methods as shown in Fig. 2. Students with dysgraphia have the largest population, while those with dyslexia have the second largest. Previous studies has focused on these two learning disabilities because they are both neurological conditions that are often connected. Although dyslexia are frequently associated with a reading-specific learning issue, it is also characterized by poor writing and spelling. Dysgraphia, on the other hand, is frequently associated with a learning disability related to writing, as seen by frequent erasing, irregular letter and word spacing, as well as poor spelling. Dyslexia-dysgraphia refers to a person that have difficulty with the act of writing and reading at the same time.

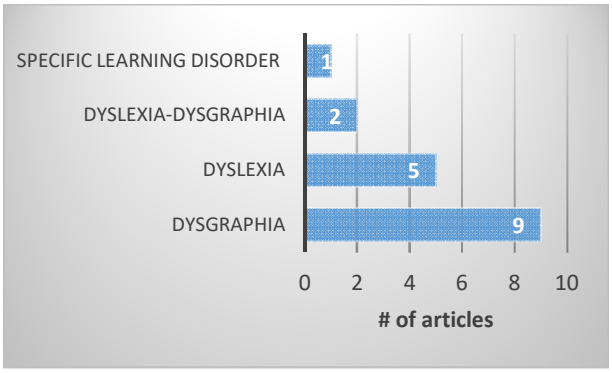

Fig. 2. Articles Retrieved According to its Type of Learning Disabilities

Additionally, the publication timeline spans from 2016 to 2024 with significant increase in publications published from 2021 onwards after realizing the need of automated diagnosis of learning disabilities utilizing handwriting images rather than traditional score-based evaluations, which obviously requires more time and specialized manpower.

IV. DISCUSSIONS

The process of automating the diagnosis of learning disabilities through handwriting image classification can be delineated into six (6) discrete stages, as depicted in Fig. 3.

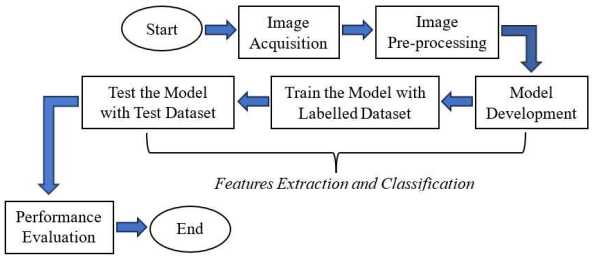

Fig. 3. The General Workflow of Handwriting Image Classification for Automated Diagnosis of Learning Disabilities

Image acquisition of handwriting is comparable to data collection. Two options are available: the publicly available dataset from the Kaggle database, or involves collecting handwriting samples on-site from both students with and without learning disabilities. Following the handwriting image acquisition is the pre-processing step. The typical pre-processing involves rotating and resizing the input handwriting images, as well as exchanging the foreground and background [5, 6].

The development, training, and testing of models are intricately interconnected processes that enhance the effectiveness of deep learning. In the process of model development, deep learning is incorporated, while during the training phase, the labelled dataset is exhaustively leveraged to discern the underlying patterns of the handwriting images and to mitigate the occurrence of misclassification. After training, the model is tested with a new set of images, known as the test dataset. The trained model evaluates these unseen handwriting images, and its performance is systematically assessed, providing critical insights into its ability to generalize and accurately classify data in real-world scenarios.

A. Convolutional Neural Networks

The convolutional neural network (CNN) is a prominent and extensively employed deep learning model within the field of computer vision and image processing. The classic CNN architecture comprises of convolutional layers that incorporate activation functions, succeeded by a pooling layer. This process is iterated for several layers. Subsequently, the final layer involves a fully connected dense layer with a SoftMax activation function [3]. Fig. 4. depicts the architecture of CNN's classic model.

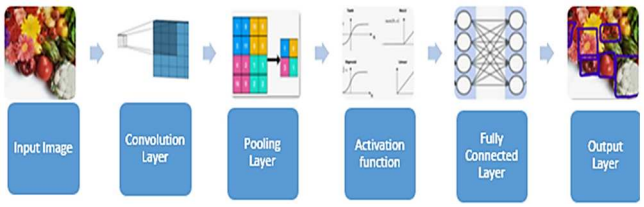

Fig. 4. Classic Architecture of the CNN Model

Using the classic CNN architecture as a foundation, previous studies have introduced several architectural enhancements to further improve the extraction of handwritten image features and reduce misclassification rates. These enhancements build upon the original design to better adapt CNNs to the complexities of handwriting analysis. The key improvements are as follows:

- adding more convolutional and pooling layers to increase the network's depth and capacity for learning intricate feature hierarchies, allowing for improved representation of complex handwriting patterns [4-6].
- replacement of the Rectified Linear Unit (ReLU) activation function with alternative functions to enhance gradient flow and mitigate issues like dying neurons, leading to more robust feature extraction [7].
- applying the 'transfer-learning via fine-tuning' where a pre-trained CNN is adapted to the task of handwriting analysis, significantly reducing training time and improving generalization across different handwriting styles [8-13].
- utilization of region proposal network for text areas to focus the model's attention on text regions, ensuring more accurate localization and classification of handwritten characters [14].

Adding More Convolutional and Pooling Layers

The first architectural enhancement pertains to the incorporation of additional convolutional and pooling layers within the CNN model. Both layers constitute as the feature extractors for the input image. As more convolutional layers are added, the architecture becomes hierarchical with higher convolutional layers extract more abstract, higher-level features (patterns) of the input image. The final output is commonly referred to as a feature map. Pooling layers are often incorporated after convolutional layers in order to reduce the spatial dimensions of the feature map and remove extraneous spatial information [15].

Reference [5] and [6] utilized 3-layer and 5-layer convolutional blocks, respectively. While adding more convolutional blocks can improve the CNN model's performance by extracting more image features (patterns) and making an image more interpretable, it can also cause the gradient to be too small for successful training.

CNNs are closely related to tuning of its parameters such as filter size, learning rate, and optimizers as they can influence the model's ability to learn and generalize from the data. Researchers often conduct extensive parameter tuning experiments to find the best combination that maximizes model accuracy and generalization to unseen data. The best parameters selected by Reference [4] and [6] are shown in Table II.

TABLE II. PARAMETERS SETTINGS IN THE CNN MODEL PERFORMED BY REFERENCE [4] AND [6]

Parameters	Values	
	Reference [4]	Reference [6]
Optimizer	Stochastic Gradient Descent with Momentum (SGDM)	Stochastic Gradient Descent with Momentum (SGDM)
Learning rate	0.01	0.001
Epochs	12	8
Iteration each epoch	Not stated	1251
Frequency		30 iterations

Replacement of the Rectified Linear Unit (ReLU) Activation Function

An activation function is a mathematical operation applied to the output of a neuron in a neural network. Its primary role is to introduce nonlinearity into the network, enabling it to model complex input-output relationships beyond simple linear mappings.

Reference [7] proposes a CNN architecture composed of four convolutional layers, two max-pooling layers, three dense layers, and one dropout layer. Uniquely, the researchers substitute the widely used ReLU (Rectified Linear Unit) activation function with Leaky ReLU, an enhanced variant designed to address specific limitations of ReLU. While ReLU is favored for its computational efficiency and effectiveness in mitigating the vanishing gradient problem, it can suffer from the "dying ReLU" issue, where neurons can become inactive for negative inputs, thereby obstructing effective training. Leaky ReLU counteracts this by introducing a small, non-zero slope for negative values, which keeps neurons active even for negative inputs. Mathematically, the Leaky ReLU function is defined as $f(x) = max(ax, x)$, where x is the input to the neuron, and a is a small constant, typically set to a value like 0.01. When x is positive, the Leaky ReLU function behaves like the ReLU function, returning x. However, when x is negative, the Leaky ReLU function returns a small negative value proportional to the input x, preserving some gradient and preventing neuron deactivation.

This adjustment in activation function yielded improved model performance in Reference [7], as reflected in the testing results: the CNN model with the standard ReLU activation achieved a test accuracy of 0.9768 and a test loss of 0.0827, whereas the model incorporating Leaky ReLU attained a higher test accuracy of 0.9791 and a reduced test loss of 0.0721. The improvement underscores Leaky ReLU's advantage in sustaining active neurons and enhancing gradient flow, which promotes more effective training and model accuracy.

Applying the 'Transfer Learning via Fine-tuning'

The third architectural enhancement observed is related to applying the 'transfer learning via fine-tuning' using the pre-trained CNN variant models. Transfer learning is the process of taking a CNN model that has been trained on a large dataset and applying its knowledge to a smaller dataset with a similar purpose. This technique provides a better starting point and can accomplish tasks at a certain level even without training. Furthermore, it saves time and requires less computational resources.

References [8] and [9] chose the MobileNet-v2 (a pre-trained CNN with 53 layers deep) and LeNet-5 (a simpler and one of the first pre-trained models) models, respectively. Subsequent fine-tuning were made to the selected pre-trained models in both studies. Reference [8] eliminated the last SoftMax layer of the MobileNet-v2 and replaced with three (3) hidden layers of ReLU neurons: Layer 1 of 800 neurons, Layer 2 of 400 neurons, and Layer 3 of 200 neurons. Meanwhile, Reference [9] conducted initial experiments on LeNet-5, exploring various hyperparameters such as activation functions, optimization algorithms, and the placement of batch normalization layers. Subsequently, the superior performing hyperparameters were chosen and amalgamated to form an improved model. Table III presents the optimal performance hyperparameters that have been incorporated into the pre-trained LeNet-5 model [9].

TABLE III. BEST HYPERPARAMETERS INTEGRATED INTO THE LENET-5 PRE-TRAINED MODEL

Hyperparameter	Best Hyperparameter	Test Accuracy
Type of Pooling Layer	Max-pooling	-
Position of Batch Normalization Layers	After every convolutional layer	0.9389
Optimization Algorithm	Adam optimizer	0.9291
Activation Function	Swish function	0.9039
Dropout Layer	Added after the third convolutional layer	-

Reference [10], on the other hand, chose the ResNet50 as its base model due to the benefit of convolutional block attention module (CBAM), which effectively learns the image's channel and spatial position information resulting in an improved robustness and feature extraction capabilities. A few changes were made by Reference [10], including the addition of the GlobalAveragePooling2D layer and the dense layer with activation equal to ReLU. The Adam optimizer and category cross entropy loss function are then added to the improved ResNet50 during compilation. Unfortunately, it typically needs more memory and processing power and is prone to overfitting, particularly with small datasets.

Utilization of Region Proposal Network for Text Areas

Instead of employing convolutional sliding windows to extract features from the input image, the Region Proposal Network (RPN) is used to propose candidate text areas (axis-aligned bounding boxes) on a feature map. Then, for each RPN-generated text region, three (3) ROI pooling of varying sizes are applied to observe additional text features [15]. The illustration of RPN and ROI pooling are shown in Fig. 5.

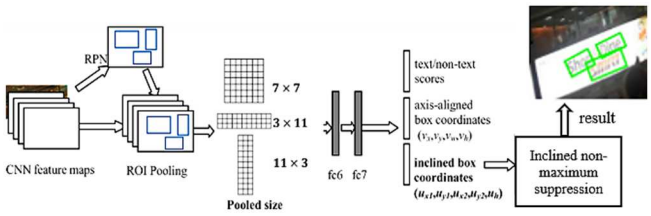

Fig. 5. The Architecture of RPN and Three (3) ROI Pooling as Specified in Reference [15]

With reference to the RPN architecture, researchers in [14] embedded non-discriminatory regularization and multi-task loss techniques to address the limitations associated with overfitting. The primary goal of non-discriminatory regularization is to take into account all non-discriminative word analysis, which leads to an effective feature analysis. In addition, an inclined non-maximum suppression is added in the post process, along with both non-discriminatory regularization and multi-task loss techniques.

As a conclusion, CNNs are widely used in computer vision and image processing tasks due to their advantages, including strong inductive bias, hierarchical representation, parameter sharing, and end-to-end training. However, it has significant shortcomings, including high computational requirements, lengthy training time, particularly for large labelled datasets, and susceptibility to overfitting.

B. *Vision Transformer*

The Vision Transformer (ViT) is a more recent deep learning-based model that is backboned by the Transformer's self-attention-based architecture [16]. Motivated by its successful application in Natural Language Processing (NLP), researchers have explored its potential in analyzing the handwriting of pre-school and primary school students to determine whether they exhibit symptoms of dysgraphia or not [17] by following the architecture presented by [18].

To accommodate 2D input images, the input image $x \in R^{H\times W\times C}$ must be reshaped into a sequence of flattened 2D patches $x_p \in R^{N\times((P2).C)}$. Here, *(H, W)* denotes the resolution of the original image, *C* represents the number of channels, *(P, P)* signifies the resolution of each image patch, and $N=HW/P^2$ represents the resulting number of patches. This *N* value also serves as the effective input sequence length for the Transformer. Subsequently, the patches are augmented with the embedding position to preserve their positional information. Fig. 6. presents a comprehensive illustration of the architecture of ViT.

The main components of Transformer encoder are multi-head attention layer and Multilayer Perceptron (MLP) layer, which is commonly referred to as a feed-forward network layer. The technique of layer normalization is implemented to both the layers individually. Self-attention is the mechanism employed in multi-head attention. Self-attention involves the utilization of query (Q), key (K), and value (V) as input. The resulting output is obtained by taking the weighted sum of the value vector, with the weights being determined through the utilization of the SoftMax function. The definition of attention is shown in equation (1).

$$Attention\ (Q,K,V) = SoftMax\left(\frac{QK^T}{\sqrt{d}}\right)V \qquad (1)$$

where *d* represents the hidden dimensions.

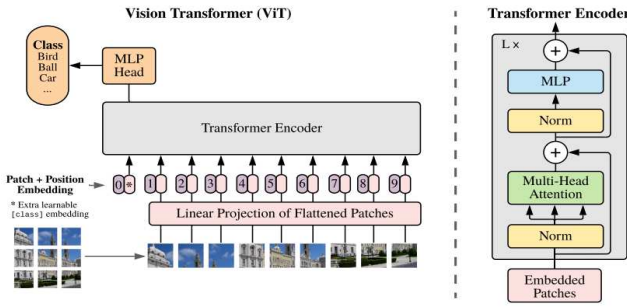

Fig. 6. The Vision Transformer Architecture

The efficacy of ViT resides in its self-attention mechanism, which facilitates the establishment of long-range contextual dependencies among pixels in images. Consequently, ViT shows the potential to generate output values of greater precision when trained on larger datasets [17] as well as allow faster training and inference [19]. Another notable advantage of the ViT architecture is its flexibility in handling images of varying sizes and aspect ratios without compromising resolution. This scalability makes it adaptable to diverse datasets and applications, ranging from simple object recognition to complex scene understanding [19].

C. *Performance Analysis*

Table IV and V present the testing accuracy results for handwriting image classification related to learning disabilities, utilizing both CNN and ViT models. However, it is important to note that performance results from all previous studies cannot be provided, as some utilized non-standardized handwriting image datasets, including self-collected primary datasets that are not publicly accessible due to ethical considerations. The results presented in these tables are limited to studies that employed the same publicly available dyslexia handwriting dataset accessible through Kaggle at: *https://www.kaggle.com/datasets/drizasazanitaisa/dyslexia-handwriting-dataset*. This dataset comprises handwriting samples from children with and without dyslexia, collected from three different sources: the NIST Special Database 19, the Kaggle Database, and dyslexic children from Seberang Jaya Primary School, totaling 151,433 handwriting images.

TABLE IV. THE ACCURACY RESULTS OF HANDWRITING IMAGE CLASSIFICATION REPORTED BY PREVIOUS STUDIES UTILIZING THE CNN MODEL

Paper / Year	Deep Learning Model	Architectures		Testing Accuracy Result
		Conv Layer / Pooling Layer	Activation Function / Loss Function	
[13] / 2023	MobileNet V2	• 3x3 depth-wise separable conv layers. • Inverted residual stride. • Global average pooling layer.	ReLU / Not Stated	99.20%
[11] / 2023	VGG-16	• 13-conv layers. • 5 max pooling layers.	ReLU / Categorical Cross Entropy	97.98%
[7] / 2022	CNN	• 4-conv layers. • 3 dense layers. • 2 max pooling layers.	Leaky ReLU / Categorical Cross Entropy	97.91%
[9] / 2021	Modified LeNet-5	• 3-conv layers. • 2 max pooling layers.	Swish / Not Stated	95.34%
[6] / 2022	CNN	• 5-conv layers. • 5 batch normalization layers. • 5 max pooling layers.	ReLU / Categorical Cross Entropy	87.44%
[10] / 2022	Modified ResNet-50	• 5 stages each with convolution and identity blocks. • Each convolution and identity block has 3 conv layers. • 1 max pooling layer at the end of stage 1. • 1 average pooling layer at the end of stage 5.	ReLU / Categorical Cross Entropy	85.00%
[16] / 2022	CNN	• 3-conv layers. • 3 max pooling layers.	ReLU / Binary Cross Entropy	79.47%

TABLE V. THE ACCURACY RESULTS OF HANDWRITING IMAGE CLASSIFICATION REPORTED BY PREVIOUS STUDY UTILIZING THE VISION TRANSFORMER (VIT) MODEL

Paper / Year	Deep Learning Model	Architectures / Parameters				Testing Acc Result
		Image Size	Image Patch Size	Dim / Depth / Dropout Rate	Num of Attn Heads	
[16] / 2022	ViT	28x28	4x4	128 / 12 / 0.1	8	86.22%

V. FUTURE DIRECTIONS AND CONCLUDING REMARK

The first future research relates to adopting more comprehensive input data by fusing features from both offline and online handwriting to enhance accuracy. Offline handwriting, also known as image-based handwriting, is frequently associated with possessing static features [8]. On the other hand, online handwriting, often referred to as digital handwriting, has the capacity to capture the dynamic features of handwriting with the help of recorded trajectory of the digital pen using digitizing tablets [4]. By combining both features, researchers can perform a more comprehensive analysis. Current studies rarely fuse these two features, focusing on one or the other. Table VI indicates the complementing handwriting features of offline and online handwriting.

TABLE VI. OFFLINE VS ONLINE HANDWRITING FEATURES

Offline (Image) Handwriting Features	**Online (Digital)** Handwriting Features
Static - purely geometric characteristics of the written text [20]	Dynamic - typically captured while using the digitizing tablets [4, 22]
Includes: the writing size, non-aligned left-margin, skewed writing, insufficient space	Includes: pressure, altitude, azimuth, pen lifts, temporal duration of stroke, length of

between words, sharp angles, broken links between letters, collisions between two letters, irregular size of letters, atypical letters, ambiguous letters, and unstable track [19].	stroke, length of stroke in vertical and horizontal directions, on airtime, and velocity [22]

Next future research in dyslexia classification can benefit from hybrid models that integrate Convolutional Neural Networks (CNNs) with Vision Transformers (ViTs), leveraging the strengths of each to address two critical limitations: overfitting and restricted global feature comprehension. CNNs' hierarchical convolutional layers are highly effective at capturing localized, fine-grained handwriting features but are prone to overfitting, especially on expansive datasets with nuanced local variations. ViTs, on the other hand, utilize multi-head self-attention, establishing expansive interpixel relationships and capturing essential long-range dependencies across handwriting samples. A hybrid model fuses these strengths, balancing CNN-driven local feature recognition with ViTs' capacity for global context extraction and reducing overfitting in the process. Such an architecture not only refines the accuracy of dyslexia classification but establishes a robust framework adaptable to diverse handwriting data.

Finally, our literature discovered that existing works are still constrained to binary classifications of "at-risk" and "no-risk" learning disabilities. Because no two (2) students with learning disabilities are the same, dividing them into only two (2) groups is insufficient to meet their learning needs, personalization, and provision of adequate interventions. As a result, we believe that extending the binary classification into multi-classifications based on severity levels (normal, mild, moderate, and severe) can provide a more nuanced understanding of the variations in handwriting patterns and, as a result, achieve more detailed diagnoses of learning disabilities, and provide highly beneficial research insights, particularly for learning institutions.

As a conclusion, an automated diagnosis of learning disabilities using handwriting images is a domain that still have rooms for further research works. Despite its strong testing accuracy performance, current models still suffer from high computational costs, prone to overfitting, and lengthy training time. Thus, more research work should be undertaken to achieve greater accuracy in classification while lowering computational costs and shortening training time.

ACKNOWLEDGEMENT

This work is supported by the Ministry of Higher Education under Grant FRGS FRGS/1/2020/ICT02/UPM/02/2, project code 08-01-20-2315FR.

REFERENCES

[1] J. Kunhoth, S. Al-Maadeed, S. Kunhoth, and Y. Akbari, "Automa and potentialted Systems for diagnosis of dysgraphia in children: A survey and novel framework," arXiv.org, https://arxiv.org/abs/2206.13043 (accessed Jun. 14, 2023).

[2] I. H. Sarker, Deep learning: A comprehensive overview on techniques, taxonomy, applications and Research Directions, 2021. doi:10.20944/preprints202108.0060.v1.

[3] D. Bhatt *et al.*, "CNN variants for Computer Vision: History, architecture, application, challenges and future scope," *Electronics*, vol. 10, no. 20, p. 2470, 2021. doi:10.3390/electronics10202470.

[4] J. Skunda, B. Nerusil, and J. Polec, "Method for dysgraphia disorder detection using convolutional neural network," *CSRN*, 2022. doi:10.24132/csrn.3201.19.

[5] Y. Pratheepan and B. Braj, "Deep Learning Approach to Automated Detection of Dyslexia-Dysgraphia," in *25th IEEE International Conference on Pattern Recognition*, 2020.

[6] S. A. Ramlan, I. S. Isa, M. K. Osman, A. P. Ismail, and Z. H. Che Soh, "Investigating the impact of CNN layers on dysgraphia handwriting image classification performance," *Journal of Electrical & Electronic Systems Research*, vol. 21, no. OCT2022, pp. 73–83, 2022. doi:10.24191/jeesr.v21i1.010.

[7] S. Sreekumar and L. A, "Comparative study of CNN models on the classification of dyslexic handwriting," *2022 IEEE Bombay Section Signature Conference (IBSSC)*, 2022. doi:10.1109/ibssc56953.2022.10037428.

[8] N. S. Mor and K. L. Dardeck, "Applying a Convolutional Neural Network to Screen for Specific Learning Disorder," *Learning Disabilities: A Contemporary Journal*, vol. 19, no. 2, pp. 161–169, 2021.

[9] M. S. Rosli, I. S. Isa, S. A. Ramlan, S. N. Sulaiman, and M. I. Maruzuki, "Development of CNN transfer learning for dyslexia handwriting recognition," *2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2021. doi:10.1109/iccsce52189.2021.9530971.

[10] A. Sasidhar, G. K. Kumar, K. Yoshitha, and N. Tulasi, "Dyslexia discernment in children based on handwriting images using residual neural network," *2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, 2022. doi:10.1109/csitss57437.2022.10026368.

[11] C. Sharmila *et al.*, "An automated system for the early detection of dysgraphia using deep learning algorithms," *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, 2023. doi:10.1109/icscds56580.2023.10105022.

[12] H. A. Rashid, T. Malik, I. Siddiqui, N. Bhatti, and A. Samad, "DYSIGN: Towards computational screening of dyslexia and dysgraphia based on handwriting quality," *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 2023. doi:10.1145/3585088.3593890.

[13] Y. Alkhurayyif and A. R. Sait, "Deep learning-based model for detecting dyslexia using handwritten images," *Journal of Disability Research*, vol. 2, no. 4, 2023. doi:10.57197/jdr-2023-0059.

[14] F. Ghouse, R. Vaithiyanathan, and K. Paranjothi, "Dysgraphia classification based on the non-discrimination regularization in rotational region convolutional neural network," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 1, 2022. doi:10.22266/ijies2022.0228.06.

[15] Y. Jiang *et al.*, "R2 CNN: Rotational Region CNN for arbitrarily-oriented scene text detection," *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018. doi:10.1109/icpr.2018.8545598.

[16] V. Ashish *et al.,* "Attention is All You Need," *31st Conference on Neural Information Processing Systems,* 2017.

[17] V. Vilasini, B. Banu Rekha, V. Sandeep, and V. Charan Venkatesh, "Deep Learning Techniques to Detect Learning Disabilities Among children using Handwriting." *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*, 2022, doi: 10.1109/icicict54557.2022.9917890.

[18] D. Alexey *et al.,* "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv.org, https://doi.org/10.48550/arXiv.2010.11929.

[19] J. Maurício, I. Domingues, and J. Bernardino, "Comparing vision transformers and convolutional neural networks for Image Classification: A Literature Review," *Applied Sciences*, vol. 13, no. 9, p. 5521, Apr. 2023. doi:10.3390/app13095521.

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

[21] G. Dimauro, V. Bevilacqua, L. Colizzi, and D. Di Pierro, "Testgraphia, a software system for the early diagnosis of dysgraphia," *IEEE Access*, vol. 8, pp. 19564–19575, 2020. doi:10.1109/access.2020.2968367.

[22] J. Kunhoth, S. Al Maadeed, M. Saleh, and Y. Akbari, "Machine learning methods for dysgraphia screening with online handwriting features," *2022 International Conference on Computer and Applications (ICCA)*, 2022. doi:10.1109/icca56443.2022.10039584.