Pertanika J. Sci. & Technol. 32 (5): 2013 - 2032 (2024)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Comparing CNN-based Architectures for Dysgraphia Handwriting Classification Performance

Siti Azura Ramlan, Iza Sazanita Isa*, Muhammad Khusairi Osman, Ahmad Puad Ismail and Zainal Hisham Che Soh

Electrical Engineering Studies, College of Engineering, Universiti Teknologi MARA, Cawangan Pulau Pinang, Permatang Pauh Campus, 13500 Permatang Pauh, Penang, Malaysia

ABSTRACT

Deep learning algorithms are increasingly being used to diagnose dysgraphia by concentrating on the issue of uneven handwriting characteristics, which is common among children in the early stage of basic learning of reading and writing skills. Convolutional Neural Network (CNN) is a deep learning model popular for classification tasks, including the dysgraphia detection process in assisting traditional diagnosis procedures. The CNN-based model is usually constructed by combining layers in the extraction network to capture the features of offline handwriting images before the classification network. However, concerns have been expressed regarding the limited study comparing the performance of the Directed Acyclic Graph (DAG) and Sequential Networks in handwriting-related studies in identifying dysgraphia. The proposed method was employed in this study to compare the two network structures utilized for feature extraction in classifying dysgraphia handwriting To eliminate this gap. Therefore, a new layer structure design in the Sequential and DAG networks was proposed to compare the performance of two feature extraction layers. The findings demonstrated that the DAG network outperforms the Sequential network with 1.75% higher accuracy in classification testing based on confusion matrix analysis. The study provides valuable insights into the efficiency of various network structures in recognizing inconsistencies identified in dysgraphia handwriting, underlining the need for additional research and improvement in this field. Subsequently, these findings highlight the necessity of deep learning approaches to advance dysgraphia identification and establish the framework for future research.

Keywords: Convolutional Neural Network, deep learning, directed acyclic graph, dysgraphia handwriting, handwriting analysis

ARTICLE INFO

Article history:
Received: 14 August 2023
Accepted: 07 March 2024
Published: 08 August 2024

DOI: https://doi.org/10.47836/pjst.32.5.05

E-mail addresses:
sitiazura@uitm.edu.my (Siti Azura Ramlan)
izasazanita@uitm.edu.my (Iza Sazanita Isa)
khusairi@uitm.edu.my (Muhammad Khusairi Osman)
ahmadpuad127@uitm.edu.my (Ahmad Puad Ismail)
zainal872@uitm.edu.my (Zainal Hisham Che Soh)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

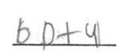
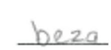
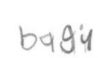
INTRODUCTION

Dysgraphia is a complex learning disability that affects language skills, including writing, spelling, and comprehension, and causes difficulties in a child's academic and social life (Deuel, 1995). Writing skills are key abilities that children must develop during their school years. Dysgraphia children, on the other hand, suffer from handwriting difficulties and lack the writing skills that are expected for his or her age and cognitive level (Chung et al., 2020; Vlachos & Avramidis, 2020). Dysgraphia children's handwriting products exhibit indications of inconsistent handwriting, improper letter size, reversed letter form, and corrected handwriting (Biotteau et al., 2019). Some scholars have also detected dysgraphia in children presented with spelling impairment (Šafářová et al., 2021; Vlachos & Avramidis, 2020) that affects writing skills, prohibiting children from writing words quickly and consistently. Table 1 presents examples of dysgraphia and normal handwriting images by children aged 7 to 12.

Addressing dysgraphia and finding appropriate interventions are critical in increasing learning and ensuring success in education. Furthermore, traditional techniques of detecting dysgraphia rely primarily on subjective assessments, such as scoring tests and observations-based methods, which can be time-consuming, biased, and lacking objectivity (Dimauro et al., 2020). These constraints emphasize the need for more accurate and efficient dysgraphia detection systems. In recent years, computer-based approaches have emerged as a potential solution for dysgraphia detection. These approaches leverage various features and algorithms to analyze and interpret dysgraphia symptoms, such as inconsistent handwriting with redundant form, reversal and corrected letters (Vaivre-Douret et al., 2021). The accuracy and efficiency of dysgraphia detection are improved by computer-based methods, providing an objective and quantitative evaluation.

By automatically learning and extracting information from dysgraphia-related data, machine learning and deep learning algorithms have improved dysgraphia identification. These methods, particularly Convolutional Neural Networks (CNNs), have been performed in image classification tasks with various input images. CNNs comprise learnable layers that extract hierarchical information from input images (Almisreb et al., 2022). The Directed Acyclic Graph Network (DAGN) and Sequential Network (SN) are two leading CNN architectures that could be implemented to extract features of layer network construction. SN is a simple network built on a single or multi-layer architecture with no shortcuts between the layers. In contrast, DAGN uses skip

Table 1
Dysgraphia and normal handwriting images by school-age children

Dysgraphia handwriting	Normal handwriting
	
	
	

connections, allowing direct connections between non-adjacent layers and improving gradient flow, resulting in improved classification performance. Moreover, by capturing specific features from dysgraphia handwriting, CNNs have shown the potential to be explored in detecting dysgraphia. Nonetheless, the performance comparison of DAGN and SN designs in dysgraphia detection is limited and current research lacks a detailed analysis of various network designs and their effects on classification. The same number of convolutional layers in the network could be used to better understand the advantages and disadvantages of DAGN and SN, which could help improve dysgraphia detection methods based on the performance of both networks.

This study compares the performance of DAGN and SN architectures for dysgraphia handwriting classification. By evaluating accuracy, precision, recall, and F1-score on a large dataset, it seeks to provide insights into the effectiveness of these network designs. The findings will advance dysgraphia detection methods and guide the development of more accurate deep-learning models for identifying dysgraphia based on handwriting characteristics. The proposed layer structure network could achieve significant output as a considerable network structure to integrate with another model in improving the dysgraphia detection approach.

RELATED WORK

Numerous methods have been presented for detecting the presence of dysgraphia symptoms in children and adults using various input domains. The domain of online and offline handwriting was utilized and extracted to obtain an accurate diagnosis. Most researchers focus on accurate diagnosis by implementing a machine-learning algorithm. Online handwriting frequently extracts acceleration, pressure, and pen tilt, typically using additional instruments such as a tablet (Asselborn et al., 2020; Dankovicova et al., 2019; Kunhoth et al., 2023). In the meantime, digitized offline handwriting could be extracted as static features (letter shape, missing letters/words, uneven slanting, and inconsistent letter size) based on the output of handwriting on paper. Researchers have investigated both domains, exponentially demonstrating the potential of machine learning techniques for detecting dysgraphia symptoms.

More research in machine learning techniques has been conducted on classification algorithms and feature extraction. Machine learning algorithms, support vector machines (SVM), AdaBoost, and decision trees were applied to extracted features, and a dysgraphia diagnostic model was developed, as Kunhoth et al. (2023) demonstrated. From the study, the AdaBoost classifier has shown high accuracy with 80.8% accuracy, 1.3% more than the state-of-the-art method, similar to the research of Devillaine et al. (2021) that presented a machine learning algorithm-based pre-diagnosis tool for dysgraphia in France. From the study, Random Forest obtained the best accuracy score of 73.4% compared to Extra Trees and Multi-Layers Perceptron (MLP). In a similar perspective of the input domain, Sihwi et al. (2019) developed a Support Vector Machine (SVM) with several kernels, including Linear, Polynomial, and Radial Base Functions (RBF), to classify the collected data and identify dysgraphia. By handling the Synthetic Minority Over-sampling Technique (SMOTE) for unbalanced data, the RBF kernel produced the highest accuracy score of 82.51%.

A study by Dankovicova et al. (2019) explored the application of various machine learning techniques, which are random forest, support vector machine, and adaptive boosting, to analyze and extract attributes from irregular handwriting and identify dysgraphia characteristics. While the study employs hyperparameter tuning, 3-fold stratified cross-validation, and normalized data to predict and assess the handwriting data, the principal component analysis has been used to visualize these attributes in a two-dimensional space with a success rate of approximately 67%. The performance indicates a significant finding based on the various input domains and extracted features used in the previous study. It might be improved more accurately by aiding the diagnosis procedure with minimal consumption. The knowledge of machine learning has grown, and a more advanced concept known as deep learning promises high performance in computer vision.

As deep learning research continues to grow, it offers the promise of additional revolutionary advancements in a variety of disciplines, including dysgraphia identification. Deep learning has shown outstanding performance in a variety of input domains used and across various disciplines. In computer vision, convolutional neural networks (CNNs) have achieved state-of-the-art performance in image classification, object detection, and image segmentation tasks (Chai et al., 2021). The effectiveness of CNNs in extracting feature spatial hierarchies has resulted in advances in image classification. Based on previous studies, several research studies have been conducted using the CNN-based model to classify dysgraphia and non-dysgraphia handwriting symptoms. A hybrid CNN-LSTM Random Forest model has been performed in classifying handwriting characters for Parkinson's patients (Masood et al., 2023). The combination of CNN and LSTM layer construction captured spatial and sequential information from the images, while Random Forest enhanced the classification performance. With different feature extraction methods, Devi and Kavya (2023) implemented a combined feature extraction formula of Kekre-Discrete Cosine Transform with Deep Transfer Learning (K-DCT-DTL) to select prominent handwriting and geometrical features. Another CNN-based model has been presented by Vilasini et al. (2022), which was performed to identify and detect abnormal handwriting among children with learning disability. The CNN model and Vision Transformers (ViT) were utilized, and they have contributed to developing an efficient approach to dysgraphia detection research.

A study by Ghouse et al. (2022) demonstrated the use of balancing parameters in the loss function to balance the class during training and eliminating the features to reduce overfitting problems in classifying dysgraphia and non-dysgraphia by implementing Non-Discrimination Regularization in Rotational Region Convolutional Neural Network (NDR-R2CNN) model. By using graphic tablets to capture the dynamic features of letter writing, Zolna et al. (2019) developed a Recurrent Neural Network model (RNN) to identify children with dysgraphia. The sequential CNN-based model was explored and experimented with using letter handwriting images, and it was shown that the performance could be improved through different layer construction, such as the number of convolutional layers and different activation function layers (Ramlan et al., 2022). The experimental results predicted that the classification performance could increase if the convolutional layers increase. Prior research indicated that numerous implementations of input domain features, such as offline or online handwriting image features, have utilized CNN-based models to detect and identify dysgraphia in children and adults.

CNN-based models have shown gradual growth in dysgraphia identification and classification based on handwriting features extracted during model training. Besides the model design, the datasets used in classifying dysgraphia handwriting are one of the crucial parameters to be considered. These datasets often include a diverse range of writing styles, ages, and severity levels of dysgraphia. Additionally, some studies employed data augmentation techniques to increase the size and diversity of the training dataset, such as rotation, scaling, and noise injection. The use of assistive tools such as a graphic tablet to preserve the dynamic features of handwriting aids in the performance of classification tasks, but it comes at a cost. Static features of handwriting products (offline handwriting) can supplement the limitations of employing online handwriting. However, functional CNN layer building is required to ensure that effective feature extraction can be done and classified appropriately. Most research studies focused on the performance accuracy of the model developed. However, limited research has investigated the impact of layer construction and hyperparameter comparison in handwriting image classification, especially in dysgraphia identification.

To summarize, deep learning has emerged as a dominant framework in machine learning, adapting several fields and delivering outstanding performance in complex tasks. A CNN-based model is a focus model designed to identify dysgraphia symptoms and assist in dysgraphia diagnosis. Limited research has been conducted on a Directed Acyclic Graph (DAG) network to investigate the effect of the layer construction on classification performance. This study compares the performance of two models in classifying dysgraphia offline handwriting products, namely Sequential and DAG network designs.

METHODOLOGY

An experimental study is used to achieve the research objective of comparing the performance between two different construction layers in extracting the features of

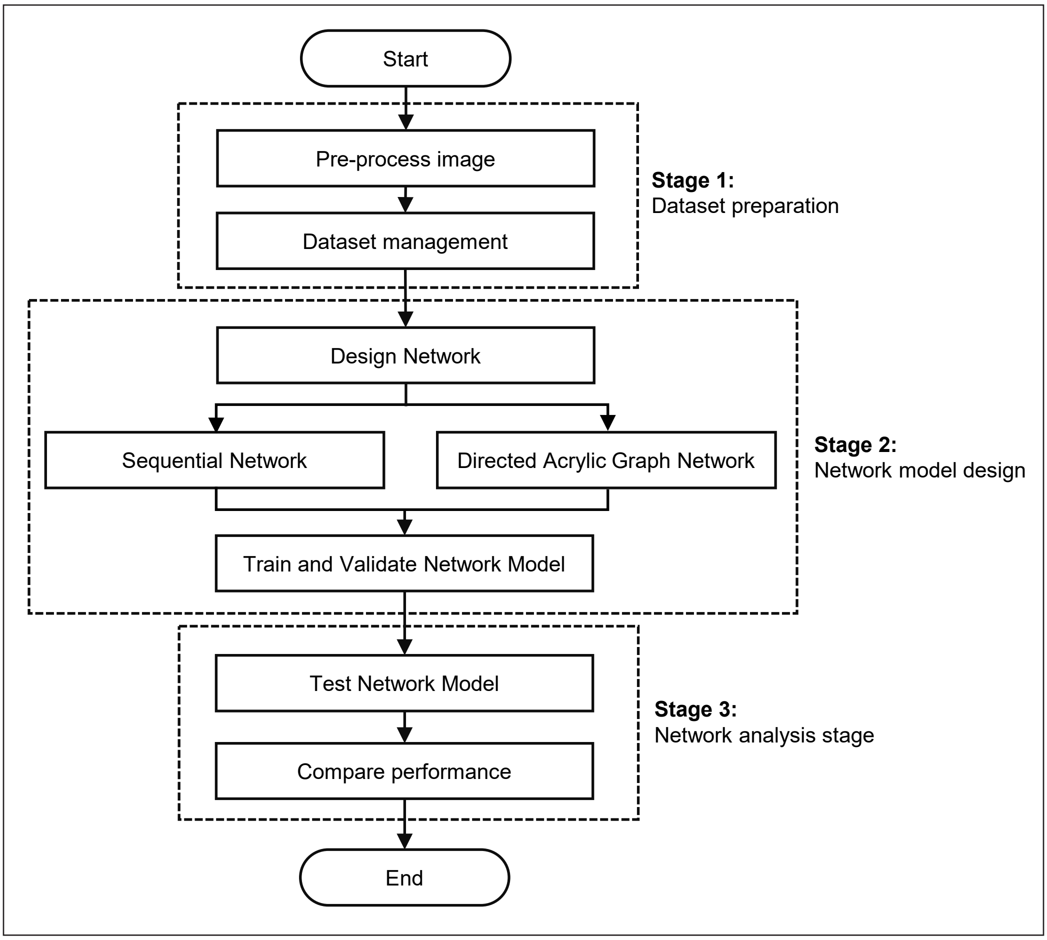


Figure 1. Overall experimental procedure for comparing networks performance of dysgraphia classification

handwriting images. This methodology focuses on the novel network layer structures proposed for Sequential and DAG feature extraction networks.

The overall experimental procedure depicted in Figure 1 is divided into three major stages: (1) preparing the dataset, (2) continuing network development, (3) and network analysis activities to complete the procedure. The first stage entails preprocessing the images using image processing techniques, including resizing and rotating to normalize the data. Then, the dataset is split into train and test portions to prepare it for input to the next stage.

The network architecture is developed in the second stage by considering the layer construction, activation function, and connectivity patterns. A model is developed in this second step, which could effectively learn and extract useful features from the input data. The hyperparameters, the adjustable parameters that control the network's learning process, include learning rate, batch size, number of epochs, optimized function and regularization techniques. Tuning these hyperparameters is critical for optimizing network performance and generalization capability. The network model is trained and validated using the prepared dataset after the hyperparameters have been determined. The model learns from the training data during training and adjusts its parameters based on the set loss function and optimization method—the validation phase aids in monitoring the model's performance on unseen data and preventing overfitting.

The model is then tested using a separate dataset in the third stage. This step assesses the network's generalization capabilities, offering insight into how well it operates on unseen data. The test results evaluate the model's performance and usefulness in addressing the target problem. When the execution is complete, the next step is to compare the network model's performance against other models or benchmarks. Based on this comparison, decisions about the feasibility and effectiveness of the established network can be made, marking the end of the process. If the execution is incomplete, the workflow returns to the training and validation phase. This iteration allows further model refining by altering hyperparameters, changing the network topology, or experimenting with other training procedures.

Following this iterative process, the proposed workflow ensures a systematic approach to designing, training, and assessing network models for a specific task. MATLAB 2021a environment was used with supported hardware that included a 2.50GHz Intel® CoreTM i5-10500H CPU and an NVIDIA GeForce RTX 3060 graphics processing unit to complete the overall experimental procedure.

Dataset Preparation

The dataset utilized in this experiment was obtained from the Kaggle database (https://www.kaggle.com/datasets/drizasazanitaisa/dyslexia-handwriting-dataset), which consists of an established image dataset (Rosli et al., 2021). The dataset preparation process establishes the image dataset and aims to ensure that it is suitable for execution as an input in deep neural networks. It consists of two key parts: preprocessing all images and the dataset management process.

Preprocess Image

The preprocessing steps collectively prepare the image data for training deep learning models, making the input more suitable for the classification task. Figure 2 illustrates the preprocessing stages, which include binarization, inversed black-and-white pixels, image resizing, noise injection, and rotation.

Initially, the image undergoes binarization, where pixel values are simplified to binary form based on a threshold. The image's colors are then inverted, changing white pixels into black and vice versa. The image is then resized to 32 × 32 dimensions to meet standardized input sizes. Controlled amounts of random noise are injected to enhance the model's robustness by exposing it to variations in real-world data. Finally, the image undergoes rotation to a specified degree to diversify the training dataset, ensuring the model's ability to handle variations in object orientation. These preprocessing stages work together to produce a refined and optimal input for further analysis and CNN model training (Rosli et al., 2021).

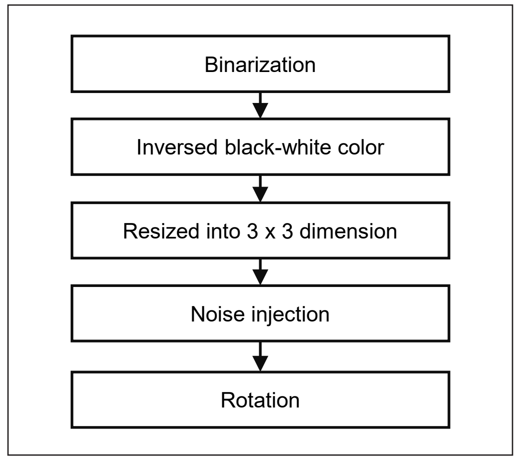


Figure 2. Preprocessing steps

Dataset Management

This procedure organizes all preprocessed data images into specified classes before they are entered into the network models. After completing the preprocessing activities, the established collection now contains 267,930 images, which include noise injection and rotated images. Therefore, each class has a balanced representation of the collection. The dataset input size is 32 × 32 pixels, encompassing rows and columns for dysgraphia and normal class. Table 2 presents the percentage proportion and numerical distribution of the training and testing datasets. About 85% of the dataset is used for training, and the remaining 15% is used for testing. However, 30% of the training dataset was randomly selected for the validation phase.

Table 2
Dataset division of training and testing

Dataset Category	Number of Images	Percentage
Training	228816	85%
Testing	39114	15%

Network Model Design

Network Design

As illustrated in Figure 3, the overall network construction consists of an extraction network and a classification network. This experiment used two types of new layer structures for extraction networks, as shown in Figures 4 and 5. The following classification network employed a fully connected Softmax layer and yielded the predicted output class at the end of the network.

Figure 4 depicts the architecture of SN, which includes three convolutional layers, batch normalization, and ReLu as an activation function. Before proceeding to the classification network, each layer block finishes with the max pooling layer.

Figure 5 demonstrates the DAGN layer architecture, which includes three convolutional layers (Conv), batch normalization (BaN), and activation function using ReLu. A skip connection is a network connection between the two layers that enables the gradient to move directly from the output to the input levels. During the forward and backward propagation training phases, this connection allows the network to bypass one or more layers. Mixing inputs from distinct layers requires an additional layer to complete the acyclic graph. By

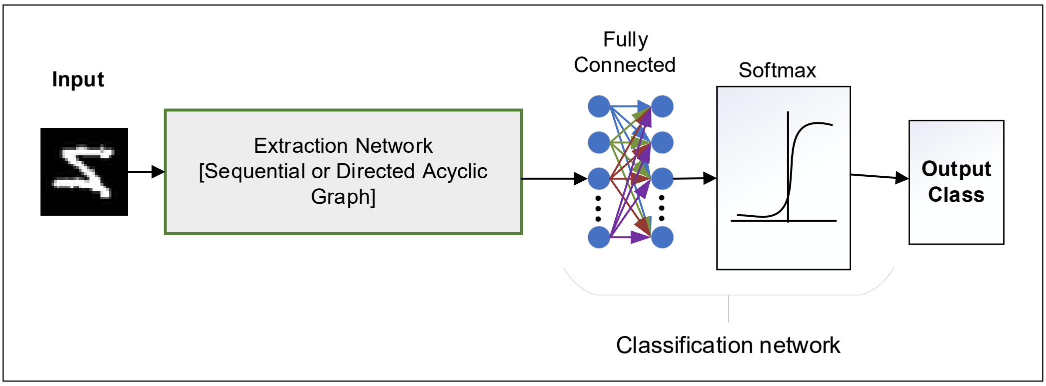


Figure 3. Overall network construction for Sequential Network and Directed Acyclic Graph Network

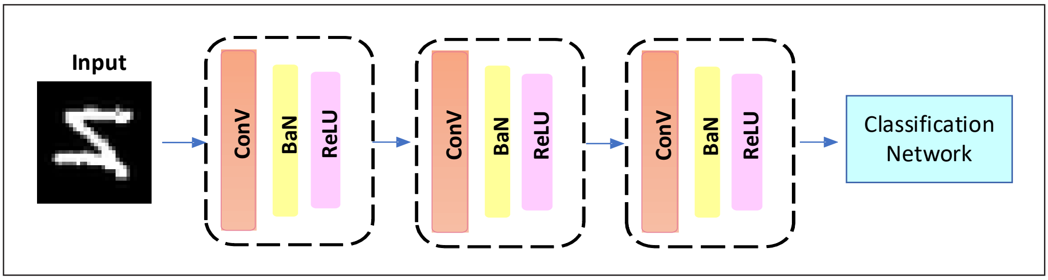


Figure 4. A new structure layer of Sequential Network architecture
Note. ConV = Convolutional layers, BaN = Batch normalization

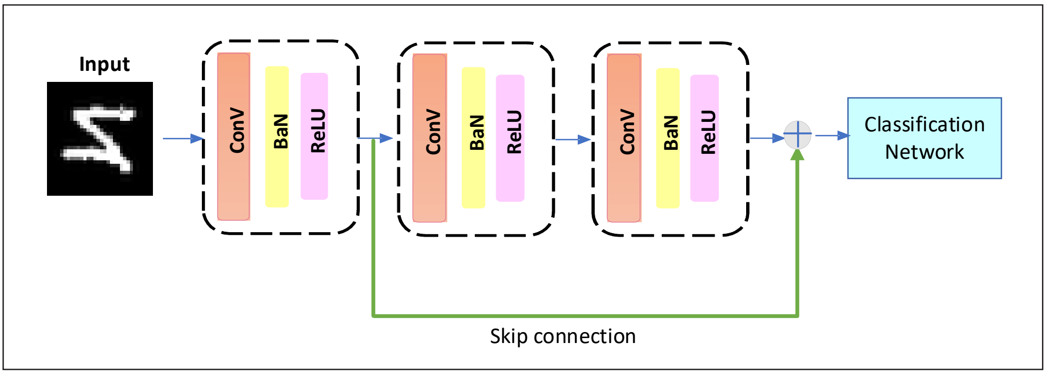


Figure 5. A new structure layer of Directed Acyclic Graph Network architecture
Note. ConV = Convolutional layers, BaN = Batch normalization

using an average pooling prior classification network, the average of the items identified in the filtered area of the feature map is determined.

For each network, the neural network model parameters for dysgraphia screening using handwriting images were independently adjusted, as presented in Table 3. Both networks used the "sgdm" optimizer (Stochastic Gradient Descent with Momentum). The learning rate was set to 0.001 for SN and 0.01 for DAGN, which defines the step size for minimizing the loss. It implies that the model iterates and adjusts the weights each time. The model is trained for eight epochs, which means the network processes the dataset eight times during training. Each epoch has 1251 iterations, with the weights adjusted every 30 iterations. The optimizer, learning rate, epochs, and iterations per epoch are all factors that affect the performance and training duration of the neural network model used to classify handwriting images as dyslexic or normal.

Table 3
Parameter set up for Sequential Network (SN) and Directed Acyclic Graph Network model (DAGN)

Parameter	SN	DAGN
Optimizer	SGDM	SGDM
Learning Rate	0.001	0.01
Number of epochs	8	8

Train and Validate Model

Training and validating a deep learning network model require a structured process to ensure the model learns meaningful representations from the input data. During the training phase, the CNN model is given a labeled dataset and iteratively modifies its internal parameters to reduce the discrepancy between predicted and actual outputs (Alzubaidi et al., 2021). This optimization is often accomplished through backpropagation and gradient descent. Meanwhile, the validation phase is important for determining the model's performance on unseen data. In this phase, the model's generalization skills are evaluated using a separate dataset not used during training. To avoid overfitting, fine-tune the model based on the validation results and repeat this iterative procedure until the model achieves satisfactory performance on both the training and validation datasets.

Network Analysis

Performance evaluation assesses the neural network model's capacity to detect potential dysgraphia and normal handwriting. This experiment uses performance evaluation to track and measure how the CNN model performs during training and testing. In this phase, network analysis involves analyzing the network's performance based on the testing results obtained from untested data and subsequent to the comparison of both SN and DAGN performances.

Network Testing

The testing phase evaluates the model's overall performance on a completely independent dataset, finalizing its capability to make accurate predictions in real-world scenarios. This process ensures that the CNN-based model is robust and accurate and can generalize well beyond the training data. The binary class confusion matrix was used to evaluate the effectiveness of each model in achieving network testing. The performance was assessed using accuracy, precision, recall, and the F1-score. All computations are based on a binary confusion matrix (Sokolova & Lapalme, 2009) to identify potential dysgraphia and normal handwriting.

Performance Comparison

The final measurement involves comparing SN and DAGN performances according to the best achievement of accuracy, precision, recall, and f1-score as harmonic values to validate the performance measurement.

RESULTS

The classification performance of both SN and DAGN on potential dysgraphia handwriting images is analyzed using the confusion matrix and loss graph. Additionally, the extraction of layer networks influences classification performance through the result of accuracy, precision, recall, F1-score and loss obtained were analyzed. It also discusses how the conclusions may improve dysgraphia detection methods based on deep learning, especially Convolutional Neural Network (CNN).

The results shown in Figure 6 emphasize the classification performance of two network models: SN and DAGN. The SN model attained an impressive training accuracy of 94.27%, but its validation accuracy was slightly lower, and its testing accuracy decreased to 86%. In contrast, the DAGN model outperformed all other models in all phases. It outperformed the SN model in terms of training accuracy, achieving 96.17%, and demonstrated improved generalization, with a validation accuracy of 95.2%. Notably, the DAGN model exceeded the SN model in testing accuracy, scoring 87.75%. These data highlight the DAGN model's improved performance and generalization capability in dysgraphia classification compared to the SN model. The finding shows that the DAGN model outperforms the SN model in terms of accuracy during training, validation, and testing. The finding shows

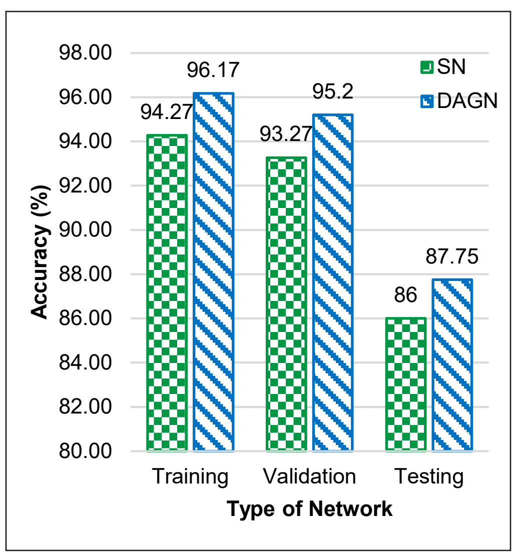


Figure 6. Accuracy performance for Sequential Network (SN) and Directed Acyclic Graph Network (DAGN)

that the DAGN model is more capable of learning and adapting properly to new unseen data. The higher accuracy achieved by the DAGN model on the testing dataset shows that it is more reliable and effective than the SN model in accurately classifying handwriting images. The DAGN model took 117 minutes and 9 seconds to train, while the SN model took somewhat longer at 121 minutes and 49 seconds. Although the SN model finished significantly faster, the time difference between the two models is relatively small.

The results presented in Figure 7 show the training and validation accuracies across multiple epochs. The training accuracy of the SN model begins at 46.09% at epoch 0 and steadily increases over consecutive epochs, reaching the highest at 94.53% at epoch 4. In subsequent epochs, it varied with slight decreases and increases. Similarly, the validation accuracy for the SN model began at 48.17% in epoch 0 and reached a maximum of 93.67% at epoch 8. Overall, the SN model improved training and validation accuracy over the length of the epochs. On the contrary, the DAGN model has better training accuracy across most epochs. It outperformed the SN model at 60.94% in epoch 0 and maintained relatively good training accuracy. At epoch 8, the highest training accuracy was obtained, 98.44%. Similarly, the DAGN model's validation accuracy began at 54.77% at epoch 0 and steadily rose to 95.20% at epoch 8. In terms of training and validation accuracy, the DAGN model consistently outperformed the SN model. The findings revealed that the DAGN model exhibited enhanced predictive capabilities and improved accuracy as the epoch progressed. These findings highlight the importance of network architecture, with the DAGN model's structure contributing to its higher performance over the SN model. The structure of layers, activation functions, or connection patterns in the DAGN model may allow it to extract relevant features and produce better predictions.

The loss graph for training and validation progress is shown in Figure 8. The training loss for the SN model began at 0.91 at epoch 0 and gradually reduced over subsequent

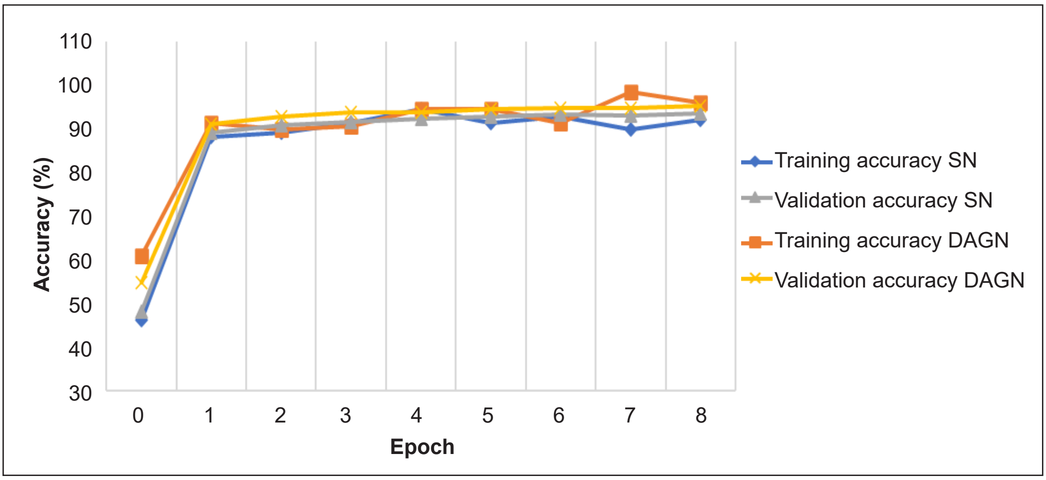


Figure 7. Accuracy performance graph for training and validation

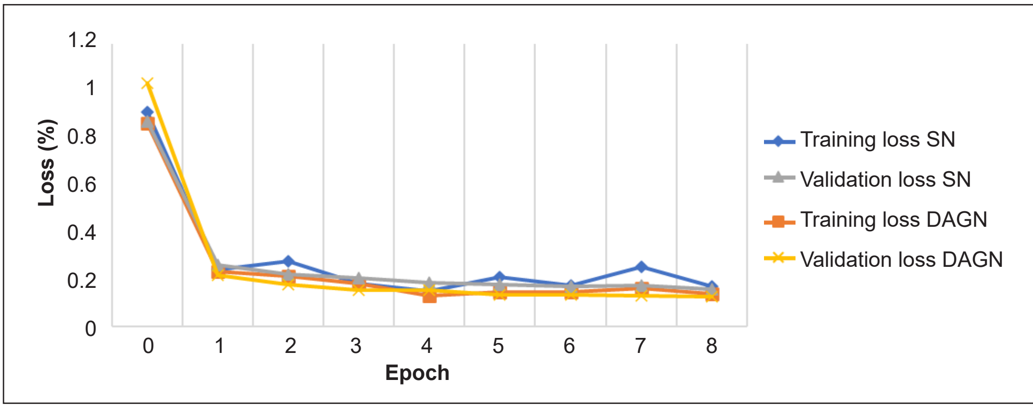


Figure 8. Loss graph for training and validation
Note. SN = Sequential Network, DAGN = Directed Acyclic Graph Network

epochs. It reached the lowest at 0.15 at epoch 4 and continued relatively low in subsequent epochs. Similarly, the validation loss for the SN model began at 0.87 at epoch 0 and decreased across the epochs, reaching a low of 0.16 at epoch 8. Over the training period, the SN model showed a reduction in training and validation losses. At the instance of the DAGN model, the training loss begins at 0.86 at epoch 0 and decreases progressively over the epochs. The validation loss for the DAGN model started at 1.03 at epoch 0 and rapidly reduced in the following epochs, reaching 0.13 at epoch 8. The DAGN model consistently reduced training and validation losses throughout the training process. The decreasing trend in losses suggests that both models are learning and adjusting their parameters to better capture the patterns in the data. These findings highlight the importance of network architecture and emphasize the potential benefits of utilizing the DAGN model for classifying handwriting images.

The data shown in Figure 9 compares the predicted classifications for DAGN to the actual classifications for two categories, namely dysgraphia and non-dysgraphia. In the first actual row, 42.0% (16,416) of the handwriting cases were correctly identified as dysgraphia, whereas 4.2% (1,652) cases of the actual dysgraphia handwriting were wrongly classified as non-dysgraphia which indicates a false negative rate. The second actual row shows that 8.0% (3,141) of the handwriting cases that were actually non-dysgraphia were wrongly labeled as dysgraphia, indicating a false positive rate. Meanwhile, 45.8 % (17,90) of cases were correctly classified as non-dysgraphia.

The testing confusion matrix of SN is depicted in Figure 10, and it was discovered that 43.9% (17,152) of the occurrences that were dysgraphia were accurately recognized as such. However, 7.8% (3,070) of those with dysgraphia were misclassified as non-dysgraphia. The second row shows that 6.1% (2,405) of the cases that were actually non-dysgraphia were misclassified as dysgraphia. However, 42.2% (16,48) of the truly non-dysgraphia examples were appropriately classified as such. Generally, the model successfully classifies

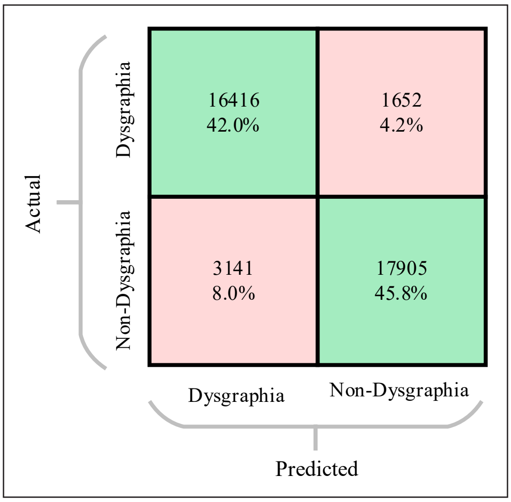


Figure 9. Testing confusion matrix of Directed Acyclic Graph Network

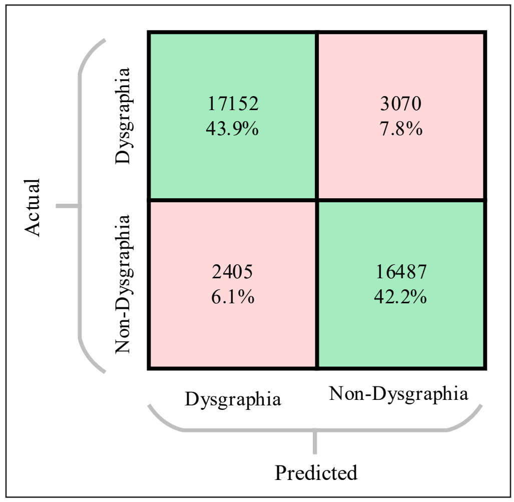


Figure 10. Testing confusion matrix of Sequential Network

dysgraphia with a higher rate of true predictions than misclassifications. However, more research is required to lower the false negative rate and ensure that all dysgraphia cases are correctly diagnosed. Similarly, with a higher percentage of accurate classifications, the model's performance in detecting non-dysgraphia cases is beneficial.

A more comprehensive insight into the model's performance is offered in Table 4 through additional analysis, including precision, recall, and F1-score. Precision for the dysgraphia class is high in training, validation, and testing for both SN and DAGN. The DAGN consistently outperforms the SN in terms of precision. During validation, the highest precision for SN was 92.90%, whereas DAGN yielded the highest precision with 97.07% accuracy. The recall scores for the dysgraphia class, SN, and DAGN are comparable, with SN having slightly higher recall values during validation and testing. The highest recall for DAGN is achieved during testing with 91.55% accuracy. For the dysgraphia class, the F1 scores for SN and DAGN are relatively close, with both models showing comparable performance during training, validation, and testing. During testing, the difference in performance between the SN and DAGN scores is only 0.12%.

For the non-dysgraphia class, the precision performance revealed that SN consistently outperforms DAGN at all training, validation, and testing. The maximum precision for SN was observed during testing, with 87.27% accuracy. Meanwhile, the recall score for DAGN increased at every stage of training, validation, and testing. The non-dysgraphia class achieved a high accuracy during testing, which was 7.15% greater than the SN class. SN and DAGN have similar F1 scores in the dysgraphia class and obtained the results with DAGN reaching slightly higher accuracy. DAGN achieved the greatest F1 score during testing and demonstrated 88.20% accuracy.

Table 4
Precision, recall and F1-score

	Type of Network	Training (%)		Validation (%)		Testing (%)	
		Dysgraphia	Non-Dysgraphia	Dysgraphia	Non-Dysgraphia	Dysgraphia	Non-Dysgraphia
Precision	SN	93.39	95.18	92.90	94.63	84.82	87.27
	DAGN	97.91	94.55	97.07	93.48	90.86	85.08
Recall	SN	95.27	93.26	94.74	92.76	87.70	84.30
	DAGN	94.35	97.99	93.22	97.18	83.94	91.55
F1 score	SN	94.32	94.09	93.81	93.56	86.24	85.29
	DAGN	95.96	96.24	94.92	95.29	86.36	88.20

Note. SN = Sequential Network, DAGN = Directed Acyclic Graph Network

The results show that SN and DAGN performance changes across the dysgraphia and non-dysgraphia classes. The DAGN model demonstrated exceptional performance throughout validation and testing, with accuracy rates of 95.2% and 87.75%, respectively. In contrast, the SN model obtained a meager 93.27% in validation and 86.0% in testing. Significantly superior in testing, the DAGN model exhibited a 1.75% improvement over the SN model. DAGN has higher precision and recall values in all classes, demonstrating a superior ability to classify handwriting. However, in terms of precision, SN outperforms the non-dysgraphia class. Both models have similar F1 scores, but the DAGN model performs better, with 86.36% and 88.20% accuracy in both classes.

DISCUSSION

The investigation results of this study demonstrate that the DAGN outperforms the SN in terms of accuracy and F1-score value. This finding indicates that the DAGN of extraction layer architecture is better suited for offline handwriting images in classifying dysgraphia and non-dysgraphia. The higher accuracy achieved by DAGN implies it is more effective at capturing and learning the underlying patterns and features than the SN.

The loss and training progress graphs presented in Figures 6 and 7 further support the superior performance of DAGN. The graphs clearly show that the DAGN model displays a faster convergence rate and lower training loss compared to the SN. It indicates that the DAGN layer construction is more efficient in optimizing the model parameters and minimizing the difference between predicted and actual classification. The consistent improvement in the loss and training progress throughout the training process indicates the efficiency and consistency of the DAGN model.

The testing confusion matrix in Figures 8 and 9 provides valuable details on the accuracy of predictions presented by both models. The confusion matrix demonstrates the DAGN's lower error rate in predicting target labels, highlighting its more accurate performance when compared to the SN. The confusion matrix demonstrates that the DAGN model has a higher accuracy in correctly classifying the target labels, with fewer instances of misclassification between different classes.

Furthermore, the precision, recall, and F1 scores indicated in Table 4 prove that DAGN outperformed SN. These metrics comprehensively evaluate the model's performance by considering true positives, false positives, and false negatives. Precision can be defined as true positives (actual dysgraphia class predicted as dysgraphia) proportion to all handwriting in actual dysgraphia class. Therefore, precision scores demonstrate the DAGN model's ability to correctly identify dysgraphia handwriting as actual instances in the dysgraphia class. Meanwhile, the recall score demonstrates that the DAGN model can effectively recognize each class from the overall handwriting input images. As a result, the F1 score presented the harmonic mean value of precision and recall score, indicating that DAGN is greater than SN for both classes.

Overall, our findings highlight the superiority of the DAGN architecture over the SN in terms of accuracy, loss optimization, prediction accuracy, and comprehensive evaluation metrics. The DAGN model's ability to capture complex patterns, faster convergence rate, and lower error rate in predicting the target labels indicate its robustness and efficacy. It is supported by the DAGN structure, which enables skip connections in layer construction. Skip connections enable CNNs to bypass some layers and connect directly to deeper or shallower ones (Mohammed et al., 2022). It could help to differentiate patterns from the image data. Additionally, skip connections can help address the problem of vanishing gradients by offering alternate paths for the gradients to flow (Qiao et al., 2018). Furthermore, they can make it easier and faster to train deeper networks with greater expressive capacity and the ability to extract more features from data images. These results contribute to the growing evidence supporting the advantages of utilizing DAGN architectures in similar problem domains. Future research should focus on exploring the underlying reasons behind the improved performance of DAGN and investigate its applicability to other domains and datasets.

Based on the investigation, several studies have been compared to the proposed study, which provides excellent results with more than 80% testing accuracy, as depicted in Table 5. According to Table 5, the proposed CNN has 87.75% testing accuracy for a simple DAG construction network with automated feature extraction and 86.0% for a simple sequential network. In a study by Devi and Kavya (2023), hand-crafted feature extraction was executed using the Kekre-Discrete Cosine Transform method and classified using deep transfer learning for offline handwriting, which yielded the highest performance at 99.75% accuracy. Hand-crafted feature extraction is not competent in representing the overall performance because it is usually not robust, and the computational requirement is high, especially for high-dimension images. The performance of the proposed CNN-based model does not show the highest percentage of accuracy. However, this proposed CNN is the simplest network,

Table 5
State-of-the-art performance comparison

Author	Model	Input Domain	Performance
Masood et al., 2023	CNN-LSTM Random Forest	Parkinson handwriting	92.6%
Devi & Kavya, 2023	Kekre-Discrete Cosine Transform with Deep Transfer Learning (K-DCT-DTL)	Offline handwriting	99.75%
Vilasini et al., 2022	Convolutional Neural Networks (CNN) and Vision Transformers (ViT)	Offline handwriting (letter form)	79.47% (CNN) 86.22% (ViT)
Ghouse et al., 2022	Non-Discrimination Regularization in Rotational Region Convolutional Neural Network (NDR-R2CNN)	Offline handwriting	98.2%
Zolna et al., 2019	Recurrent Neural Network model (RNN).	Online handwriting	>90% diagnosed as dysgraphia
Proposed CNN	Sequential CNN (three Convolutional layers (SN)	Offline handwriting	86%
	DAG network	Offline handwriting	87.75%

and it has been successfully executed with automated feature extraction and only requires minimal time to complete the training and testing. However, the proposed CNN-based model demonstrated noteworthy performance, which is substantial enough to warrant its integration with another model in the future.

CONCLUSION

This paper focused on the classification performance of SN and DAGN on potential dysgraphia handwriting images and compared both network models. According to the experimental results, this study demonstrates that DAGN can significantly improve classification performance on dysgraphia screening using children's handwriting products. This finding is consistent with previous research that found skip connections in DAGN improved classification performance by addressing the issue of vanishing gradients during backpropagation. In addition, skip connections provide a different route that acts as a shortcut, preventing information loss and distortion across the network. Hence, DAGN has been proven to be a useful tool for image classification tasks, as it uses skip connections to improve training speed, accuracy, and stability. Besides, this improved performance may be contributed by the DAGN model's layer structure, activation functions, or connection patterns, which enable more efficient feature extraction and representation. The proposed CNN-based network model shows significant performance in children's handwriting classification and could be a considerable network structure to be integrated with another model to assist the dysgraphia detection process.

The performance in this study is based on a specific dataset, limiting the generalizability of the models. Additional validation on larger and more diverse data sets is required to establish the validity of the findings. Future research could investigate the interpretability of model predictions to comprehend the underlying characteristics and patterns that contribute to dysgraphia detection in handwriting. Incorporating additional features and conducting comparative analyses with other advanced architectures or traditional algorithms would extend dysgraphia detection research while improving model precision and reliability. This study shows that using DAGN as a promising strategy for identifying dysgraphia-related handwriting symptoms has the potential to improve understanding of dysgraphia and stimulate the development of improved tools and interventions for people with this learning difference.

ACKNOWLEDGMENTS

This project is funded by the Malaysia Ministry of Higher Education Grant Scheme (FRGS), "A Formulation on Optimum Resnet-CNN Layer Architecture Based on Dilated Method for Dysgraphia Severity Classification"(Ref: FRGS/1/2021/ICT02/UITM/02/4). The authors thank members of the Research Intervention for Dysgraphia: Learning & Technology (RIDyLT) and the Centre for Electrical Engineering Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang, Malaysia, for their help and guidance during the fieldwork. Finally, the authors express their gratitude to Universiti Teknologi MARA, Cawangan Pulau Pinang, Malaysia, for their exceptional administrative support.

REFERENCES

Almisreb, A. A., Tahir, N. M., Turaev, S., Saleh, M. A., & Al Junid, S. A. M. (2022). Arabic handwriting classification using deep transfer learning techniques. *Pertanika Journal of Science and Technology*, *30*(1), 641–654. https://doi.org/10.47836/PJST.30.1.35

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*, Article 53. https://doi.org/10.1186/s40537-021-00444-8

Asselborn, T., Chapatte, M., & Dillenbourg, P. (2020). Extending the spectrum of dysgraphia: A data driven strategy to estimate handwriting quality. *Scientific Reports*, *10*(1), Article 3140. https://doi.org/10.1038/s41598-020-60011-8

Biotteau, M., Danna, J., Baudou, É., Puyjarinet, F., Velay, J. L., Albaret, J. M., & Chaix, Y. (2019). Developmental coordination disorder and dysgraphia: Signs and symptoms, diagnosis, and rehabilitation. *Neuropsychiatric Disease and Treatment*, *15*, 1873–1885. https://doi.org/10.2147/NDT.S120514

Chai, J., Zeng, H., Li, A., & Ngai, E. W. T. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, *6*, Article 100134. https://doi.org/10.1016/j.mlwa.2021.100134

Chung, P. J., Patel, D. R., & Nizami, I. (2020). Disorder of written expression and dysgraphia: Definition, diagnosis, and management. *Translational Pediatrics, 9*(Suppl 1), S46–S54. https://doi.org/10.21037/TP.2019.11.01

Dankovicova, Z., Hurtuk, J., & Fecilak, P. (2019, September 12-14). *Evaluation of digitalized handwriting for dysgraphia detection using random forest classification method*. [Paper presentation]. IEEE 17th International Symposium on Intelligent Systems and Informatics, Proceedings (SISY), Subotica, Serbia. https://doi.org/10.1109/SISY47553.2019.9111567

Deuel, R. K. (1995). Developmental dysgraphia and motor skills disorders. *Journal of Child Neurology*, *10*(1_suppl), S6-S8. https://doi.org/10.1177/08830738950100S103

Devi, A., & Kavya, G. (2023). Dysgraphia disorder forecasting and classification technique using intelligent deep learning approaches. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 120,* Article 110647. https://doi.org/10.1016/j.pnpbp.2022.110647

Devillaine, L., Lambert, R., Boutet, J., Aloui, S., Brault, V., Jolly, C., & Labyt, E. (2021). Analysis of graphomotor tests with machine learning algorithms for an early and universal pre-diagnosis of dysgraphia. *Sensors*, *21*(21), Article 7026. https://doi.org/10.3390/s21217026

Dimauro, G., Bevilacqua, V., Colizzi, L., & Di Pierro, D. (2020). TestGraphia, a software system for the early diagnosis of dysgraphia. *IEEE Access*, *8*, 19564–19575. https://doi.org/10.1109/ACCESS.2020.2968367

Ghouse, F., Paranjothi, K., & Vaithiyanathan, R. (2022). Dysgraphia classification based on the non-discrimination regularization in rotational region convolutional neural network. *International Journal of Intelligent Engineering and Systems*, *15*(1), 55–63. https://doi.org/10.22266/IJIES2022.0228.06

Kunhoth, J., Maadeed, S. A., Saleh, M., & Akbari, Y. (2023). Biomedical signal processing and control exploration and analysis of on-surface and in-air handwriting attributes to improve dysgraphia disorder diagnosis in children based on machine learning methods. *Biomedical Signal Processing and Control*, *83*, Article 104715. https://doi.org/10.1016/j.bspc.2023.104715

Masood, F., Khan, W. U., Ullah, K., Khan, A., Alghamedy, F. H., & Aljuaid, H. (2023). A hybrid CNN-LSTM random forest model for dysgraphia classification from hand-written characters with uniform/normal distribution. *Applied Sciences*, *13*(7), Article 4275. https://doi.org/10.3390/app13074275

Mohammed, A. B., Al-Mafrji, A. A. M., Yassen, M. S., & Sabry, A. H. (2022). Developing plastic recycling classifier by deep learning and directed acyclic graph residual network. *Eastern-European Journal of Enterprise Technologies*, *2*(10), 42–49. https://doi.org/10.15587/1729-4061.2022.254285

Qiao, J., Lv, Y., Cao, C., Wang, Z., & Li, A. (2018). Multivariate deep learning classification of alzheimer's disease based on hierarchical partner matching independent component analysis. *Frontiers in Aging Neuroscience*, *10*, Article 417. https://doi.org/10.3389/fnagi.2018.00417

Ramlan, S. A., Isa, I. S., Osman, M. K., Ismail, A. P., & Soh, Z. H. C. (2022). Investigating the impact of CNN layers on dysgraphia handwriting image classification performance. *Journal of Electrical and Electronic Systems Research*, *21*, 73–83. https://doi.org/https://doi.org/10.24191/jeesr.v21i1.010

Rosli, M. S. A., Isa, I. S., Ramlan, S. A., Sulaiman, S. N., & Maruzuki, M. I. F. (2021, August 27-28). *Development of CNN transfer learning for dyslexia handwriting recognition*. [Paper presentation]. 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia. https://doi.org/10.1109/iccsce52189.2021.9530971

Šafárová, K., Mekyska, J., & Zvončák, V. (2021). Developmental dysgraphia: A new approach to diagnosis. *The International Journal of Assessment and Evaluation*, *28*(1), 143–160. https://doi.org/10.18848/2327-7920/CGP/v28i01/143-160

Sihwi, S. W., Fikri, K., & Aziz, A. (2019). Dysgraphia identification from handwriting with support vector machine method. *Journal of Physics: Conference Series*, *1201*(1), Article 012050. https://doi.org/10.1088/1742-6596/1201/1/012050

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Vilasini, V., Rekha, B. B., Sandeep, V., & Venkatesh, V. C. (2022, August 11-12). *Deep learning techniques to detect learning disabilities among children using handwriting*. [Paper presentation]. Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT), Kannur, India. https://doi.org/10.1109/ICICICT54557.2022.9917890

Vaivre-Douret, L., Lopez, C., Dutruel, A., & Vaivre, S. (2021). Phenotyping features in the genesis of pre-scriptural gestures in children to assess handwriting developmental levels. *Scientific Reports*, *11*(1), Article 731. https://doi.org/10.1038/s41598-020-79315-w

Vlachos, F., & Avramidis, E. (2020). The difference between developmental dyslexia and dysgraphia: Recent neurobiological evidence. *International Journal of Neuroscience and Behavioral Science*, *8*(1), 1–5. https://doi.org/10.13189/ijnbs.2020.080101

Zolna, K., Asselborn, T., Jolly, C., Casteran, L., Johal, W., & Dillenbourg, P. (2019). The dynamics of handwriting improves the automated diagnosis of dysgraphia. *arXiv:1906.07576,* Article 1906.07576. https://doi.org/10.48550/arXiv.1906.07576