https://doi.org/10.24191/jeesr.v21i1.010

Investigating The Impact of CNN Layers on Dysgraphia Handwriting Image Classification Performance

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

***Abstract*—The diagnostic and detection process for distinguishing dysgraphia handwriting is vital in the intervention procedure to personalise the severity level of children's handwriting at an early stage. Currently, many deep learning methods and developments focus on other different domains such as handwriting signals and computer-based screenings that offers several limitations in time consuming and intricate procedures. Therefore, the dysgraphia handwriting classification using convolutional neural networks (CNNs) has seen a lot of success in the domain of handwriting image-based data that offers clear structured topology in the regular lattice of pixels in the dysgraphia handwriting patterns. However, due to the diversity and different characteristics of handwriting patterns, the convolution operations such as local connectivity layers and architectural framework in CNN model are numerous and invariant to generalize the features map in learning model classification. Thus, the comparative study of CNN layers is presented to investigate the impact of the different number of layers based on automated feature extraction for classifying dysgraphia and non-dysgraphia handwriting images. Experimentally, five CNN models that differ in structural architecture layers namely CNN-1, CNN-2, CNN-3, CNN-4, and CNN-5 are trained and validated using synthetic letter images dataset to observe the performance of each model. The CNN models are evaluated based on the confusion matrix of predicted and actual classes. Overall, the experiment shows high accuracy gained when the number of feature extraction layers is increased, which is five convolution layers outperformed others with significant training accuracy of 97.2% and 95.86% validation accuracy. The testing accuracy of 87.44% has approved that the proposed model is relatively accurate.**

***Index Terms*—Convolution Neural Network, Deep learning, Dysgraphia handwriting, Handwriting disorder.**

This manuscript is submitted on 7th June 2022 and accepted on 29th August 2022. This research is fully supported by FRGS grant, FRGS/1/2021/ICT02/UITM/02/4.

Siti Azura Ramlan, Iza Sazanita Isa, Muhammad Khusairi Osman, Ahmad Puad Ismail and Zainal Hisham Che Soh are with the School of Electrical Engineering, College of Engineering, University Teknologi MARA Cawangan Pulau Pinang, MALAYSIA (e-mail: sitiazura@uitm.edu.my, izasazanita@uitm.edu.my, khusairi@uitm.edu.my, ahmadpuad127@uitm.edu.my and zainal872@uitm.edu.my).

*Corresponding author
Email address: sitiazura@uitm.edu.my

I. INTRODUCTION

Dysgraphia is a hereditary learning problem that makes it difficult to accomplish writing tasks, resulting in lags that are not typical for their age and cognitive level [1][2]. Nevertheless, dysgraphia is no longer classified as a separate illness in the Diagnostic and Statistical Manual of Mental Disorders 5th edition (DSM-5), but rather as a "specific learning disability" [3]. In addition, the study of dysgraphia has always been known as an impairment in writing which is commonly detected as early as the beginning of school. Generally, the handwriting impairments can be observed at early school age by diagnosing letter formation which constrain the children from writing rapidly and irregular spacing of words and letters. Commonly, dysgraphia handwriting is characterised by reversed letters [1], [4], [5], due to the difficulty for remembering how to shape letters [4], [6], confusion in using lower-case or upper-case letters [6], difficulties to form written sentences with proper grammar and punctuation, frequent word omissions, incorrect word sequence, incorrect verb and pronoun usage, and word ending mistakes [1], [7]. However, there are many other signs of dysgraphia that have been observed among different age groups regarding the handwriting task as presented in Table I.

TABLE I
SIGNS OF DYSGRAPHIA PROVIDED BY UNITED STATES NATIONAL CENTER FOR LEARNING DISABILITIES [8]

Age group	Signs or symptoms
Pre-school children	• An awkward grip or body position when writing • Tire easily with writing • Avoidance of writing and drawing tasks • Written letters are poorly formed, inversed, reversed, or inconsistently spaced • Difficulty staying within margins
The school-aged child	• Illegible handwriting • Switching between cursive and print • Difficulty with word-finding, sentence completion, and written comprehension
The teenager and young adult	• Difficulty with written organization of thought • Difficulty with written syntax and written grammar that is not duplicated with oral tasks

Typically, the assessment for the detection of dysgraphia is discovered through the established procedure of the dyslexia screening session, such as an observation session by the expert or assessor and using the scoring methods [9], [10]. The assessments are administered through peer meetings and include a variety of assessments to determine a child's cognitive strengths and limitations in different areas such as phonemic awareness, visual-spatial abilities, and sound-letter identification [3,4]. For example, the United Kingdom (UK) implemented a screening test by using the Detailed Assessment of Speed of Handwriting (DASH) [2], and Netherland and French have a Concise Evaluation Scale for Children's Handwriting (BHK) as standard dysgraphia diagnosing test [13]. Meanwhile in Malaysia, the Ministry of Education has developed a checklist for monitoring the presence of suspected dysgraphia children by assessing prospective behaviours and symptoms using Senarai Semak Disleksia (ISD) [14]. In a nutshell, three factors are considered to identify the dysgraphia symptoms among children which are spelling and reading and writing abilities including the strengths and limitations. However, such conventional method has a limitation as it is highly strained since teachers must patiently observe and analyse the potential dysgraphia students [1]. Furthermore, the conventional assessment method necessitates the use of an expert to administer the test to children, and there is a high likelihood that the evaluation process will be incomplete due to distraction or inability to give complete attention. Meanwhile, many studies have been conducted in the community to raise awareness [15]–[18] and support dysgraphia in writing impairment [1], [19] as well as investigating its symptoms [6], [20], diagnosis[4], [5], and treatment[6]. Therefore, dysgraphia has recently become a favoured issue among scholars focusing on the diagnosis, detection, and intervention improvement for future enhancement of the quality of education among school children in general and learning disability specifically.

II. RELATED STUDIES

A. Dysgraphia Assessment and Screening Test

With recent advancement in technology, many researchers have proposed the best suitable method as screening tests by using a computer-based model and framework. As example, A. A. Abd Rauf et al. [18] proposed a support system for dyslexia and dysgraphia screening by comparing three tools namely the Web-based Screening System (e-ISD), the Perceptron-Based Learning Disability Detector (PLEDDOR) and the Lucid Adult Dyslexia Screening (LADS). Meanwhile, Mekyska et al. [21] suggested the handwriting proficiency screening questionnaire as a method for automating the diagnosis of this problem and determining the degree of difficulty (HPSQ). The study demonstrates that the proposed method was superior to the current way since it enables dynamic handwriting evaluation while processing the input and producing findings quickly. Furthermore, a similar finding was revealed by V. Zvoncak et al. [17] which proposed an improved computerised developmental dysgraphia using IWN (intra-writer and self-esteem normalisation methods). According to the study, the IWN approach reduced error rates but was ineffective due to the study's unfit participants. Moreover, S. Roseenblum and G. Dror [22] took a new approach by analysing handwriting output from Computerized Penmanship Evaluation Tool (ComPET) using machine learning methodology. However, the results indicated that handwriting traits are the most discriminatory and unstandardised in detecting dysgraphia. As a result, other researchers investigated ways to improve methods for identifying dysgraphia based on handwriting using various methodologies. Therefore, based on many proposed methods, the machine learning approach is highly significant to be used as dysgraphia symptoms detection among school children as early marker detection.

B. Identifying Dysgraphia Signs using Deep Learning and CNN models

In recent years, there has been an increasing research using CNN framework and model [23] which plays a vital role in many conventional image classification algorithms[24]–[26]. Over the past decade, researchers have established that the CNN model is widely used in image recognition [27], classification[24], [26], speech recognition[28], and many other applications [23]. As it has a learning capacity that has been substantially enhanced throughout the years by implementing depth and other structural improvements, the authors of [29] suggested a CNN-based model for handwritten digit recognition using the MNIST dataset tested on two convolutional layers. The study asserted that the proposed architecture could achieve a higher accuracy rate than SVM. Another study of CNN model employing MNIST dataset was proposed by [30] which used two convolutional layers with activation function ReLU and subsampling layers in the middle of the layer design. The study had demonstrated a higher classification accuracy of 99.37%.

Based on the literature, machine learning underwent significant developments with the potential to aid in dysgraphia detection. Machine learning has grown in popularity and has begun to pervade our daily lives, including the specialised field of impairment identification and assistive technology. This expert system has been studied by researchers using various inputs and approaches such as Support Vector Machine (SVM) [31]–[33] , Convolutional Neural Network (CNN) [34]–[37], Backpropagation [38], and AdaBoost Algorithm [39]. Overall, the summarized of research works conducted regarding dysgraphia detection and classification using machine learning models are presented in Table II.

Based on the review, the machine learning models, and CNN approach highlight that the investigation of dysgraphia detection and classification using machine learning model by utilising extracted children's written images as inputs are limited. Therefore, it is important to retrieve the characteristics of handwriting in images that can be used for conducting a dysgraphia assessment. Furthermore, researchers may discover the characteristic of handwriting images produced by dysgraphia children to improve the technique and neural network design. However, there is still limited study on classifying dysgraphia based on children's handwriting images.

TABLE II
SUMMARY OF CURRENT RESEARCH CONDUCTED FOR CLASSIFYING DYSGRAPHIA USING DEEP LEARNING MODELS AND CNN.

Study	Machine Learning model	Domain	Performances
Mekyska et al. 2017 [21]	CART (nonparametric learning algorithm)	Raw signal of kinematic features and nonlinear dynamic features	92.86%
Kurniawan [40]	Forward Chaining method	Handwriting images and characteristic	94.71%
Avishka [41]	Convolutional Neural Network (CNN)	Handwriting character	-
Kariyawasam [42]	Convolutional Neural Network (CNN)	Letter image (Sinhala)	90%
Samodro [38]	Backpropagation Method	Handwriting raw signal	84.7%
Sihwi [43]	Support Vector Machine (SVM)	Handwriting raw signal	-
Zvoncak [44]	Support Vector Machine (SVM)	Handwriting raw signal	85%
Mombach [45]	Optical Character Recognition (OCR) method	Mirrored and rotated letters	96%
Drotár [39]	AdaBoost, and Random Forest with stratified tenfold cross-validation	Writing signal	79.5%.

Therefore, this paper presents the investigation of CNN architecture with different feature extraction layers on dysgraphia handwriting image. The main contribution of this paper is a comparison of CNN layer architecture performance with the model's capability for classifying dysgraphia and non-dysgraphia handwriting images.

C. *CNN Framework and Architecture for Image Classification*

According to recent research, the most significant improvement in CNN performance has been realised by replacing the traditional layer structure with blocks [30]. The development of novel and effective block designs is currently one of the study themes in CNN architectures result in making architecture simple and more understandable. The concept of the block as a structural unit will continue to be used, and the CNN performance will be observed. A typical CNN design involves image feature extraction and classification and alternate composition layers of convolution and pooling, followed by one or more fully connected layers [23] as depicted in Fig. 1. The following are descriptions of the CNN building blocks that accommodate the various layers.

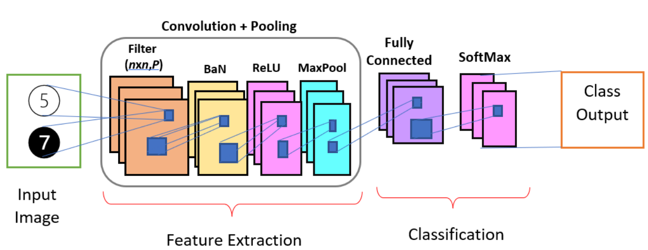


Fig. 1. Example of typical CNN architecture for image classification

1) *Convolutional Layer*

Features can be extracted from an image using filters (kernel) on the convolutional layer. It is possible to use a variety of filter sizes and numbers in this layer as needed [46]. Convolution kernels are another name for these filters. Using a convolutional algorithm, the input images are passed through a series of filters before being combined. Then, a feature map is created. The output of this operation is an integer, and the process is repeated until the entire image has been processed. The next layer will receive this image information.

2) *Batch Normalization (BaN)*

The activations and gradients propagating through a network are normalised by batch normalisation layers, making network training a more straightforward optimisation task [23]. It speeds up network training and reduces susceptibility to network initialisation, using batch normalisation layers between convolutional layers and nonlinearities, such as ReLU layers.

3) *Activation Layer (ReLU)*

The activation function in a neural network is responsible for converting the node's summed weighted input to the node's activation or output for that input. The rectified linear activation function, or ReLU, is a piecewise linear function that outputs the value directly if the input is positive; otherwise, its output is zero [46]. The rectified linear unit is the most widely used as an activation function (ReLU).

4) *Pooling Layer (MaxPool)*

Down sampling is occasionally used after convolutional layers (with activation functions) to minimise the spatial size of the feature map and eliminate superfluous spatial information [46]. Down sampling allows adding more filters to deeper convolutional layers without increasing the amount of computing required per layer. Max pooling is one method of down sampling. The max-pooling layer returns the maximum values of rectangular input areas.

5) *Fully connected layer (FC)*

One or more fully connected layers could be implemented after the convolutional and down-sampling layers at the end of layer composition. A fully connected layer is functioning in which all the neurons in the previous layers are connected. This layer aggregates all the information learnt by the preceding layers throughout the image to identify the larger patterns [46]. The final fully connected layer integrates the characteristics for image classification and classifies the input images.

6) *SoftMax Layer*

The fully connected layer's output is normalised using the SoftMax activation function. The SoftMax layer produces a set of positive integers that add up to one, where the classification layer can be utilised as classification probabilities [47].

Based on the CNN architectures and configurations, classification of dysgraphia and non-dysgraphia could be developed using the block structure of architecture. Then, the performance of each architecture that has a different number of local connectivity layer designs is observed.

III. EXPERIMENTAL SETUP

This section discusses the entire process of the proposed methodology in detail. Fig. 2 illustrates an overall proposed work in this experiment. In general, the method comprises three (3) main stages: dataset preparation, pre-processing, and network comparison. Firstly, the dataset preparation has been made, dividing the dataset into training and testing purposes. The pre-processing stage involves rotating and resizing images to obtain a balancing dataset for both classes of dysgraphia and non-dysgraphia. Lastly is the network comparison stage, which comprises the development of the CNN layer, classification of dysgraphia and non-dysgraphia handwriting images, and evaluation of the performance using a confusion matrix. Finally, the data from the network evaluation are compared and presented. The following sub-sections will discuss each part of the methodology in detail:

A. Dataset Preparation

The available public dataset prepared by [48] was used for classifying dysgraphia handwriting images into two classes. The dataset contains 315,021 dysgraphia and non-dysgraphia handwriting images, divided into 83,225 non-dysgraphia and 231,796 dysgraphia handwriting images. The dysgraphia handwriting images contain 96444 reversed and 135352 corrected letters. For both classes, the original dataset is 28 x 28, 31 x 31 and 29 x 29 with varying size pixels and an unbalanced number of image data. Then, the dataset consists of letters ‘a’, ‘b’, ‘c’, ‘d’, ‘f’, ‘g’, ‘h’, ‘j’, ‘k’, ‘m’, ‘p’, ‘q’, ‘r’, and ‘z’ which are commonly reversed and corrected by students [49].

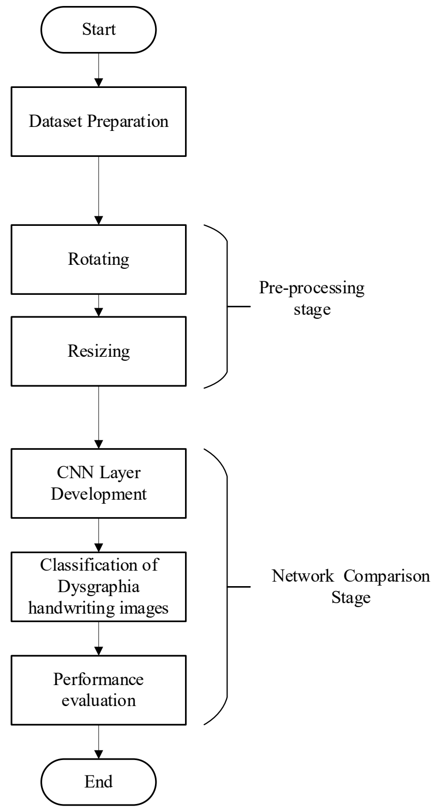


Fig. 2. Overall proposed work for CNN model development

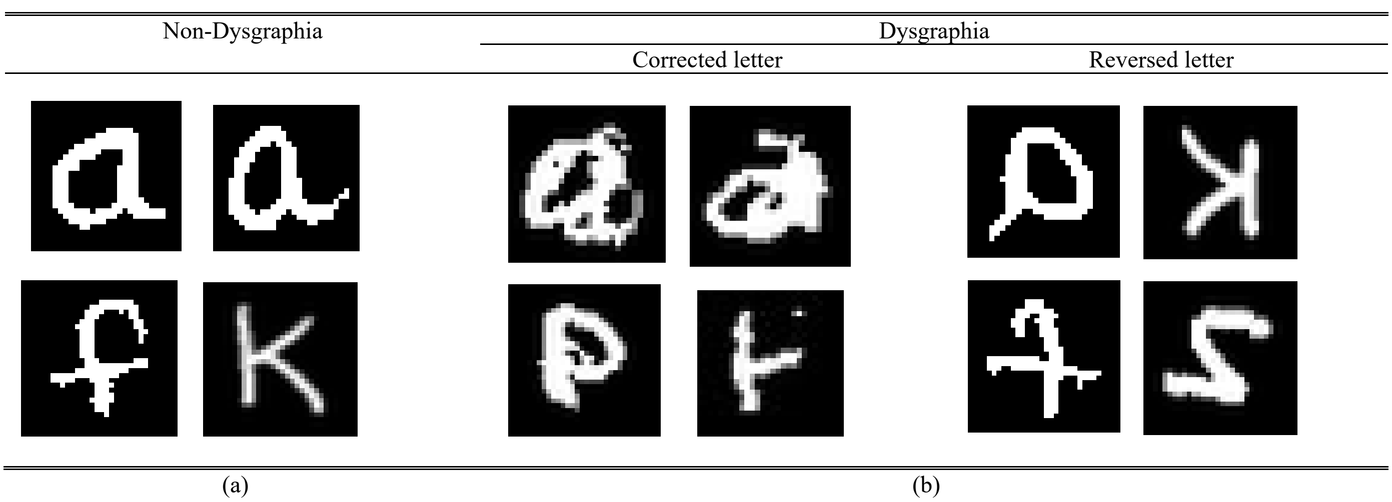


Fig. 3. Sample images of handwriting image by (a) non-dysgraphia and (b) dysgraphia that contains corrected and reversed letters

B. Pre-processing stage

Since the number of datasets distributed across both classes of dysgraphia and non-dysgraphia was imbalanced, the classification predictions could lead to bias for the classification between classes. Therefore, image pre-processing is conducted to balance the number of images for each class. To achieve a balanced and standardised input image size of 32 x 32 pixels, this study used image rotation of 5 degrees and 10 degrees, as well as image resizing. Table III shows the number of datasets configurations after the pre-processing process. The images of datasets are randomly split into training and testing for all CNN models based on 80:20 ratio as proposed by [47]. The training dataset was then divided into two parts: 70% training and 30% testing. 160,172 images were used for training, 68,644 for validation, and 39,114 images were used for testing out of a total of 267,930 images.

TABLE III
DATASET ARRANGEMENT FOR TRAINING, VALIDATION, AND TESTING

Dataset Class	Training	Validation	Testing
Dysgraphia	80086	34322	19557
Non-dysgraphia	80086	34322	19557
Total	160,172	68,644	39,114

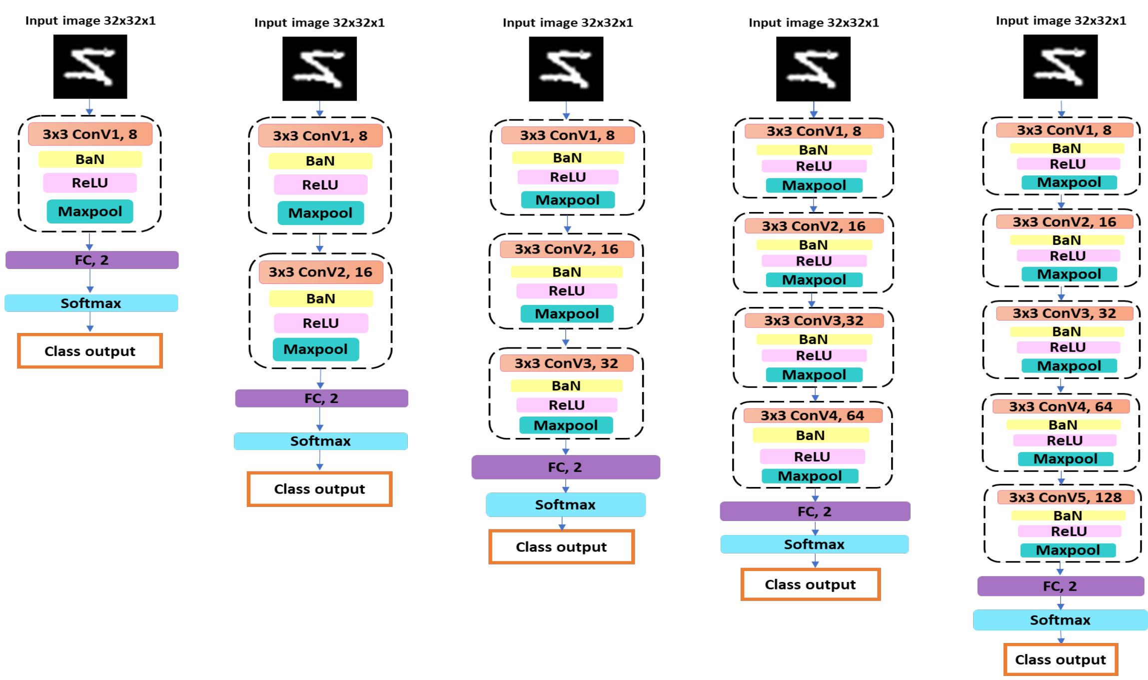
C. *CNN Layer Development*

In this study, five different architectures of CNN models were developed, and all architectures have different layer designs to observe the performance capability for classifying two classes of handwriting images. All the proposed CNN architectures used the same dataset input prepared to perform the features extraction and classification on one channel input image of 32 x 32 size. Fig. 4 shows the proposed five architectures of CNN models, with convolutional layers, batch normalisation (BaN) layers, ReLU, pooling layers, fully connected layers, Softmax layers, and finally, output layers. All architectures have a fully connected layer and a Softmax layer in their classification network structure. Fig. 3(a) is the simplest network architecture with one convolutional layer and max-pooling layer with a reLU activation function. Fig. 3(b) shows a network with two convolutional layers, two batches of normalisation, two ReLU layers, and one Maxpooling layer. As a features extraction network for CNN-3, Fig. 3(c) illustrates an architecture with three convolutional layers, three batches of normalisation layer, three ReLU, and two Maxpooling layers. Meanwhile, Fig. 3(d) has the same network series as Fig. 3(b), but the difference is the number of layers and filter, which has 19 layers. Four convolutional layers with 64 filters for the fourth convolutional layer, four batches of normalisation layer, four ReLU, and three Maxpooling 3x3 kernel size and 32 filters make up this architecture. The fifth architecture as shown in Fig. 3(e), has 23 layers, five of which are convolutional, and the fifth layer has 128 filters. The batch normalisation layer is stacked after the convolutional layer and consecutive with ReLU as an activation function. Batch normalisation is a method for increasing the speed and stability of artificial neural networks by normalising the inputs to the layers. The stride on each convolutional layer is 1x1, and the stride in the pooling layer is 2x2 with no padding. Experimentally, different networks of features extraction in all CNN models were used to observe the performance effect on accuracy and error rate value.

Some parameters can obtain the best performance of accuracy in training, namely optimizer, learning rate, and epoch. Depending on the optimizer's formulation, optimizers can be described as a mathematical function that updates the network's weights given the gradients and additional information [51]. For instance, SGD, ADAM, RMSProp, Adagrad, Adadelta, and Adamax. The increment where weights are increased during training is known as a learning rate. While epochs indicate the number of times, the neural network has processed all training data during training. Based on success work by H. Hafeez [50], this experiment has used the same parameter setting for all CNN architectures as tabulated in Table IV [50]

TABLE IV
PROPOSED PARAMETERS SETTINGS OF PLAIN CNN ARCHITECTURE

Parameter	Value
Optimizer	Stochastic Gradient Descent with Momentum 'sgdm'
Learning Rate	0.001
Epochs	8
Iteration per epoch	1251
Frequency	30 iterations



(a) (b) (c) (d) (e)

Fig. 4. The proposed CNN architectures for dysgraphia handwriting classification (a) CNN-1 (b) CNN-2 (c) CNN-3 (d) CNN-4 (e) CNN-5

D. Classification of Dysgraphia Handwriting Images

This experiment used the MATLAB 2021a environment and equipment with an Intel® CoreTM i5-10500H CPU running at 2.50GHz and an NVIDIA GeForce RTX 3060 GPU. Five architectures of convolutional neural networks are developed and observed with suitable architecture designs in various layers and learning parameter setups to maximise recognition performance and accomplish the class prediction of handwriting.

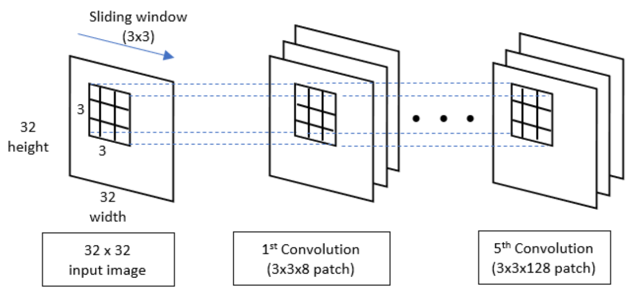


Fig. 5. Features map in convolutional layers

Fig. 5 shows the features map in each convolutional layer involved in this experiment. In CNN training, the input image is patched into 3x3 and slid along the image (from left to right) according to stride length in the convolution layer of the neural network. In this experiment, the CNN has multiple convolution layers from 1 to 5, and each convolutional layer generates many alternate convolutions, so the weight matrix is a tensor of $3 \times 3 \times n$, where n is the number of convolutions. As the convolution layer generates 8 convolutions by sliding a 3x3 window, the parameters generated for each CNN model is remarkably fewer parameters than a fully connected network (32x32) as presented in Table V. Since the number of parameters is independent of the size of the original image, CNN can be run on any image size and the number of parameters did not change in the convolution layer.

TABLE V
PARAMETERS GENERATED ON FEATURES MAP FOR EACH CNN MODELS

Model	Features Map parameters
CNN-1	72
CNN-2	144
CNN-3	288
CNN-4	576
CNN-5	1152

E. Performance Evaluation using Confusion Matrix

An analysis will be conducted in the evaluation part. This performance evaluation aims to ascertain the effectiveness of a neural network model in classifying the desired classes. The performance metrics are used in this study to monitor and quantify the CNN model's performance during training, validation, and testing. Numerous classification metrics can assess model performance, but this research focuses exclusively on each architecture's accuracy and error rate. Both metrics' performance was determined using the binary class confusion matrix. A confusion matrix is a tabular representation of the model predictions and their corresponding ground-truth labels. Table VI depicts the confusion matrix for the model architecture proposed in this study and represents two classes.

TABLE VI
CONFUSION MATRIX

		Predicted	
		Dysgraphia	Normal
Actual	Dysgraphia	TP	FP
	Normal	FN	TN

a. True Positive (TP): Actual dysgraphia class was correctly predicted as dysgraphia class
b. False Negative (FN): Actual normal class but was incorrectly predicted as dysgraphia class
c. True Negative (TN): Actual normal class was correctly predicted as normal class
d. False Positive (FP): Actual dysgraphia class but was predicted as normal class

The accuracy values are observed during training and testing execution to determine the performance of all proposed CNN models. The results of training, validation, and testing accuracy were calculated based on Precision, Recall and F1 Score for each output class as expressed by Equations 1 to Equation (3) respectively.

$$Accuracy=\frac{TP+TN}{TP+FP+TN+FN}\times 100\% \quad (1)$$

Precision is a confusion matrix that measures the accuracy relative to the prediction of a specific class. It is calculated as the ratio of the True Positives of the class in question to the sum of its True Positives and False Positives. In other words, precision is the number of correct positive predictions. For example, in this study, this metric will determine whether or not all of the handwriting that has been labelled as dysgraphia is dysgraphia. The precision is Equation (2).

$$Precision=\frac{TP}{TP+FP}\times 100\% \quad (2)$$

Recall or sensitivity is the process of evaluating the performance of the ground truths for positive outcomes. It is how well predicted outcome is positive when the result is positive. For example, in this research, from all the dysgraphia handwriting images, how many were the model able to label its class correctly.

$$Recall=\frac{TP}{TP+FN}\times 100\% \quad (3)$$

The harmonic mean achieved by the model is the F1 Score, which indicates that the model has vital Precision and Recall values. F1-score is the Harmonic mean of the Precision and Recall, as shown in (4).

$$F1\ Score=2\times\frac{Precision\times Recall}{Precision+Recall}\times 100\% \quad (4)$$

The loss function is shown in Equation (5) where N is the number of samples, K is the number of classes, w_i is the weight for class *i*, t_{ni} is the indicator that the n^{th} sample belongs to the i^{th} class, and y_{ni} is the output for sample *n* for class *i*, which in this case, is the value from the Softmax function. In other words, yni is the probability that the network associates the n^{th} input with class *i*.

$$loss = -\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{K} w_i t_{ni} \ln Y_{ni} \qquad (5)$$

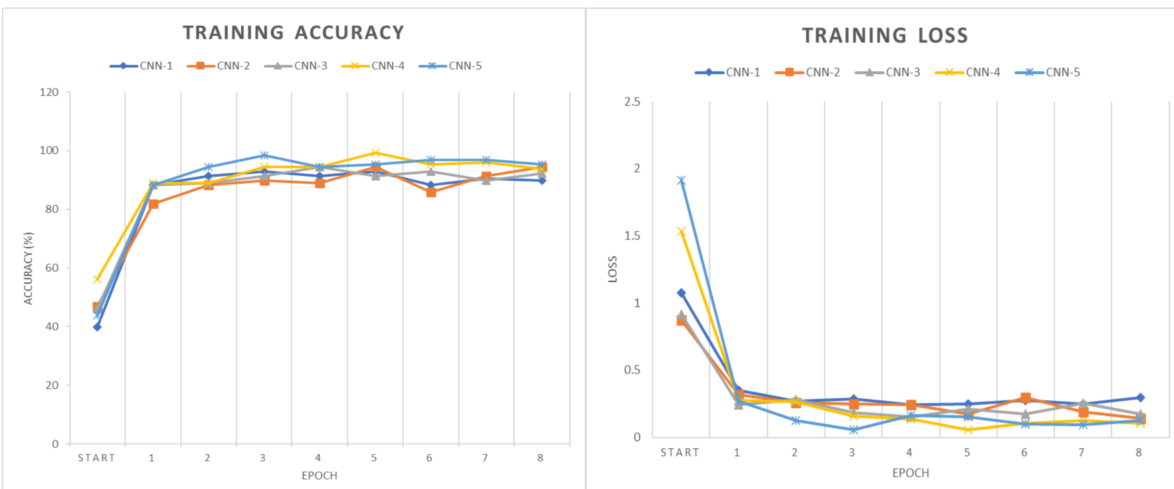
IV. RESULTS AND DISCUSSION

This section analyses the model performance in terms of accuracy for all five architectures tested in this study. To obtain a comprehensive evaluation, we examined the performance of all five CNN models on different layer architectures. The five architectures presented in Section III are successfully implemented, and the simulation produces validation values and losses for the graph. The training, validation accuracy, and loss contingency of the proposed architecture are shown in Fig. 6 and Fig. 7.

A. CNN Training and Validation Performances of Accuracy and Loss

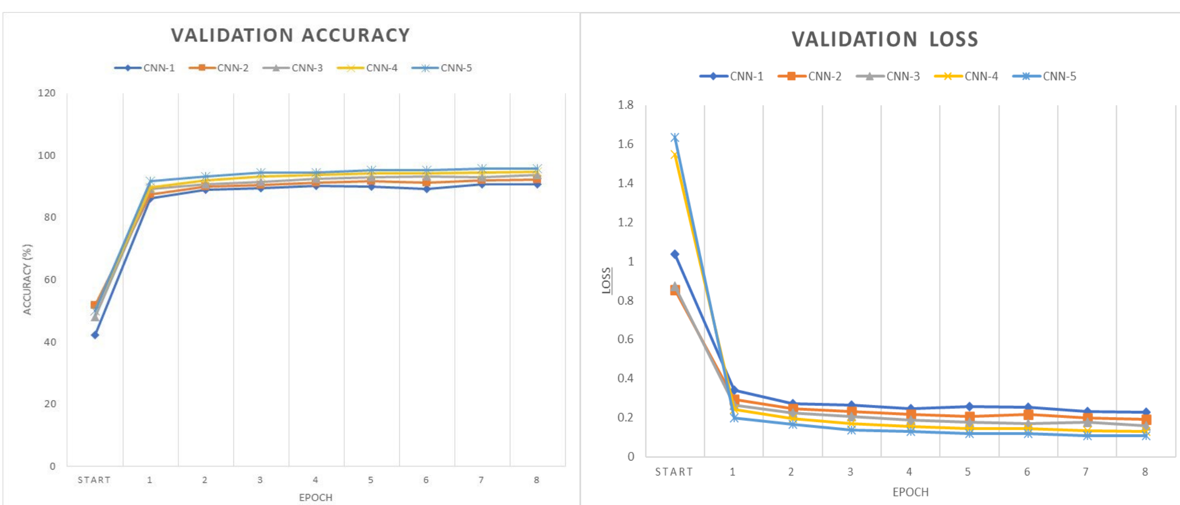
The training progress graph in Fig. 6(a) depicts five different CNN architectures with different convolutional layers whilst, the model performance in term of loss progress is presented in Fig. 6(b).

The accuracy of each design architecture was observed through the validation and loss graphs as illustrated in Fig. 7(a) and Fig. 7(b), and all training parameters are constant for each architecture using the same dataset. Based on the progress graphs, CNN-5 architecture showed the highest accuracy value and less data loss than other architecture graphs for training and validation. The experiment results show that more layers of feature extraction are likely to be a better architecture and show good performance for classifying dysgraphia or not having dysgraphia.



(a) (b)

Fig. 6. Performance of CNN models (a)Training Accuracy and (b) training loss



(a) (b)

Fig. 7. Performance of CNN Models (a)Validation Accuracy and (b) Validation Loss

Tables VII illustrates the performance analysis of each proposed CNN model in terms of accuracy for all architectures. This table is provided to help analyses the optimal result and identify the optimal model among all proposed models. Starting with the CNN-1 model, the accuracy value has shown to be relatively good. However, when the convolution layer is added, the accuracy value improves and rises. It also indicates that performance improves by increasing the convolution layer based on the accuracy of training, validation and testing value.

TABLE VII
THE ACCURACY PERFORMANCE ANALYSIS SUMMARY

MODEL	CNN-1	CNN-2	CNN-3	CNN-4	CNN-5
Training	0.9100	0.9241	0.9427	0.9597	**0.9720**
Validation	0.9084	0.9223	0.9327	0.9487	**0.9586**
Testing	0.8158	0.8379	0.8600	0.8660	**0.8744**

TABLE VIII
SUMMARY OF PERFORMANCE ANALYSIS

		Training		Validation		Testing	
		Dysgraphia	Non-Dysgraphia	Dysgraphia	Non-Dysgraphia	Dysgraphia	Non-Dysgraphia
Precision	CNN-1	0.9079	0.9120	0.9063	0.9105	0.8111	0.8207
	CNN-2	0.9202	0.9281	0.9174	0.9274	0.8312	0.8449
	CNN-3	0.9339	0.9518	0.9290	0.9463	0.8482	**0.8727**
	CNN-4	0.9712	0.9487	0.9612	0.9368	0.8726	0.8597
	CNN-5	**0.9851**	**0.9596**	**0.9730**	**0.9449**	**0.8861**	0.8635
Recall	CNN-1	0.9124	0.9075	0.9109	0.9059	0.8234	0.8083
	CNN-2	0.9288	0.9195	0.9282	0.9164	0.8480	0.8278
	CNN-3	0.9527	0.9326	**0.9474**	0.9276	**0.8770**	0.8430
	CNN-4	0.9475	0.9719	0.9351	0.9623	0.8572	0.8748
	CNN-5	**0.9585**	**0.9855**	0.9432	**0.9738**	0.8593	**0.8895**
F1 score	CNN-1	0.9102	0.9093	0.9086	0.9077	0.8172	0.8116
	CNN-2	0.9245	0.9231	0.9228	0.9210	0.8396	0.8330
	CNN-3	0.9432	0.9409	0.9381	0.9356	0.8624	0.8529
	CNN-4	0.9582	0.9602	0.9466	0.9494	0.8625	0.8672
	CNN-5	**0.9709**	**0.9724**	**0.9566**	**0.9591**	**0.8687**	**0.8763**

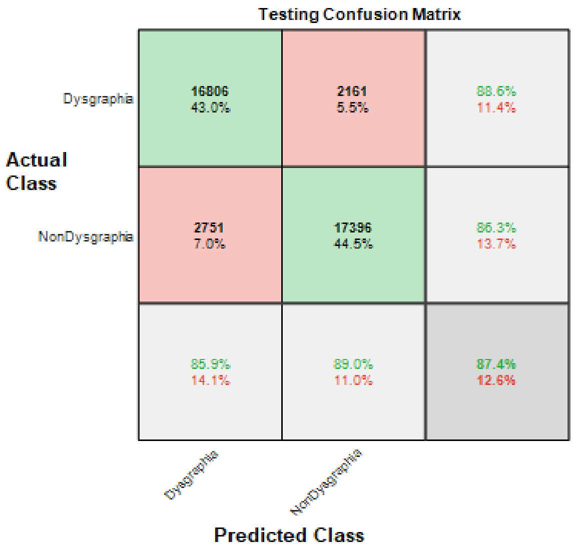
The precision, recall and f1 Score of each class regarding the training, validation and testing of the proposed CNN models for classifying dysgraphia handwriting are recorded in Table VIII. This table is provided to aid in discovering the best result and displaying the best model among all those architectures proposed. The precision percentages reflect how many samples of each class are correctly categorised. Meanwhile, the recall percentages display all the properly categorised samples from each class. Precision and recall percentages should be close to one to gain optimum performance; that is F1-score of one, will yield an accuracy of 100 per cent, but this is rarely the case with machine learning model. When the performance metrics are compared, precision, recall, and F1 Score provide more insight into the prediction.

B. Confusion Matrix of The Selected CNN model Performance

The confusion matrix is presented in Fig. 8, with the best accuracy percentage shown in CNN-5 architecture with accuracy values of 97.2% for training and 95.86% for validation. As a result, the accuracy for testing displayed the highest accuracy among five architectures, with an accuracy number of 87.44%. The number of layers affects the accuracy values, and feature extraction produces more accurate information as the layer depth increases.

This study demonstrates that the CNN-5 architecture is the most accurate model for identifying dysgraphia and non-dysgraphia. Using a total of 267,930 datasets, this model achieved the best validation accuracy of 95.86%. The CNN-5 comprises five blocks of layers, each having two 2d convolutional layers, a batch normalisation layer, a ReLU activation function layer, and a max-pooling layer. Then, the last fully connected layer functions as the output layer for two classifications and employs the SoftMax activation function. The network comprises the input size of 32 x 32 images and followed by a convolutional layer at filter size 3x3, which remained the same for all convolutional layers in the architecture CNN-5 model. 8 numbers of 3×3 kernels were applied, with the same size of kernel in both batch normalisation and ReLU layers.

The subsequent layer is max pooling with 2x2 kernel size stride 2; the feature maps were obtained and convoluted again. The second block of the layer is then sub-sampled with 16 filters followed by 32 filters in the third block. Forth block generated 64 filters, and the final blocks of feature extraction obtained 128 filters. The final fully connected layer used two neurons for the two-class classification. Within eight epochs, this model was trained and validated. To further categorise, fully connected layers were employed for labelling. The feature value was acquired and forwarded to the next layer to make the final prediction.



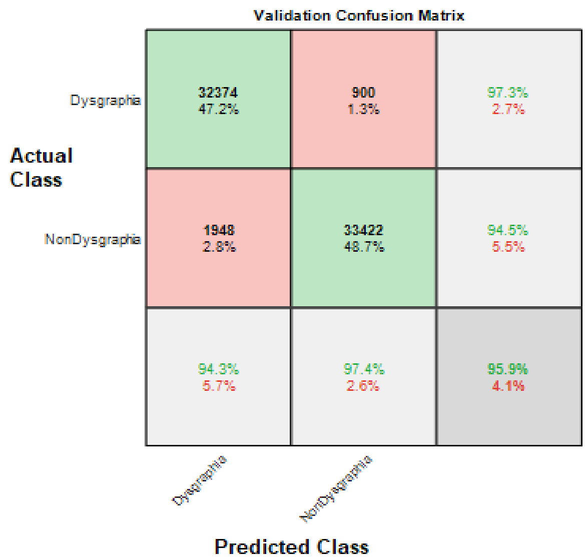


Fig. 8. Confusion matrix of the validation and testing of CNN-5 architecture

The feature information collected from the preceding convolution layers was reduced using the max-pooling layer. It facilitated the reduction of network size and enabled quick training. In order to extract the characteristics from an image properly, this experiment, CNN-5 employs the deepest convolutional layers compared to other architectural models. The structured CNN-5 model is successful enough to classify dysgraphia and non-dysgraphia handwriting accurately.

According to Table VII, the accuracy value increases proportionally with layer depth. Consequently, with the conditions that CNN-6 is adopting, the prediction of the outcome may achieve better performance with a high percentage of accuracy. Based on this work, we may estimate that performance will rise as the depth of the feature extraction layer increases.

V. Conclusion

This study aims to investigate the performance of CNN model in classifying dysgraphia and non-dysgraphia handwriting images by comparing with the different number of feature extraction layers. Based on the findings obtained, the performance was achieved in terms of accuracy, loss, precision, recall and F1 score for training, validation, and testing. The experiment involved 267,930 letter images divided into three purposes; 160,172 images used for training, 68,644 images for validation 39,114 data images was used for testing. All proposed CNN models were successfully developed and obtained the output with performance value respectively. The architectures are designed with different block compositions of layers comprising convolutional, batch normalisation, ReLU activation function, and max-pooling layer. From this study, the proposed architecture with five blocks of layers shows the higher performance with validation accuracy of 95.86% and training accuracy obtained of 97.2%. The best possible convolutional layers of the CNN model have been proven by testing accuracy of 87.44%, which is higher than other architectures. The results demonstrate that the biggest number of layers in CNN model is relatively accurate in identifying dysgraphia symptoms. Nevertheless, the dataset was only implemented for this conventional CNN layer designed and did not experiment with the other pretrained network models. Furthermore, the dataset was not investigated with various hyperparameters to see how the dysgraphia severity classification performed. Our future work will optimise the classification of dysgraphia handwriting image performance by employing different deep learning models.

ACKNOWLEDGEMENT

The authors would like to express their appreciation to the School of Electrical Engineering, Universiti Teknologi MARA, Cawangan Pulau Pinang, for the information and direction they received throughout the completion of this study. Special thanks to the FRGS Grant of Higher Education Ministry (FRGS/1/2021/ICT02/UITM/02/4) for the financial support provided throughout this research.

REFERENCES

[1] P. Chung and D. R. Patel, “Dysgraphia,” vol. 8, no. 1, pp. 27–36, 2015.

[2] M. Prunty and A. L. Barnett, “Understanding handwriting difficulties: A comparison of children with and without motor impairment,” *Cognitive Neuropsychology*, vol. 34, no. 3–4, pp. 205–218, 2017, doi: 10.1080/02643294.2017.1376630.

[3] American Psychiatric Association. “DIAGNOSTIC AND STATISTICAL MANUAL OF DSM-5 ™.” (5th ed). Washington, DC: American Psychiatric Publishing, 2013.

[4] 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.

[5] V. Z. Katarína Šafářová, Jiří Mekyska, “Developmental Dysgraphia: A New Approach to Diagnosis Katarína,” *Educational Psychology: Concepts, Research and Challenges*, vol. 28, no. 1, pp. 143–160, 2021, doi: https://doi.org/10.18848/2327-7920/CGP/v28i01/143-160.

[6] M. Biotteau *et al.*, “Developmental coordination disorder and dysgraphia: Signs and symptoms, diagnosis, and rehabilitation,” *Neuropsychiatric Disease and Treatment*, vol. 15, pp. 1873–1885, 2019, doi: 10.2147/NDT.S120514.

[7] M. McCloskey and B. Rapp, “Developmental dysgraphia: An overview and framework for research,” *Cognitive Neuropsychology*, vol. 34, no. 3–4, pp. 65–82, 2017, doi: 10.1080/02643294.2017.1369016.

[8] P. J. Chung, D. R. Patel, and I. Nizami, “Disorder of written expression and dysgraphia: definition, diagnosis, and management,” *Translational Pediatrics*, vol. 9. AME Publishing Company, pp. S46–S54, Feb. 01, 2020. doi: 10.21037/TP.2019.11.01.

[9] S. Rosenblum, P. L. Weiss, and S. Parush, “Handwriting evaluation for developmental dysgraphia: Process versus product,” *Reading and Writing*, vol. 17, no. 5, pp. 433–458, Jul. 2004, doi: 10.1023/B:READ.0000044596.91833.55.

[10] B. Engel-Yeger, L. Nagauker-Yanuv, and S. Rosenblum, “Handwriting performance, self-reports, and perceived self-efficacy among children with dysgraphia,” *American Journal of Occupational Therapy*, vol. 63, no. 2, pp. 182–192, 2009, doi: 10.5014/ajot.63.2.182.

[11] H. M. Al-Barhamtoshy and D. M. Eldeen Motaweh, “Diagnosis of Dyslexia using Computing Analysis,” *Journal of Engineering Technology*, vol. 6, no. 2, pp. 462–482, 2017, [Online]. Available: http://www.joetsite.com/wp-content/uploads/2017/07/Vol.-62-37-2017.pdf

[12] L. Fostick and H. Revah, “Dyslexia as a multi-deficit disorder: Working memory and auditory temporal processing,” *Acta Psychologica*, vol. 183, no. April 2017, pp. 19–28, 2018, doi: 10.1016/j.actpsy.2017.12.010.

[13] [de B. J. den B. B. Hamstra-Bletz L, “Concise evaluation scale for children’s handwriting,” 1987.

[14] I. S. Isa, W. N. Syazwani Rahimi, S. A. Ramlan, and S. N. Sulaiman, “Automated Detection of Dyslexia Symptom Based on Handwriting Image for Primary School Children,” *Procedia Computer Science*, vol. 163, pp. 440–449, 2019, doi: 10.1016/j.procs.2019.12.127.

[15] S. M. K.P, “Awareness on Learning Disabilities among Elementary School Teachers,” *Asian Journal of Education and Training*, vol. 2, no. 2, pp. 78–83, Dec. 2016, doi: 10.20448/journal.522/2016.2.2/522.2.78.83.

[16] M. Brown, “FireScholars Dysgraphia Part of the Disability and Equity in Education Commons, Elementary Education Commons, and the Special Education and Teaching Commons.” [Online]. Available: https://firescholars.seu.edu/ccplus

[17] Rabih Nabhan and Nouhad Kamel, “Technologically Enhanced Art Therapy: Introducing a New Synergistic Model to Enhance Morphological Awareness in Students with Dysgraphia,” *LingLit Journal Scientific Journal for Linguistics and Literature*, vol. 2, no. 1, pp. 1–8, Mar. 2021, doi: 10.33258/linglit.v2i1.416.

[18] A. A. Abd Rauf, M. A. Ismail, V. Balakrishnan, and K. Haruna, “Dyslexic Children: The Need for Parents Awareness,” *Journal of Education and Human Development*, vol. 7, no. 2, pp. 91–99, 2018, doi: 10.15640/jehd.v7n2a12.

[19] U. The and G. Of, *Identification of Developmental Dysgraphia By Handwriting*. 2018.

[20] F. Ashraf and N. Najam, “An epidemiological study of prevalence and comorbidity of non-clinical dyslexia, dysgraphia and dyscalculia symptoms in public and private schools of pakistan,” *Pakistan Journal of Medical Sciences*, vol. 36, no. 7, pp. 1659–1663, 2020, doi: 10.12669/pjms.36.7.2486.

[21] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, “Identification and Rating of Developmental Dysgraphia by Handwriting Analysis,” *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 235–248, 2017, doi: 10.1109/THMS.2016.2586605.

[22] S. Rosenblum and G. Dror, “Identifying Developmental Dysgraphia Characteristics Utilizing Handwriting Classification Methods,” *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 293–298, 2017, doi: 10.1109/THMS.2016.2628799.

[23] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, *A survey of the recent architectures of deep convolutional neural networks*, vol. 53, no. 8. Springer Netherlands, 2020. doi: 10.1007/s10462-020-09825-6.

[24] X. Lei, H. Pan, and X. Huang, “A dilated cnn model for image classification,” *IEEE Access*, vol. 7, pp. 124087–124095, 2019, doi: 10.1109/ACCESS.2019.2927169.

[25] L. H. Thai, T. S. Hai, and N. T. Thuy, “Image Classification using Support Vector Machine and Artificial Neural Network,” *International Journal of Information Technology and Computer Science*, vol. 4, no. 5, pp. 32–38, 2012, doi: 10.5815/ijitcs.2012.05.05.

[26] Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv, “Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification,” *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3840–3854, 2020, doi: 10.1109/TCYB.2020.2983860.

[27] Y. Ouyang, K. Wang, and S. Wu, “SAR Image Ground Object Recognition Detection Method based on Optimized and Improved CNN,” *Proceedings of 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2019*, no. Iaeac, pp. 1727–1731, 2019, doi: 10.1109/IAEAC47372.2019.8997680.

[28] M. Sidi Yakoub, S. ahmed Selouani, B. F. Zaidi, and A. Bouchair, “Improving dysarthric speech recognition using empirical mode decomposition and convolutional neural network,” *Eurasip Journal on Audio, Speech, and Music Processing*, vol. 2020, no. 1, pp. 1–7, 2020, doi: 10.1186/s13636-019-0169-5.

[29] R. Jantayev and Y. Amirgaliyev, “Improved handwritten digit recognition method using deep learning Algorithm,” *2019 15th International Conference on Electronics, Computer and Computation, ICECCO 2019*, no. Icecco, pp. 2019–2022, 2019, doi: 10.1109/ICECCO48375.2019.9043235.

[30] S. Ali, Z. Sakhawat, T. Mahmood, M. S. Aslam, Z. Shaukat, and S. Sahiba, “A robust CNN model for handwritten digits recognition and classification,” *Proceedings of 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications, AEECA 2020*, pp. 261–265, 2020, doi: 10.1109/AEECA49918.2020.9213530.

[31] J. Szalma and B. Weiss, “Data-Driven classification of dyslexia using eye-movement correlates of natural reading,” *Eye Tracking Research and Applications Symposium (ETRA)*, pp. 10–13, 2020, doi: 10.1145/3379156.3391379.

[32] A. Jothi Prabha and R. Bhargavi, “Predictive Model for Dyslexia from Fixations and Saccadic Eye Movement Events,” *Computer Methods and Programs in Biomedicine*, vol. 195, 2020, doi: 10.1016/j.cmpb.2020.105538.

[33] Z. Mahmoodin, K. Y. Lee, W. Mansor, and A. Z. A. Zainuddin, “Support vector machine with theta-beta band power features generated from writing of dyslexic children,” *International Journal of Integrated Engineering*, vol. 11, no. 3, pp. 42–50, 2019, doi: 10.30880/ijie.2019.11.03.005.

[34] F. Latifoğlu, R. İleri, and E. Demirci, “Assessment of dyslexic children with EOG signals: Determining retrieving words/re-reading and skipping lines using convolutional neural networks,” *Chaos, Solitons and Fractals*, vol. 145, 2021, doi: 10.1016/j.chaos.2021.110721.

[35] R. Ileri, F. Latifoglu, and E. Demirci, “New Method to Diagnosis of Dyslexia Using 1D-CNN,” *TIPTEKNO 2020 - Tip Teknolojileri Kongresi - 2020 Medical Technologies Congress, TIPTEKNO 2020*, pp. 1–4, 2020, doi: 10.1109/TIPTEKNO50054.2020.9299241.

[36] S. Zahia, B. Garcia-Zapirain, I. Saralegui, and B. Fernandez-Ruanova, “Dyslexia detection using 3D convolutional neural networks and functional magnetic resonance imaging,” *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105726, 2020, doi: 10.1016/j.cmpb.2020.105726.

[37] O. L. Usman and R. C. Muniyandi, “CryptoDL: Predicting dyslexia biomarkers from encrypted neuroimaging dataset using energy-efficient residue number system and deep convolutional neural network,” *Symmetry*, vol. 12, no. 5, 2020, doi: 10.3390/SYM12050836.

[38] P. W. Samodro, S. W. Sihwi, and Winarno, “Backpropagation implementation to classify dysgraphia in children,” *Proceeding - 2019 International Conference of Artificial Intelligence and Information Technology, ICAIIT 2019*, pp. 437–442, 2019, doi: 10.1109/ICAIIT.2019.8834520.

[39] P. Drotár and M. Dobeš, “Dysgraphia detection through machine learning,” *Scientific Reports*, vol. 10, no. 1, pp. 1–11, 2020, doi: 10.1038/s41598-020-78611-9.

[40] D. A. Kurniawan, S. W. Sihwi, and Gunarhadi, “An expert system for diagnosing dysgraphia,” *Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017*, vol. 2018-Janua, pp. 468–472, 2018, doi: 10.1109/ICITISEE.2017.8285552.

[41] I. Avishka, K. Kumarawadu, A. Kudagama, M. Weerathunga, and S. Thelijjagoda, “Mobile App to Support People with Dyslexia and Dysgraphia,” *2018 IEEE 9th International Conference on Information and Automation for Sustainability, ICIAfS 2018*, 2018, doi: 10.1109/ICIAFS.2018.8913335.

[42] R. Kariyawasam, M. Nadeeshani, T. Hamid, I. Subasinghe, and P. Ratnayake, “A Gamified Approach for Screening and Intervention of Dyslexia, Dysgraphia and Dyscalculia,” *2019 International Conference on Advancements in Computing, ICAC 2019*, pp. 156–161, 2019, doi: 10.1109/ICAC49085.2019.9103336.

[43] S. W. Sihwi, K. Fikri, and A. Aziz, "Dysgraphia Identification from Handwriting with Support Vector Machine Method," *Journal of Physics: Conference Series*, vol. 1201, no. 1, 2019, doi: 10.1088/1742-6596/1201/1/012050.

[44] V. Zvoncak, J. Mekyska, K. Safarova, Z. Smekal, and P. Brezany, "New approach of dysgraphic handwriting analysis based on the tunable Q-factor wavelet transform," *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2019 - Proceedings*, pp. 289–294, 2019, doi: 10.23919/MIPRO.2019.8756872.

[45] J. Mombach, C. Ferreira, J. Felix, R. Salvini, and F. Soares, "Mirrored and Rotated Letters in Children Spellings: An Automatic Analysis Approach," *Canadian Conference on Electrical and Computer Engineering*, vol. 2020-Augus, pp. 4–7, 2020, doi: 10.1109/CCECE47787.2020.9255765.

[46] G. Habib and S. Qureshi, "Optimization and acceleration of convolutional neural networks: A survey," *Journal of King Saud University - Computer and Information Sciences*, 2020, doi: 10.1016/j.jksuci.2020.10.004.

[47] H. M. Najadat, A. A. Alshboul, and A. F. Alabed, "Arabic Handwritten Characters Recognition using Convolutional Neural Network," *2019 10th International Conference on Information and Communication Systems, ICICS 2019*, no. January, pp. 147–151, 2019, doi: 10.1109/IACS.2019.8809122.

[48] M. S. A. bin Rosli, I. S. Isa, S. A. Ramlan, S. N. Sulaiman, and M. I. F. Maruzuki, "Development of CNN Transfer Learning for Dyslexia Handwriting Recognition," no. August, pp. 194–199, 2021, doi: 10.1109/iccsce52189.2021.9530971.

[49] J. P. Fischer and C. Luxembourger, "A synoptic and theoretical account of character (Digits and capital letters) reversal in writings by typically developing children," *Education Sciences*, vol. 8, no. 3. MDPI AG, Sep. 01, 2018. doi: 10.3390/educsci8030137.

[50] H. Hafeez, C. A. Abbas, and M. U. Khan, "Improved Recognition of Hand-Written Digits," *UW J. Sci. Technol. Vol. 5 28-33 ISSN 2523-0123 2616-4396*, vol. 5, no. Vol 5 No 1 (2021): Vol 5 (2021): University of Wah Journal of Science and Technology (UWJST), pp. 28–33, 2021.

[51] D. J. S. Naidu and T. M. Rafi, "Handwritten Character Recognition Using Convolutional Neural Networks," *Int. J. Comput. Sci. Mob. Comput.*, vol. 10, no. 8, pp. 41–45, 2021, doi: 10.47760/ijcsmc.2021.v10i08.007