IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 2, April 2026, pp. 1418~1427
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i2.pp1418-1427

Novel convolution neural network model for dysgraphia-affected handwriting classification

Nisha Vanjari[1,2], Prasanna J. Shete[1]
[1]Department of Computer Engineering, K. J. Somaiya School of Engineering, Mumbai, India
[2]Department of Computer Engineering, K. J. Somaiya Institute of Technology, Mumbai, India

Article Info

Article history:

Received May 16, 2025
Revised Jan 8, 2026
Accepted Jan 25, 2026

Keywords:

Convolution neural network
Deep learning
Dysgraphia
Handwriting classification
Learning disability

ABSTRACT

It is estimated that 10% of the population in the world suffers from learning disabilities like dyslexia, dysgraphia, and dyscalculia. Learning disabilities are neurological disorders in which children struggle with reading, writing and mathematical skills. Dysgraphia disorder impacts on writing abilities of students and thus may be a hurdle in their learning and evaluation of subject matter. Hence early detection/prediction of learning disability (LD) in school going children will greatly help in providing necessary accommodations so as to ease their future learning curve. In recent years researchers have used several deep learning algorithms that produce automated and trained models which can be useful in the handwriting classification. To properly capture the distinct handwriting inconsistencies linked to dysgraphia, this study contains experiments that determine how various convolution neural network (CNN) model layers contribute to performance. To address it, this research focused on the improved novel model based on CNN and targeted dysgraphia English handwriting classification with 98% accuracy with 102,691 trainable parameters. The model is trained on both normal and dysgraphia-affected handwriting, increasing its accuracy in identifying individual differences.

Corresponding Author:

Nisha Vanjari
Department of Computer Engineering, K. J. Somaiya Institute of Technology
33/A, Nandadeep CHS, Chembur, Mumbai, India
Email: nvanjari@somaiya.edu

1. INTRODUCTION

It is estimated that 10% of the population in the world suffers from learning disabilities like dyslexia, dysgraphia, and dyscalculia [1]–[4]. Primary impact of dysgraphia, a learning disability (LD), is on an individual's capacity for written expression. Dysgraphia and particular learning deficit in written expression refer to people who, despite proper education, write below their cognitive level and age. Dysgraphia is a specialized LD that makes it difficult to duplicate both letters and numbers. Depending on the child's age and developmental stage, dysgraphia can present a variety of symptoms. Multiple cues must be taken into account for an accurate diagnosis of dysgraphia, which presents major obstacles. As handwriting is a way to convey thoughts and emotions, many researchers have worked on handwriting character recognition systems to analyze images of written material and try to find the pattern from it. Researchers have used several deep learning algorithms like convolution neural network (CNN), LeNet5, and customized CNN that can be useful in predicting the symptoms of learning disabilities and produce automated and trained models that significantly minimize the need of human intervention [5]–[9].

CNNs are built from a multi-layered feed-forward neural network, which consists of stacking several hidden layers in a specific order. With its sequential architecture, CNN is able to acquire knowledge about hierarchical characteristics. CNN models are based on mathematical operations on every layer and the model performance can be improved by tuning the parameters. For developing the custom model, some researchers proposed architectural changes to improve the performance of model. With this experience, in order to classify dysgraphia symptoms using handwritten alphabet images, this study will compare basic CNN models with our proposed model that comprises hyper parameter tuning and architectural changes. Additionally, the accuracy and loss analysis of the CNN network will be used to validate the comparisons. This paper proposes, design and develop a novel nested convolution CNN model named Nested_Conv_CNN for dysgraphia symptoms classification and evaluate its performance on DysData handwriting dataset (based on NIST special database19 and available at Kaggle) [5]–[10]. Investigation was conducted to compare the novel model with the basic CNN model.

The rest of this paper is organized as follows. Section 2 presents the related work, which reviews recent studies. Sections 3 and 4 describe the dataset and the proposed methods. Section 5 presents the results and analysis. Section 6 provides the ablation study. Finally, section 7 concludes the paper.

2. RELATED WORK

A government study news release from 2015 states that 70–80% of children with learning impairments have reading difficulties. The most prevalent type of LD are dyslexia and dysgraphia. Worldwide, the estimated affected population ranges from 5–20%. In India, 15% of the population is estimated to be suffering from LD. Researchers have used several deep learning algorithms that can be useful in predicting the symptoms of learning disabilities like dyslexia and dysgraphia. The main objective of deep learning is to produce automated and trained models that significantly minimize the need for human interaction. Spoon *et al.* [4] attempted optical character recognition (OCR) tools like Tesseract (Google) to transcribe the writing in the photographs before turning to machine learning to detect dyslexia based on spelling errors and other textual characteristics and also created LeNet-5 to classify dyslexia handwriting. The 138,500 images in the dataset were pre-processed and enhanced with data before being fed into the network. Isa *et al.* [8] compared various CNN models and experimented CNN1, CNN2, CNN3, and LeNet-5 and got 87% accuracy in CNN1 model to detect the LD symptom. Recent studies on dysgraphia increasingly support the use of CNN-based pipelines and feature/classifier fusion [11]–[16]; however, they frequently lack comprehensive ablations and clear justifications for architectural and training decisions. Recent studies comparing architectures and releasing pediatric datasets highlight that accuracy depends on multi-scale stroke modelling and effective training configurations, rather than solely on deeper backbones.

Table 1 will give summary of relevant papers that used CNN technique to classify the handwriting to achieve dysgraphia handwriting classification. Mainly handwriting classification is performed on data collected in the form of online handwriting (data collected using digital tablet with digital pen or stylus) or offline handwriting (scanned images). In the Table 1, focus is only those papers which used handwritten data in offline mode or taken via tablet and digital pen. Researcher used DysData dataset has 78,275 samples for the normal class, 52,196 samples for the reversal, and 8,029 samples for the corrected class. In order to distinguish people with mild cognitive impairment (MCI) from healthy controls, Kawa *et al.* [10] focused on spatial and dynamic handwriting analysis. It emphasizes how well-reliable handwriting measures are as markers of cognitive impairment. Ribeiro *et al.* [17] published a study that presented the "bag of samplings" method, which uses recurrent neural networks to help diagnose Parkinson's disease by analyzing handwriting. The technique captured minute motor abnormalities, highlighting the promise of cutting-edge neural network techniques for medical diagnosis.

Zanuy *et al.* [18] group uses of biometrics of handwriting for health monitoring are covered in this study, along with potential future developments. It highlights the growing application of handwriting analysis outside of medical diagnosis. Using quantitative electroencephalography (qEEG) data and handwriting dynamics, Chai *et al.* [19] describe a novel method for identifying MCI. Combining various physiological data types appears to have advantages, as this multimodal approach greatly improves detection accuracy. According to Devi and Kavya [20], dysgraphia disorder can be predicted and classified using an intelligent deep learning approach. Deep learning approaches have been shown to be effective in managing learning problems, and the method has shown potential in early identification and intervention. Kunhoth *et al.* [21] worked to diagnose dysgraphia in children by utilizing a novel modified image dataset in conjunction with CNN feature extraction and classifier fusion. The potential of AI-driven diagnostic tools to increase the accuracy of developmental problem diagnosis is highlighted by this method [21], [22].

Handwriting has long been studied to detect conditions such as MCI, Parkinson’s disease, and dysgraphia, with early approaches relying on handcrafted or shallow features that often-limited generalization and lacked clinical validation. More recent CNN-based models demonstrated stronger performance, but many still relied on adult data rather than pediatric samples, reducing dataset diversity, and they were rarely tested in real screening contexts. To address these limitations, researchers began exploring ensemble and multimodal approaches. For instance, one study combined both online handwriting signals and offline images through conditional feature fusion, achieving 88.8% accuracy [23], while another used CNN-based feature and classifier fusion to push performance as high as 97.3% [24]. A recent review further highlighted the value of multimodal integration for developing more generalizable systems, though explicit ethical concerns such as privacy and informed consent remained largely unaddressed. Building on this trend, researchers have increasingly turned to transformer-based models for dysgraphia detection. A vision transformer (ViT) outperformed traditional CNNs such as visual geometry group (VGG16), ResNet50, and InceptionV3, reaching a macro F1-score of 0.92 in early dysgraphia tasks. These studies clearly demonstrate the power of deeper and more complex architectures, but they also highlight recurring challenges: transformers typically require large annotated datasets, which raises questions of dataset diversity and representativeness, as well as ethical concerns around data collection and resource accessibility.

Table 1. Review on modality used with used classifier and research opportunities

Reference	Modality	Dataset	Classifier	Group	Accuracy (%)	Research opportunities
[6]	Offline	NIST+special dataset19	CNN	Children	84	No feature fusion, minimal ablations
[7]	Offline	NIST+special dataset19	CNN	Children	87	Shallow features, limited generalization
[8]	Offline	Private	CNN	Children	>80	Adult data, not pediatric
[9]	Offline	NIST+special dataset19	CNN	Children	87	Adult data, not pediatric
[11]	Online	Private	SVM	Adult	96.3	Calls for richer, clinical datasets
[12]	Online	Private	Used web app	Adult	-	Non-pediatric, no glyph focus
[13]	Online	Private	RCNN	Elder	-	Dataset transparency issues
[14]	Offline	Private	App	Adult	-	High accuracy, little ablation
[15]	Offline	Private	K-DCT-DTL	6 to 12	>80	Supports stroke-texture analogy
[16]	Offline	Private	CNN	6 to 12	>84	Entropy feature idea transferable
[20]	Offline images	Dysgraphia dataset	ViT	Transformer	F1 =0.92	High compute, needs large data
[21]	Online+offline	Handwriting signals	CNN+RNN (fusion)	Multimodal	88.80	Limited generalization
[25]	Offline	Private	SVM, RF	6 to 12	~90	Can apply multimodal for severity levels

In contrast, our proposed Nested_Conv_CNN model is designed with practicality in mind. By stacking small kernels across convolutional blocks, the model is able to capture subtle handwriting details such as stroke thickness, slant, and reversals while drastically cutting down on trainable parameters and computation. This lightweight design enables accuracy levels comparable to larger models, but in a form that is both efficient and accessible, directly addressing real-world challenges such as limited pediatric dataset availability, modest hardware in educational environments, and the need for clinically practical solutions. Showcasing the broad applicability of deep learning in both occupational safety and educational health contexts, our model offers a strong balance between accuracy, speed, and practicality for large-scale dysgraphia detection.

3. DATASET

For research purpose, English handwritten dataset is considered. The dataset used for the experimentation was collected by Isa *et al.* [8], this dataset has total 138,500 images (78,275 images for the normal class, 52,196 for the reversal, and 8,029 for the corrected class). Following images will give more details about it. In Figures 1 to 3, dataset contains black and white alphabets images. Normal handwritten character means no symptom of dysgraphia but corrected handwritten character means it’s a beginning stage of dysgraphia as strokes are not firm and trying to rewrite or some time height and width is not appropriate like normal handwritten characters. Reversal/dysgraphic handwritten character means strong symptom of dysgraphia as child used to write mirror image characters and stroke pattern is always different for same character. We have applied augmentation to try to balance the dataset and also preprocessed to obtain noiseless clean images. For processing the image from dataset, the size of image is made 29×29 pixels and input to the proposed model.

Figure 1. Normal handwritten characters image

Figure 2. Corrected handwritten characters image

Figure 3. Reversal/dysgraphic handwritten characters image

To handle the imbalance in our handwriting dataset, we upsampled the minority classes and applied data augmentation to increase diversity and variability. Specifically, images were augmented with rotations up to 20°, horizontal and vertical shifts up to 20%, shearing up to 20%, and zooming up to 20%. These steps ensured that all classes had sufficient representation and helped the model learn more robust and generalized features, improving overall classification performance and to make our model more robust and generalizable.

4. METHOD

In proposed model of Figure 4 shows research majorly focused on convolution; as convolution has kernel/filter, stride to slide kernel on image and applying elementwise multiplication and then summation. By using convolution back-to-back, model will learn more complex hierarchical features. The proposed model has 4 convolution layers, 2 max pooling layers, flatten layer followed with fully connected or dense layer. Max pooling will be used for dimensionality reduction and it extracts maximum values from patch of each feature map and eliminates unwanted values to reduce the computational stress of model and also avoids overfitting problem. In the model, SoftMax activation function is used on output layer and rectified linear unit (ReLU) on internal layer as ReLU focuses on 0 to max value and it helps us to focused relevant data and send it to next layer for processing.

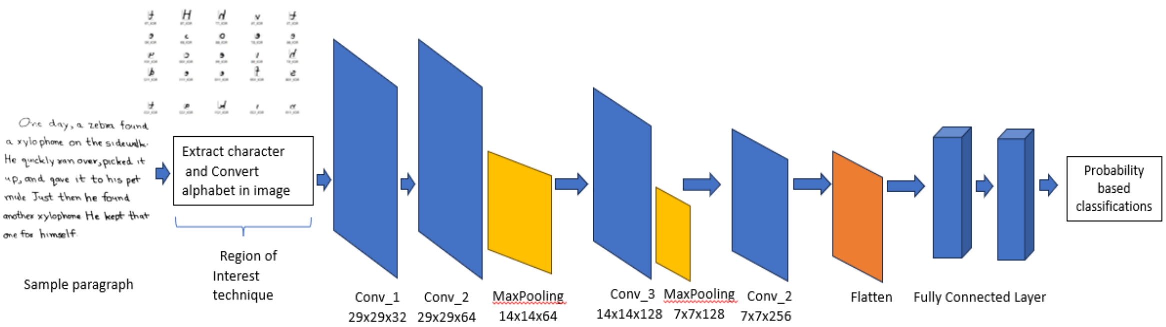

Figure 4. Proposed Nested_Conv_CNN model

After flatten layer, dense layer is added with ReLU activation function and output layer with SoftMax activation function. ReLU is not deeply affected by the vanishing gradient issue, and permits gradients to stay big, ensuring efficient backpropagation and quicker convergence. SoftMax is appropriate for multi-class classification jobs since it transforms a neural network's raw output scores into probabilities that add up to 1. CNN model uses SoftMax activation function in classification layer and sparse categorical cross-entropy (SCE) as the loss function. SoftMax function is using (1).

$$Y_j = \frac{e^{zj}}{\sum_{k=1}^{3} e^{zk}} \quad (1)$$

zj is the output from dense layer, Yj is class j predicted probability, and the denominator is the addition of probabilities of all 3 classes is 1. The objective function or loss function is using (2).

$$L = \sum_{i=1}^{S} \log Y_{ci} \quad (2)$$

S is number of training samples, ci is the index of correct predicted class, and Y_{ci} is the SoftMax calculated probability of correct class. By substituting (1) and (2), the final objective function is obtained as (3).

$$f = \sum_{i=1}^{S} \log\left(\frac{e^{Zj}}{\sum_{k=1}^{3} e^{zk}}\right) \quad (3)$$

It is the final equation of proposed Nested_Conv CNN architecture. This loss function value is minimized during training to recover classification accuracy. Our aim is to minimize the loss as it will boost to classify the correct class. Table 2 will show the used layers with functionalities and Table 3 shows hyper parameters used. In order to identify characteristics like edges, textures, and other spatial patterns, 2D convolution applies a collection of kernels to the input image. The convolution operation's output is subjected to the ReLU function. ReLU adds non-linearity to the model and aids in its learning of intricate patterns by setting all negative values in the output feature map to zero. The layer would effectively function as a linear filter in the absence of this activation. The model can learn more complicated patterns than a single convolution layer could, by using two convolutional layers back-to-back. So, our model will learn automatically the effective features and it designs the structure of it as it gains information from image and classify with the probability distribution.

Table 2. Proposed Nested_Conv_CNN model layers with functionalities

Sr No	Block	Layers	Benefits
1	Block 1–2-convolution functions followed by max pooling	Conv2D (32 filters, 3×3, ReLU) Conv2D (32 filters, 3×3, ReLU) MaxPooling2D (2×2)	Extract low-level handwriting features and refine and reduce feature maps
2	Block 2–1-convolution function followed by max pooling	Conv2D (64 filters, 3×3, ReLU) MaxPooling2D (2×2)	Capture high-level handwriting patterns and refine and reduce feature maps
3	Block 3–1-convolution	Conv2D (64 filters, 3×3, ReLU)	Capture high-level handwriting patterns
4	Flattening	Flatten()	Convert feature maps into a 1D vector
5	Fully connected layers	Dense (64 neurons, ReLU)	Learn high-level handwriting features
6	Output layer	Dense (3 neurons, softmax)	For multi-class classification

Table 3. Hyper parameter used in Nested_Conv CNN

Parameter	Value
Optimizer	Adam
Learning rate	0.001 (default)
Loss function	SCE
Metrics	Accuracy
Number of classes	3
Input size	29×29×1

4.1. Experimental environment

For experimentation purpose, we have used Intel (R), Core (TM) i5 Processor, 8.00 GB RAM, 64-bit operating system, ×64-based processor, no pen or touch input is available; rather input is taken from images of handwritten text on paper. The images are preprocessed to separate out characters on the lines. A region of interest (ROI) is applied to capture portion of an image which contains alphabet that we want to filter and save it as an image in dataset. So, we have represented an ROI as a binary mask image for the experimentation we run the code on Google Colab with Python and TensorFlow (including Keras), NumPy, Matplotlib (for visualization), OpenCV (for image processing), scikit-learn (for preprocessing), and Pandas (for datasets).

5. RESULTS AND DISCUSSION

To extract information and categorize handwritten characters, a simple CNN for handwriting classification consists of convolutional layers followed by dense layers. While the max-pooling layers lower dimensionality and preserve significant features, the earliest convolutional layers capture crucial patterns like edges and curves. Following feature extraction, the features are run through dense layers that are fully connected. This aids in the creation of final predictions using learnt representations. For simple handwriting recognition applications, this architecture works well and is computationally efficient.

5.1. Scenario 1: basic convolution neural network model performance analysis

In this scenario, we have used basic CNN model with 3 convolution, 3 max pooling, 1 dense layer, and 1 output layer. We tried to evaluate the performance of this model on handwriting dataset for its classification. Table 4 presents the performance analysis of basic CNN model for varying number of epochs. Over epochs, the simple CNN model consistently improves its training and validation accuracy, demonstrating successful learning. The model appears to generalize well even in its early stages, as evidenced by the training accuracy commencing at 82.13% and the validation accuracy being marginally higher at 88.27%. Both measures keep rising as training goes on, culminating at 95.9% for validation accuracy and 96.39% for training accuracy. This pattern shows that the algorithm is learning handwriting patterns that are relevant and getting better at making predictions over time. Losses for training and validation both steadily decline; training loss falls from 0.4511 to 0.1022, while validation loss falls from 0.3104 to 0.1175. As the discrepancy between actual and anticipated values decreases, this shows that the model is growing more confident in its predictions. But there is a small oscillation around the fifth epoch, when validation loss marginally rise and validation accuracy falls from 94.02% to 93.67%. This implies that the model may be starting to overfit the training data a little bit or facing a complex variety in handwriting styles. However, the model recovers its performance and achieves high validation accuracy in later epochs.

Table 4. Basic convolution neural network [9] performance analysis

Epochs	Model	Training accuracy	Training loss	Validation accuracy	Validation loss
1	(1 Conv.+1 Max pool+1 Conv.+1 Max pool+1 Conv.+1	0.8213	0.4511	0.8827	0.3104
2	Max Pool+1 flatten layer+1 dense layer+1 output layer)	0.9026	0.2659	0.9181	0.2279
3		0.9232	0.2131	0.9305	0.1906
4		0.9339	0.1829	0.9402	0.1697
5		0.9411	0.1618	0.9367	0.1723
6		0.9483	0.1442	0.9498	0.1405
7		0.9528	0.1315	0.9506	0.1395
8		0.9566	0.1215	0.9517	0.1355
9		0.9605	0.1109	0.959	0.1164
10		0.9639	0.1022	0.9589	0.1175

Internal layers use ReLU function on each convolution layer and SoftMax on last layer for classification. ReLU considered positive or zero and SoftMax used to calculate probability distribution. Figure 5 shows the training and validation accuracy and loss statistics. As per observations, CNN model performs well on training dataset with 96.39% accuracy and 10% loss. It also works well on validation dataset with 95.89% accuracy with 11% loss. We have applied early stop in code and it observe that this is the best accuracy of this model. We tried to test it on real-time images and we got significant change as it used to classify most of samples in corrected class only and as per our observation, it is the sign of overfitting as data is not classified properly.

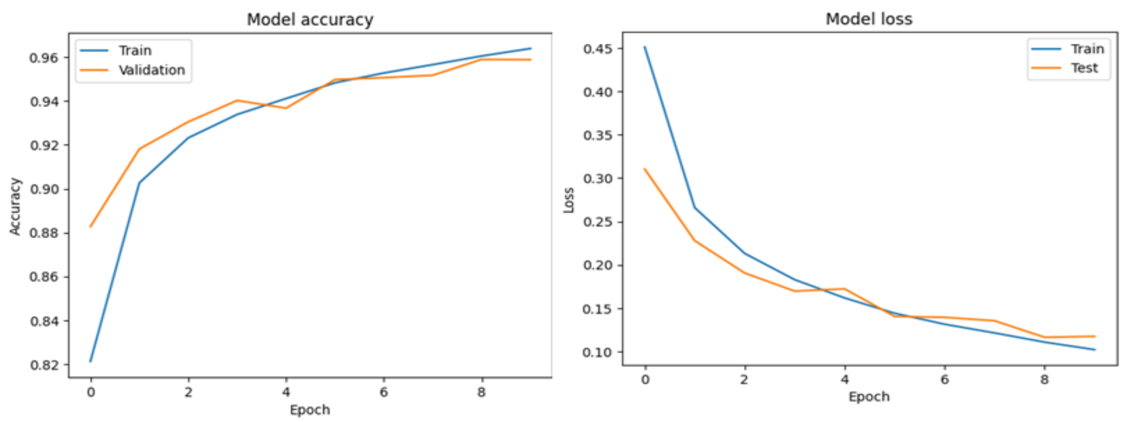

Figure 5. Plots of CNN model training and validation accuracy and loss

Although the basic CNN achieves excellent accuracy, it has drawbacks, such as the possibility of modest overfitting and possible sensitivity to changes in handwriting styles. This can be fixed by proposing a custom CNN with an extra first convolutional layer. The network would be able to capture finer details and increase classification robustness if this additional layer improved feature extraction earlier. The proposed model's architecture can be improved to increase accuracy, reduce overfitting, and efficiently handle a greater variety of handwriting styles.

5.2. Scenario 2: Nested_Conv_CNN model performance analysis

Table 5 shows the model is able to learn a hierarchy of features by sequentially applying convolutional layers, starting with basic edges and textures and moving up to more intricate forms and objects in the deeper layers. After experimentation, we obtained accuracies and losses of training and validation task. The system took 65 seconds in total to finish the epochs. Since minimal loss is a measure of the model's effectiveness, lower values correspond to greater performance. The training loss in this case is 0.0515 as shown in Figure 6.

Table 5. Nested_Conv_CNN model performance analysis

Epochs	Model	Training accuracy	Training loss	Validation accuracy	Validation loss
1	(2 Conv.+1 Max pool+1 Conv.+1 Max pool+1	0.8611	0.3639	0.9177	0.2339
2	Conv.+1 flatten layer+1 Dense layer+1 Output layer)	0.9304	0.1948	0.9391	0.1689
3		0.9467	0.1477	0.9516	0.1366
4		0.9576	0.1188	0.9530	0.1313
5		0.9639	0.1014	0.9614	0.1093
6		0.9686	0.0869	0.9623	0.1080
7		0.9732	0.0755	0.9668	0.0934
8		0.9766	0.0661	0.9687	0.0867
9		0.9797	0.0580	0.9697	0.0838
10		0.9817	0.0515	0.9726	0.0772

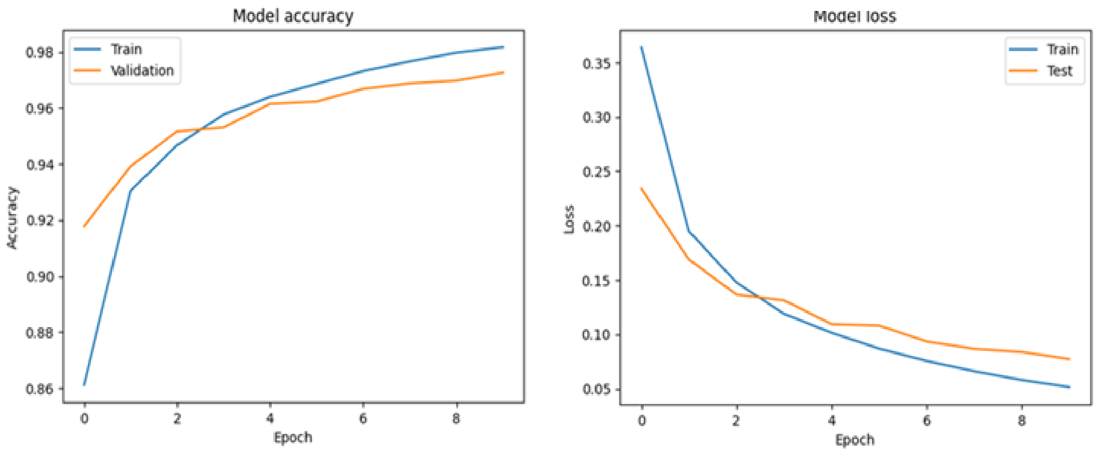

Figure 6. Plots of Nested_Conv_CNN model accuracy and loss

We have set default learning rate with batch size 8 and Adam optimizer. The model's validation loss is computed using an independent validation dataset and unused data samples. It shows the loss is less than CNN [9] and accuracy is improving. This is the sign of robust model. The basic CNN [9] has a low training loss 10% but the high validation loss 11%, as well as a discrepancy between training accuracy 96.39% and validation accuracy 95.89%. Given this disparity, it is probably overfitting the training set. This results in the model performing poorly on the validation set by capturing noise and particular characteristics that do not categorize to new data. Table 6 shows the training and validation accuracy and loss statistics.

Table 6. Comparison of performance metrics

Sr No	Model	T. Acc.	T. Loss	V. Acc.	V. Loss	Trainable parameters	Precision	Recall	Accuracy	F1-score
1	CNN	0.96	0.10	0.95	0.11	60,675	85.30	85.71	85.71	0.85
2	Nested_Conv_CNN	0.98	0.05	0.97	0.07	102,691	88.44	88.35	88.35	0.88

The proposed Nested_Conv_CNN model achieves a balanced performance, with both high training accuracy and high validation accuracy. Its training loss 0.0515 and validation loss 0.0772 are both low and close to each other, indicating that the model has learned the patterns in the data well without overfitting. Thus, the proposed model demonstrates strong generalization, meaning it performs well on new data. Loss-validation is an important technique for producing a robust and trustworthy assessment of a model's performance, ensuring it generalizes well to new data, and assisting with model selection and hyperparameter tuning. Testing accuracy is also 0.95 and loss is 0.10. We performed cross validation and the mean accuracy: 0.96, standard deviation: 0.002 of basic model and novel CNN model is having mean accuracy: 0.97, standard deviation: 0.001. As shown in Table 6, the basic CNN model is too simplistic and unable to generalize adequately to new data, overfitting may happen. The number of trainable parameters in the Nested_Conv_CNN model loss to 102,691 by including an extra convolutional layer. With the addition of additional filters and weights from this new convolutional layer, the model was able to extract more intricate and abstract features from the input data. In order to balance overfitting, the model's complexity is also increased. This enhancement emphasizes how optimizing the model architecture may greatly affect performance by finding the ideal ratio of layers.

5.3. Performance metrics and statistical testing

Figure 7 of confusion matrix shows that the model performs well overall, with high true positive counts for each class: 17,879 for class 0, 16,825 for class 1, and 15,411 for class 2. The confusion matrix shows that the model performs well overall. These numbers indicate that the model has been adequately trained and is capable of correctly predicting the various classes in the dataset. The high true positive counts for each class demonstrate the model's robustness and ability to handle several classes with reasonable accuracy. This performance is a good indicator of the model's overall reliability and efficacy at classifying tasks.

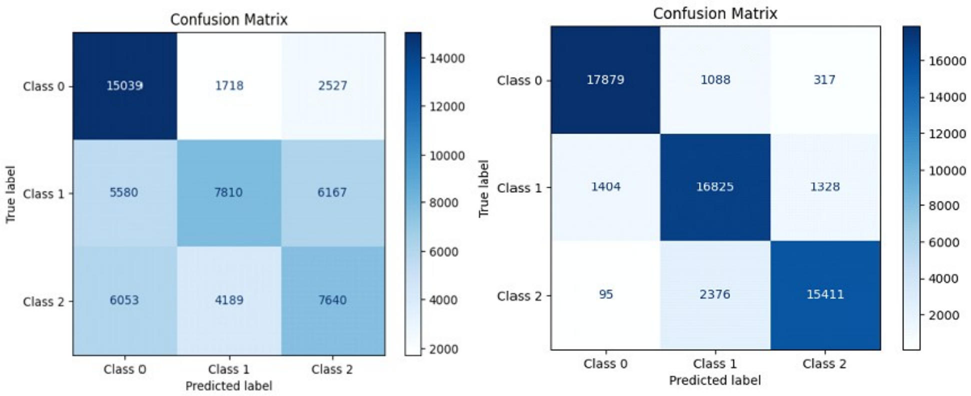

Figure 7. Confusion matrix basic CNN and Nested_Conv_CNN

There is a noticeable performance difference between the basic CNN model and our Nested_Conv_CNN model in terms of classification measures, among other important performance indicators. With an accuracy of 96.39%, existing model is able to classify slightly more than half of the cases correctly. Our model, on the other hand, performs noticeably better, properly classifying the great majority of cases with an accuracy of 88.35%. With a recall of 88.35%, it correctly detects the majority of the real positive cases, and its precision of 88.44% shows that almost all of its positive predictions are accurate. The F1-score of 0.88 indicates that our model is more adept at striking a balance between recall and precision. In general, our model performs far better than existing one, making it a lot more dependable and efficient classifier. The performance of the proposed nested cascaded CNN was evaluated using 5-fold cross-validation and compared with a baseline CNN [9]. Across the folds, the nested CNN achieved a mean validation accuracy of 0.9726 (range: 0.971–0.973) with a corresponding mean validation loss of 0.0772, outperforming the baseline CNN, which recorded a mean validation accuracy of 0.9589 (range: 0.956–0.960) and mean validation loss of 0.1175. It despite having only 102,691 trainable parameters, demonstrated superior performance while maintaining a lightweight design compared to the baseline's 60,675 parameters. Statistical analysis using a paired t-test across the 5 folds confirmed that the improvement in accuracy is statistically significant ($p < 0.05$). It also proven with use of ROC curves and AUC values, which show how effectively the model can distinguish between different handwriting classes. Encouragingly, the model maintained high AUC scores even under challenging conditions, suggesting it can reliably recognize handwriting patterns despite small distortions.

6. ABLATION STUDY

When we tried simplifying our model to just a single convolutional layer, we noticed a clear drop in performance. Accuracy fell by nearly 5%, and the loss increased by 3–5%, showing that the model struggled to truly understand the handwriting patterns. It was as if the network could only see the rough outlines but missed the finer details—the curves, loops, and strokes that make each letter unique. Multiple convolutional layers act like layers of understanding: the first sees edges, the next starts to recognize shapes, and deeper layers capture the full structure of letters. Without this hierarchy, the model couldn't generalize well, and learning became unstable. This experience reinforced why stacking convolutional blocks is crucial, they allow the network to gradually gives detailed understanding and helping it achieve higher accuracy and lower loss.

7. CONCLUSION

This paper presented improved novel CNN model for dysgraphia classification based on English handwriting. We have experimented with two models like basic CNN and proposed Nested_Conv_CNN and proposed model worked well as the training accuracy is 98.17%. The model's validation loss is computed using an independent validation dataset and unused data samples. It measures the model's ability to generalize to new untested data with validation loss is 0.0772 and validation accuracy is 97.26%. As CNN is mainly used to solve the issues of feature learning in non-linearity-based samples and effective down sampling, the network performs better on classification tasks. These features boost the network's power and capacity to identify intricate patterns. We have tried architectural changes and it worked well. We have used default learning rate and batch size of 8 with Adam optimizer. While the computational complexity slightly increases, the benefits of improved recognition performance make the Nested_Conv_CNN more appropriate for complex handwriting datasets. This helps extract lower-level features like edges and textures before moving to deeper feature representations.

FUNDING INFORMATION

The authors declared that they have not received any funding for this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Nisha Vanjari	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Prasanna J. Shete	✓	✓				✓		✓	✓	✓	✓	✓		

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

The authors declared that they have no known competing interest.

DATA AVAILABILITY

The datasets analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

[1] E. S. Shapiro, "Preventing academic failure," *School Psychology Review*, vol. 17, no. 4, pp. 601–613, 1988, doi: 10.1080/02796015.1988.12085378.

[2] W. W. Lim, K. J. Yeo, and L. Handayani, "A systematic review on interventions for children with dyslexia," *International Journal of Evaluation and Research in Education*, vol. 12, no. 3, pp. 1674–1682, 2023, doi: 10.11591/ijere.v12i3.25099.

[3] P. Westwood and L. Graham, "How many children with special needs in regular classes? official predictions VS teachers' perceptions in South Australia and New South Wales," *Australian Journal of Learning Disabilities*, vol. 5, no. 3, pp. 24–35, 2000, doi: 10.1080/19404150009546632.

[4] K. Spoon, D. Crandall, and K. Siek, "Towards detecting dyslexia in children's handwriting using neural networks," in *Proceedings of International Conference on Machine Learning AI for Social Good Workshop*, 2019, pp. 1–5.
[5] M. S. A. B. Rosli, I. S. Isa, S. A. Ramlan, S. N. Sulaiman, and M. I. F. Maruzuki, "Development of CNN transfer learning for dyslexia handwriting recognition," in *2021 11th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2021*, 2021, pp. 194–199, doi: 10.1109/ICCSCE52189.2021.9530971.
[6] N. S. L. Seman, I. S. Isa, S. A. Ramlan, W. L. -Chih, and M. I. F. Maruzuki, "Notice of removal: classification of handwriting impairment using CNN for potential dyslexia symptom," in *2021 11th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2021*, 2021, pp. 188–193, doi: 10.1109/ICCSCE52189.2021.9530989.
[7] I. S. Isa, M. A. Zahir, S. A. Ramlan, W. L. -Chih, and S. N. Sulaiman, "CNN comparisons models on dyslexia handwriting classification," *ESTEEM Academic Journal*, vol. 17, pp. 12–25, 2021.
[8] I. S. Isa, W. N. S. 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.
[9] P. J. Grother and K. K. Hanaoka, "NIST special database 19," *National Institute of Standards and Technology*. Accessed: Apr. 20, 2024. [Online]. Available: https://www.nist.gov/srd/nist-special-database-19
[10] J. Kawa, A. Bednorz, P. Stępień, J. Derejczyk, and M. Bugdol, "Spatial and dynamical handwriting analysis in mild cognitive impairment," *Computers in Biology and Medicine*, vol. 82, pp. 21–28, Mar. 2017, doi: 10.1016/j.compbiomed.2017.01.004.
[11] K. Spoon, K. Siek, D. Crandall, and M. Fillmore, "Can we (and should we) use AI to detect dyslexia in children's handwriting?," in *Proceedings of Artificial Intelligence for Social Good (NeurIPS 2019)*, 2019, pp. 1–6.
[12] S. Patel, "A-Z handwritten alphabets in .csv format," *Kaggle*. Accessed: Apr. 20, 2024. [Online]. Available: https://www.kaggle.com/sachinpatel21/az-handwritten-alphabets-in-csv-format
[13] I. S. Hussein and N. A. Jasim, "Iris recognition scheme based on entropy and convolutional neural network," *Advances in Technology Innovation*, vol. 9, no. 3, pp. 197–209, 2024, doi: 10.46604/aiti.2024.13516.
[14] S. Rangasrinivasan, M. S. S. Suresh, A. Olszewski, S. Setlur, B. Jayaraman, and V. Govindaraju, "AI-enhanced child handwriting analysis: a framework for the early screening of dyslexia and dysgraphia," *SN Computer Science*, vol. 6, no. 5, 2025, doi: 10.1007/s42979-025-03927-0.
[15] B. Manimekala, D. Umamaheswari, J. Rozario, M. Kannan, and P. M. Savitha, "Dysgraphia disorder detection and classification using deep learning technique," *SN Computer Science*, vol. 6, no. 3, Mar. 2025, doi: 10.1007/s42979-025-03825-5.
[16] A. Fallah, Y. Z. Vakili, and H. Sajedi, "AI-driven approaches for dysgraphia diagnosis using online and offline handwriting data: a comprehensive scoping review," *PLOS One*, vol. 20, no. 12, Dec. 2025, doi: 10.1371/journal.pone.0328722.
[17] L. C. Ribeiro, L. C. Afonso, and J. P. Papa, "Bag of samplings for computer-assisted Parkinson's disease diagnosis based on recurrent neural networks," *Computers in Biology and Medicine*, vol. 115, 2019, doi: 10.1016/j.compbiomed.2019.103477.
[18] M. F. -Zanuy, J. Fierrez, M. A. Ferrer, M. Diaz, R. Tolosana, and R. Plamondon, "Handwriting biometrics: applications and future trends in e-security and e-health," *Cognitive Computation*, vol. 12, no. 5, pp. 940–953, 2020, doi: 10.1007/s12559-020-09755-z.
[19] J. Chai *et al.*, "Classification of mild cognitive impairment based on handwriting dynamics and qEEG," *Computers in Biology and Medicine*, vol. 152, 2023, doi: 10.1016/j.compbiomed.2022.106418.
[20] A. Devi and G. Kavya, "Dysgraphia disorder forecasting and classification technique using intelligent deep learning approaches," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 120, 2023, doi: 10.1016/j.pnpbp.2022.110647.
[21] J. Kunhoth, S. A. Maadeed, M. Saleh, and Y. Akbari, "CNN feature and classifier fusion on novel transformed image dataset for dysgraphia diagnosis in children," *Expert Systems with Applications*, vol. 231, 2023, doi: 10.1016/j.eswa.2023.120740.
[22] S. M. S. Islam, E. K. Dey, M. N. A. Tawhid, and B. M. M. Hossain, "A CNN based approach for garments texture design classification," *Advances in Technology Innovation*, vol. 2, no. 4, pp. 119–125, 2017.
[23] R. Sharma and A. Singh, "An integrated approach towards efficient image classification using deep CNN with transfer learning and PCA," *Advances in Technology Innovation*, vol. 7, no. 2, pp. 105–117, 2022, doi: 10.46604/aiti.2022.8538.
[24] P. Sharma, B. Agarwal, G. S. Yadav, and S. Jain, "Vision transformer-based model for early detection of dysgraphia among school students," *Microsystem Technologies*, vol. 31, no. 3, pp. 775–785, 2025, doi: 10.1007/s00542-024-05741-9.
[25] B. Agarwal, S. Jain, K. Beladiya, Y. Gupta, A. S. Yadav, and N. J. Ahuja, "Early and automated diagnosis of dysgraphia using machine learning approach," *SN Computer Science*, vol. 4, no. 5, 2023, doi: 10.1007/s42979-023-01884-0.

BIOGRAPHIES OF AUTHORS

Nisha Vanjari is assistant professor in Department of Computer Engineering, KJSIT, Mumbai, since January 2009. She is pursuing her Ph.D. in Computer Engineering from KJSSE. She is M.E. in Information Technology, P. G. D. B. M. from NMIMS Deemed University, B.E. in Computer Engineering from North Maharashtra University, and Diploma in Computer Technology from BTE. She has done 'software quality testing' certification–CST from SQTL. She has published more than 20 papers in conferences, book chapters, journals. For the last five years, her main research was in the field of networking, data science and artificial intelligence. This has led her interest in the deep learning and cloud computing as well. She has published 2 books for Distance and Open Learning Course MCA. She can be contacted at email: nvanjari@somaiya.edu.

Prasanna J. Shete is associate professor in Computer Engineering, KJSSE. He is Ph.D. (Science and Technology) V. J. T. I., University of Mumbai, M. Tech. (Electronics and Telecommunication), V. J. T. I., University of Mumbai and B.E. (Electronics Engineering), Dr. B. A. Marathwada University, Aurangabad. He is working in K. J. Somaiya School of Engineering since June 2003. He is the convener of the Equal Opportunities Cell, Somaiya Vidyavihar University, coordinator, KJSSE resource center for pwd students, Former in-charge, KJSSE data center. He has published more than 25 papers in conferences, book chapter, and journals. His research specialization in mobile computing, IoT, emerging wireless networks, WSNs, MANETs, VANETs, accessibility, and assistive technologies. He can be contacted at email: prasannashete@somaiya.edu.