A Cognitive-Driven Stacking Ensemble Approach for Dyslexia Handwriting Classification using LeNet-Based Deep Learning Models

Tabassum Gull Jan[1]*, Sajad Mohammad Khan[1], Sajid Yousuf Bhat[1]

[1] *Department of Computer Sciences,*
   University of Kashmir, J&K, Srinagar, 190006, INDIA

*Corresponding Author: tabassumgull.scholar@kashmiruniversity.net

Abstract

Dyslexia, a specific learning disability, affects cognitive processing in approximately 5-20% of children worldwide. In India, its prevalence is alarmingly high, affecting about 10% of the population, yet early screening remains limited despite growing technological advancements. Dyslexic individuals often struggle with interpreting words and visual stimuli, but timely intervention can significantly improve learning outcomes. Recently, handwriting analysis has gained attention for dyslexia detection because it uses cognitive-motor patterns linked with dyslexia. Although several studies have explored handwriting-based dyslexia detection, there is still need for more accurate and efficient classification methods for dyslexic handwriting using advanced machine learning techniques. Our proposed design is cognitive-driven wherein each architectural choice targets well-known grapho-motor manifestations linked to dyslexia such as mirror reversals (b/d, 6/9, p/q), stroke formation, spacing and size variability. In this study, we introduce a novel approach utilizing a stacking ensemble deep neural network for classifying dyslexic handwriting. Our method employs four variants of the LeNet-5-based Convolutional Neural Network models, distinguished by different numbers of triple convolution layers used to extract cognitive-motor features from handwriting patterns. The triple convolution layers capture multi-scale stroke information like edges, curvatures and junctions. All models were trained and validated on a dataset using standard performance metrics. Experimental results revealed increasing depth enhances overall model performance. By stacking the four variants of the models, the proposed approach attained an impressive accuracy of 96.86% in classifying the three classes of dyslexic handwriting. Notably, the ROC curves demonstrate perfect classification for the Corrected and Reversed classes with an AUC of 1.00, and an AUC of 0.96 for the Normal class, indicating the generalizability and robustness of the proposed approach. These findings highlight the potential of deep learning and cognitive-driven handwriting analysis in advancing dyslexia screening, emphasizing the effectiveness of stacking ensemble in addressing this critical issue.

1. Introduction

Among specific learning disabilities, dyslexia is the most prevalent accounting for 80% of the cases [1]. Dyslexia has been placed in the top five most frequently occurring learning disabilities as per the reports of National Center for Learning Disabilities (NCLD) [2]. The word "Dyslexia" has been derived from Greek word "Dys" meaning "lack of" and "Lexicon" meaning "word or verbal language"

Dyslexia is a hidden learning disability said to be neurobiological in origin, wherein students usually experience problems in reading writing, spelling, decoding words and comprehension [3]. Such individuals find it difficult to distinguish between "b" and "d" likewise "6" and "9" and so on. It has been reported that 1 in every 10 individuals suffers from dyslexia. In America 40 million adults are dyslexic and only 2 million know it1.

There are several other categories of dyslexia such as phonological dyslexia, rapid automatized naming (RAN), surface dyslexia, directional dyslexia, double deficit dyslexia and visual deficit dyslexia [4]. Anyone can have dyslexia irrespective of age and gender, the only thing that varies is the process of testing. It may be different for adults and different for children. Research studies reveal that an individual with dyslexia is more creative and intelligent, yet they suffer from basic reading skills [4]. Dyslexia has nothing to do with ones IQ. In fact, such students have been found out to be more intelligent and smarter than normal intelligence ones. Not a single test can diagnose dyslexia alone. Series of tests have been designed by practitioners to diagnose dyslexia more accurately during early stages of development. Furthermore, different approaches have been used by researchers to detect dyslexia viz machine learning (ML) based approaches, Image processing-based approaches, EEG based approaches, MRI based approaches, Eye movement-based approaches, Handwritten Image based approaches.

Since dyslexia is the most prominent specific learning disability, so detection and evaluation of it during early stages of development is the most crucial problem that needs to be addressed. Although it is challenging to detect it by a single test, yet researchers have developed series of dyslexia detection systems incorporating inputs from advanced techniques like ML, image processing, artificial intelligence, deep learning (DL), EEG-based, Eye movement based and so on.

In the field of machine learning and predictive modelling, stacking ensemble techniques have garnered considerable attention due to their capability to significantly enhance predictive accuracy and model robustness. Stacking ensembles have been used by researchers for variety of applications like forecasting wind power [5], hepatocellular carcinoma detection [6], air pollution prediction [7], electricity consumption bill prediction [5] and so on. Stacking involves amalgamating multiple base learners, which could be distinct machine learning algorithms or variations of the same algorithm, with the goal of achieving a more precise final prediction. The predictions from each base learner are supplied as an input to the meta-learner, which subsequently combines these predictions to yield the final predicted output. This approach is based on combining the strengths of the individual base learners, compensating their weakness and surpassing the performance of single models.

The present study proposes a novel stacking ensemble learning approach for dyslexia detection using handwritten images. The proposed method is based on an ensemble of four convolutional neural network (CNN) models that differ in architectural configurations. The proposed approach is driven by cognitive principles, aligning well with the effective feature extraction focusing mainly on cognitive-motor correlates of dyslexia related to handwriting: such as left-right confusion, orientation styles that yields mirror reveals (b/d, 6/9, p/q), variable spacing and letter size inconsistency. The model is specifically designed to prioritize orientation-sensitive, curvature aware representations that map onto these behaviours more closely associated with dyslexia handwriting.

The experimental observations demonstrate that the proposed stacking ensemble model achieved higher accuracy than all the four CNN models used, which verifies the effectiveness of approach in classification process.

The noteworthy contribution of the paper includes:

- Code optimization is adopted to improve the accuracy and stability of the proposed approach.
- An improved stacking ensemble is proposed for dyslexia handwriting classification.
- Stacked CNNs are combined in an ensemble using a logistic regression meta-learner to achieve better generalization.
- The proposed model enhanced classification accuracy and reduced error rate.
- The proposed model outperforms existing state-of-art-approaches achieving higher accuracy and faster prediction time.

2. Related Works

Recent technological advancements have led researchers to explore various machine learning based methods for early detection of dyslexia. These methods often focus on assessing brain behaviour or handwriting styles. Notably, brain behaviour analysis is prominent since dyslexia relates to differences in brain function. Identifying

---

brain behaviour is the most promising research field in the direction of dyslexia detection as the problem is pertaining to difference in functioning of brain parts itself. Past studies have utilized diverse data types for dyslexia diagnosis. like EEG data [8][9], eye tracking measures and eye fixation points [10][11][12][13], EOG data, audio signals [14], behavioral tests, MRI images [15], Handwritten images [16][17][18] etc.

Various approaches penned down in literature concerning dyslexia detection using machine learning employed variety of machine learning algorithms like KNN, SVM, Particle Swarm Optimization, Naive Bayes, Decision Trees, and Logistic Regression for classifying dyslexic and normal readers [8][9][10][14][15][19][21][22]. However, these methods typically focus on single aspects of dyslexia, limiting their broader applicability.

Recent studies on handwriting analysis, used optical color recognition (OCR) for transcripting handwritten images of dyslexic students. Employing CNN in such scenarios attained an accuracy of 77% [16] [17]. Some researchers worked on dysgraphia detection and analyzed the impact of adding layers to CNN and attained a training accuracy of 97.2% and testing accuracy 87.44% [18]. While promising, these approaches struggle with diverse handwriting styles and require significant preprocessing. Similarly, using deep learning models like LeNet for handwritten data yielded 88.73% accuracy, but these models are computationally intensive and may not be feasible for real-time applications [20][21]. Specifically, Bin Rosli et al. [20] proposed a transfer learning-based adaption of LeNet-5 model. The original LeNet-5 convolutional layers were retained, while the fully connected layers were fine-tuned using pre-trained parameters from a similar dataset. In contrast, Isa et al. [21] implemented a comparative study involving varying CNN variants, including a modified LeNet-5 architecture to improve performance. The modification primarily included a Convolutional layer, ReLU activation functions with tanh, and a dropout layer before the dense layers to overcome overfitting. Since both [20] [21] were based on LeNet-5 base architecture, they differ in training strategy (transfer learning vs architectural enhancement) and the model configuration, justifying the reported performances.

Other studies related to this comprised of neuroanatomical investigations that revealed individuals with developmental dyslexia exhibit distinct brain connectivity patterns during reading and writing tasks [22][23][24]. Nowadys, EEG analysis has been used by many researchers for investigating the functional networks involved in the brain of dyslexic and normal students while performing a particular task. Some researchers analyzed the resting state EEG of dyslexic students using graph analysis [25]-[29] Techniques like Independent Component Analysis (ICA), Principal Component Analysis (PCA), Sequential Floating Forward Selection (SFFS) algorithm were employed to extract an optimal set of spectral and phase related features. These were then subjected to classification using SVM and an accuracy of 92% was obtained [26]. These methods being effective often require professional expertise restricting their usage in normal real-world applications.

Ensemble learning coupled with multimodal feature extraction has emerged as new research direction in deep learning approaches, enabling the models to fuse multiple feature representations from varied sources for improving robustness and accuracy. The DxDetekt system developed by researchers [31] consists of CNN and CNN-LSTM architectures, extracting features from handwritten sequences and attained an accuracy of 90%, effectively capturing both spatio-temporal dynamics involved in dyslexia handwritten styles. Along the same lines, a multimodal ensemble [32] integrating handwriting data both online and offline performed conditional feature fusion and attained an accuracy of 88.8%, outperforming all single modality approaches at the cost of system complexity. In neurocognitive studies related to dyslexia detection, ensemble approaches have been used to ERP-EEG signals offering explainable insights to brain functional analysis [33], while some studies using autoencoder based approaches [34] achieved an accuracy of more than 80%, revealing altered temporal and frontal connectivity. Beyond EEG and handwritten images, ensemble approaches like bagging, boosting and voting delivered an accuracy of 90% [35]. Studies on behavioral screening and quiz-based assessments using XGBoost ensemble also demonstrated robust performance [36]. Jyakanth et al. [37] developed a CNN feature fusion-based method for transformed handwritten image dataset. With extensive resource utilization demands and complexity in developed systems, these approaches have the potential to reduce the overfitting risk, improve generalization and robustness of models built in real-world settings for varied handwriting styles, collectively opening the road for ensemble-based dyslexia detection approaches for handwriting classification.

The proposed model is designed in such a way to overcome the limitations of existing approaches and capture the complex and varying handwriting patterns using triple convolution layers. These layers are introduced in all the four variants of CNN (base learners), effectively capturing the varying dyslexic behaviour in handwriting with improved accuracy than state-of-art approaches.

3. Proposed Method

The proposed methodology for stacking ensemble based deep learning model for dyslexia handwriting classification is detailed in Fig. 1. The raw data is subjected to dataset preprocessing followed by Ensemble stacking and final prediction. The core methodology is based on the design of four different LeNet-5 architectural variants named as Model 1, Model 2, Model 3 and Model 4 as detailed in architectural block diagrams Fig. 5 to Fig.

8. The output predictions yielded from individual models are then stacked via logistic regression based meta-learner to generate the final predicted outcome.
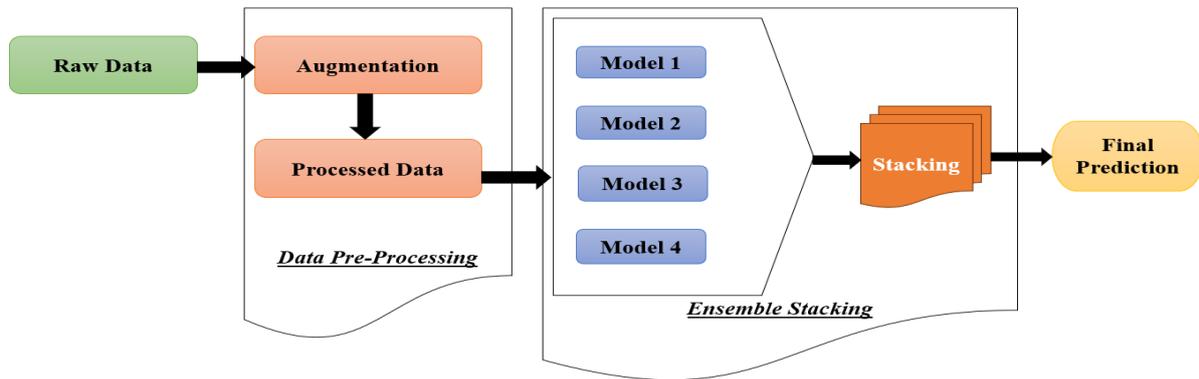

Fig. 1 *Proposed system*

Unlike existing approaches [20][21] that typically rely on modified LeNet architectures, proposed method introduces several enhancements, hyperparameter tuning, ensemble learning to improve model performance.

3.1 Architectural Enhancements

The architectural enhancements carried out in the traditional LeNet-5 architecture include altering number of layers and introduction of triple convolution layers. These architectural modifications are aimed at increasing the classification accuracy. Table 2 shows the architectural details of the base LeNet-5 architecture and Table 3 shows details of modified LeNet-5 architecture adopted by existing techniques [20][21]. The architectural enhancements for all four variants of LeNet-5 used in the study are tabulated in Table 4 to Table 7 in Section 5.

3.2 Diverse CNN Architectures

In this paper to ensure diversity in CNN architecture, four different CNN models have been designed with varying number of triple convolution layers as tabulated with layer wise specifications in Table 4 to Table 7. These diverse CNN architectures enable the proposed model to capture diverse handwriting patterns exhibited by dyslexic children.

3.3 Hyperparameter Optimization

Extensive hyperparameter tuning was conducted for all models to maximize classification accuracy. This process involved hyperparameter optimization by finding the optimal values for parameters such as learning rate (LR), batch size (BS), number of epochs (Max Epochs), and activation functions. Our goal was to achieve optimal performance across all three handwriting classes. The critical hyperparameters of all the CNN models were selected through extensive experimentation and are detailed in Table 1 below

Table 1 *Hyperparameters of base variants of CNNs and the logistic regression meta-learner in the stacking ensemble*

Hyperparameter	Optimal Value	Description
Learning Rate	0.001	Selected after tuning within the range [0.0001 to 0.1] for optimal accuracy
Batch Size	64	[8,32,64,128,256,512] empirically chosen to balance performance
Optimizer	Adam	Used for its adaptive learning rate and faster convergence
Max Epochs	50	With early stopping applied based on validation loss
Loss Function	Flattened Loss of Cross Entropy Loss	Multi-classification
Activation Function	ReLU	Introduces non-linearity
Meta Learner	Logistics Regression	Learns optimal weighting of predictions
Regularization	L2	Smoothens boundaries and reduces overfitting
Training Strategy (Meta Learner)	5-Fold	Improves accuracy and generalization of stacking models

3.4 Ensembling Technique

In this paper, an ensemble-based technique using stacking is employed to aggregate predictions of the four different CNN models viz Model 1, Model 2, Model 3 and Model 4 as shown in Fig. 5, Fig. 6, Fig. 7 and Fig 8, respectively. The proposed ensemble classification approach shows a significant improvement in predictive accuracy. By integrating multiple CNN architectures and employing a robust ensembling strategy, the proposed method offers a novel and effective solution to this complex problem.

To evaluate our method, a detailed comparative analysis with existing approaches have been carried out. The performance metrics include confusion matrix, accuracy, sensitivity, specificity, error rate, precision, recall, F1-score, and ROC curve analysis.

4. Dataset

In this study, we have carried out an image-based approach for dyslexia classification primarily applying advanced ML techniques to enhance the overall accuracy and prediction of existing models. The dataset used comprises a total of 78,275 images for the normal class, 52,196 images for the reversed class (Dyslexic Class), and 8,029 images for the corrected class. To address class imbalance issue, we have computed class weights from the train split using the effective number of samples method by Cui et. al [38]. These weights were applied to the cross-entropy loss to counteract majority class bias. This approach has effectively mitigated the class imbalance problem as is reflected by the improved performance of minority class (Corrected class accuracy= 96.86% see Table 10). The dataset² used in the study has been sourced from works [18][20] [21] and is freely accessible at Kaggle². The dataset is already partitioned into train and test folder sets by original dataset providers, and hence no manual splitting was performed. As the dataset includes sufficiently large and well-structured images in the testing set, so we didn't implement the cross validation. A snapshot of the images in the dataset is presented in Fig. 2
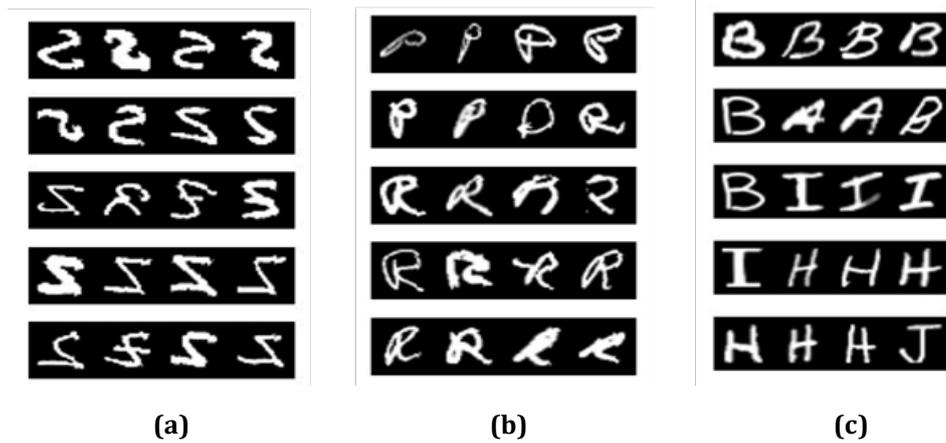
---

² https://www.kaggle.com/datasets/drizasazanitaisa/dyslexia-handwriting-dataset

(a)   **(b)**   **(c)**

Fig. 2 *Sample dataset images (a) reversed class; (b) corrected class; (c) normal class*

To evaluate the proposed model, the train part of dataset is used for training the classifiers and a test set for validation purposes. Series of experiments have been conducted to assess the impact of the four different CNN models with varying architectures. Hyperparameter tuning as shown in Table 1 was performed to optimize the models, achieving high accuracy and low error rates. Additionally, dataset preprocessing and augmentation were applied to the raw data to enhance the classification accuracy of the ensemble technique.

4.1 Data Preprocessing and Augmentation

Pre-processing is a crucial step in preparing any image dataset, transforming it into a standardized format suitable for model training. Due to varying sizes and shapes of images with uneven resolutions in the original dataset, it is necessarily important to preprocess the images first before supplying them as an input to the deep learning-based models. The whole preprocessing and augmentation pipeline is shown in Fig. 3.
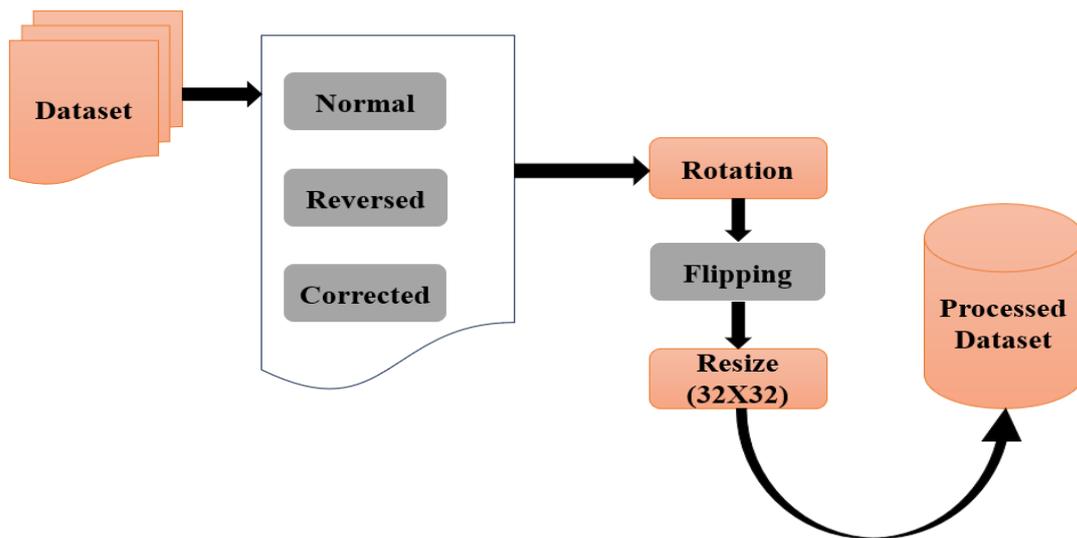

Fig. 3 *Dataset preprocessing and augmentation*

Firstly, the raw images undergo rotation. The rotation can be clockwise or anticlockwise with the rotation angle varying between 1° and 90°. This preprocessing step introduces variability, that makes the model generalize better. Now, the dataset diversity is enhanced by the technique called flipping. After this step images are resized to consistent dimensions (32x32 pixels), so that all images are compatible with all for CNN models for efficient training. Lastly, the images in the dataset are labelled based on the folder names. The "Normal" label corresponds to the Normal class, "Reversed" to the Reversed class, and "Corrected" to the Corrected class. This labelling process is implemented using FastAI data loaders, streamlining the whole workflow of the proposed model. This pre-

processing and data augmentation pipeline standardize the dataset, providing a robust foundation for all four CNN models to achieve high classification accuracy.

5. Architectural Details

LeNet-5 is a CNN architecture that was proposed by "Yann LeCun et al." in 1998, that is specifically designed for handwritten digit recognition (HDR). LeNet-5 was instrumental in demonstrating the effectiveness of CNNs for image classification tasks and laid the foundation for modern deep learning. The generalized LeNet-5 architecture for 3 class problems used as basic architecture is shown in Fig. 4 below.

The LeNet-5 architecture consists of seven layers, including three different layer types: (a)convolutional layers, (b)sub sampling layers, and (c) fully connected (FC) layers.

I.   **Input Layer:** It accepts grayscale images of fixed dimensions, typically 32x32 pixels.

II. **Convolutional Layers:** The first convolutional layer applies six different filters (or kernels) to the input image, producing six feature maps as outputs. Each filter is small in spatial size (5x5) and convolves across the input using a dot product operation. The output feature maps capture different local image features through convolutional operations. To introduce nonlinearity, sigmoid activation function is applied elementwise.

III. **Subsampling Layers:** The subsampling (or pooling) layers are meant to reduce the spatial dimensions. These layers follow each convolutional layer in order to make the network more robust to variations in position and scale. The most common subsampling operation is average pooling, where each feature map is divided into non-overlapping regions (typically 2x2) and the average value within each region is retained. Subsampling reduces the number of parameters and computational complexity while preserving important features.

IV. **Convolutional Layers (Repeated):** Two additional convolutional layers with 16 and 120 filters, respectively, are applied. Similar to the first convolutional layer, subsampling layer is used to further reduce the spatial dimensions.

V.   **Fully Connected Layers:** After the final subsampling layer, the network transitions to fully connected layers, which are similar to those in a traditional neural network. The outputs from the last convolutional layer are flattened into a 1D vector. Two fully connected layers with 84 and 3 neurons, respectively, are employed. The final FC layer yields a 3-dimensional output vector, corresponding to the 3 classes (Normal, Reversed and Corrected).

VI. **Output Layer:** This is the last layer of the architecture outputting prediction probabilities of the classes. It comprises of SoftMax activation function that is needed to convert the outputs from the previous layer into probabilities of each class. The highest-class probability is selected as the predicted class label for the input image.
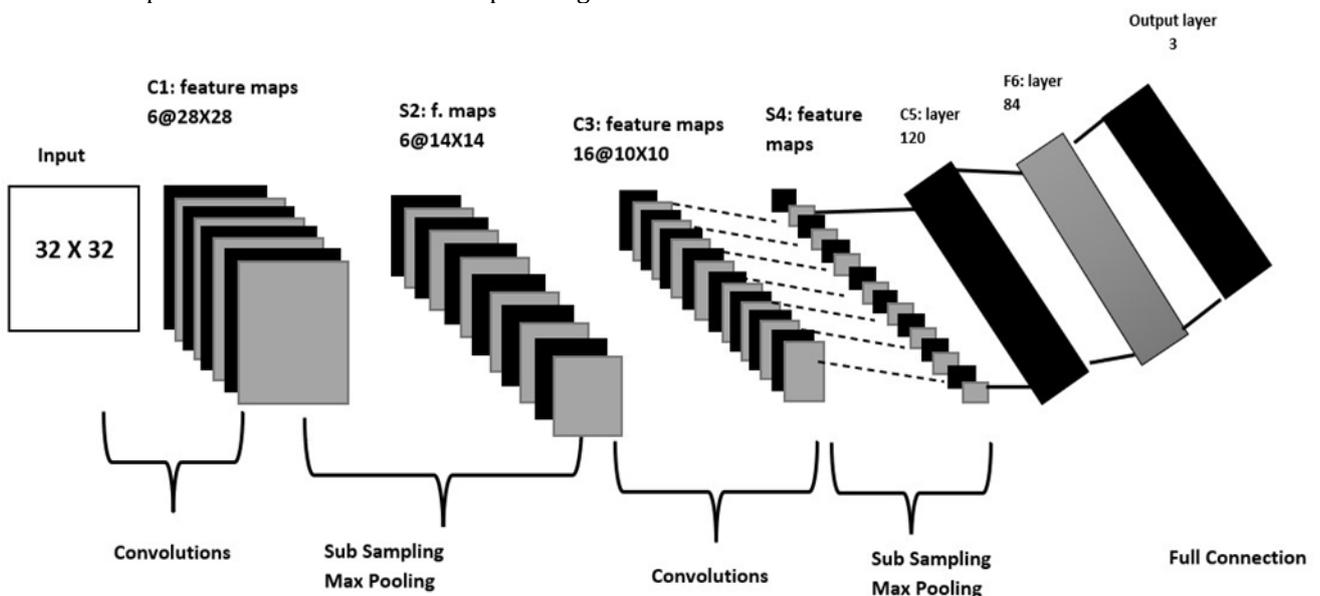

Fig. 4 *LeNet-5 architecture for 3 class problem*

The experimental setup involved executing code files using a Jupyter notebook environment on an NVIDIA server (NVIDIA Tesla V100 GPU with 32 GB of memory, an Intel Xeon processor operating at 2.30 GHz, and 128

GB of RAM. The system ran on Ubuntu 20.04 LTS) hosted at Department of Computer Science, University of Kashmir, Srinagar. The model development and training were carried out using Python 3.9 with Pytorch 1.10.1 as primary deep learning framework. FastAI 2.7 was utilized for data loading, augmentation and training workflows, while OpenCV 4.5 for image processing related tasks. We developed and evaluated four variations of the LeNet-5 architecture, carefully designing their layers and tuning hyperparameters to maximize performance in dyslexia handwriting recognition and class prediction.

Tables 2 and Table 3 provide a detailed layer-by-layer summary (including layer wise parameters such as the number of neurons, stride size, kernel size, and feature maps) of both the original and modified LeNet-5 architectures. Likewise, the architecture details of the proposed four LeNet-5 versions are presented in Tables 4-Table 7. Model 1, described in Table 4, builds upon the existing LeNet-5 architecture by adding an extra convolutional layer. Model 2, detailed in Table 5, further enhances Model 1 by incorporating one triple convolution layer, two additional convolutional layers, and two batch normalization layers. Model 3, shown in Table 6, includes two triple convolutional layers along with one standard convolutional layer. Finally, Model 4, as outlined in Table 7, features three triple convolutional layers and two standard convolutional layers, followed by batch normalization layers. Fig. 5, Fig. 6, Fig. 7 and Fig. 8 show the architectural block diagram of the four proposed LeNet-5 variants. Fig. 9 illustrates the architecture of the triple convolution layers. The predictions from all four models are combined using stacking ensemble as shown in Fig.1.

The proposed stacking ensemble model is cognitive driven in terms of its structural design inspired by the hierarchal nature of human cognitive processing during handwriting tasks. Dyslexic handwriting exhibits irregularities in motor coordination, spatial alignment and stroke sequencing, resulting in manifestations of underlying cognitive motor challenges. To capture these characteristics, the proposed model employing the triple convolution blocks is designed to learn progressively complex spatial motor representations. The first convolution layer is focusing on capturing low-level stroke orientations analogous to visual perception, second layer extracts feature related to motor coordination patterns and last layer extracts high level abstractions of handwriting rhythm. This architectural design matches the human cognitive process of transforming inputs into coordinated motor outputs, making the proposed triple convolution strategy cognitive driven. Through this biologically inspired design, the model effectively learns discriminative features linked to dyslexia.
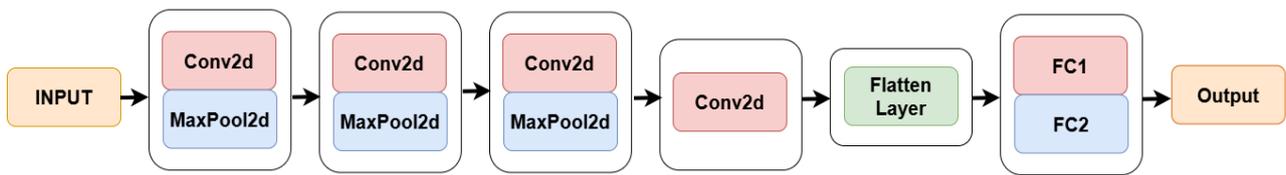

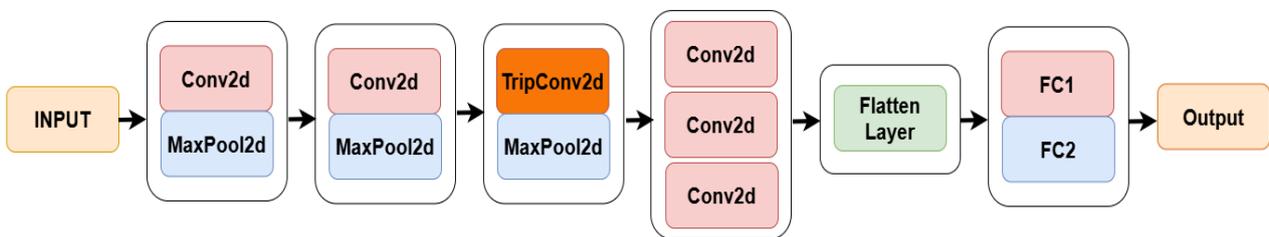**Fig. 5** *Architectural diagrams of Proposed Model 1*
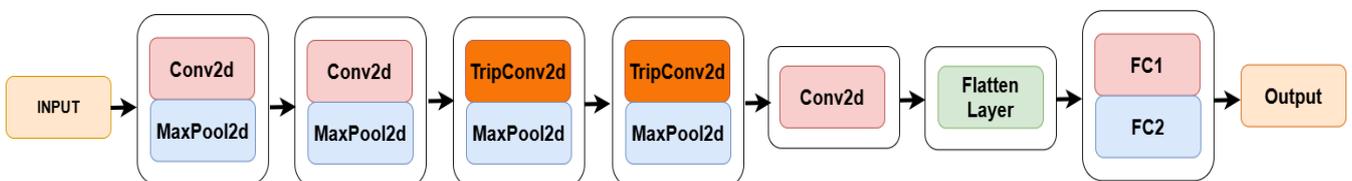

Fig. 6 *Architectural diagrams of Proposed Model 2*

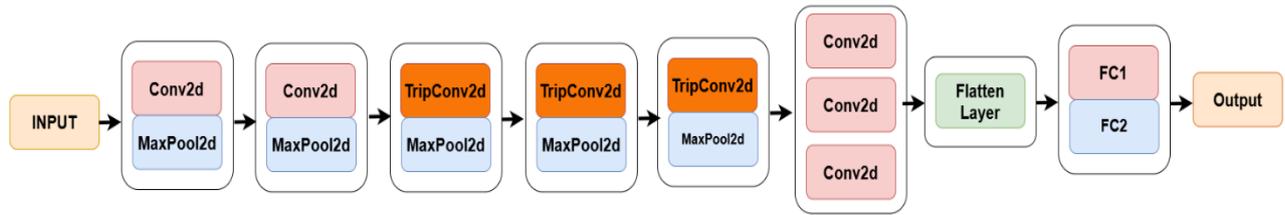**Fig. 7** *Architectural diagram of Proposed Model 3*

Fig. 8 *Architectural diagrams of Proposed Model 4*

Table 2 *LeNet-5 architecture details*

Layer	Filters /Neurons	Filter Size	Stride	Size of Feature Map
Input Image	-	-	-	32x32x3
Conv2d	6	5x5	1	28x28x6
MaxPool2d		2x2	2	14x14x6
Conv2d	16	5x5	1	10x10x16
MaxPool2d		2x2	2	5x5x16
Conv2d	120	5x5	1	120
Fully Connected 1	-	-	-	84
Fully Connected 2	-	-	-	3

Table 3 *Modified LeNet-5 architecture used by existing approach [20]*

Layer	Neurons	Filter Size	Stride	Size of Feature Map
Input Image	-	-	-	32x32x3
Conv2d	6	5x5	1	28x28x6
MaxPool2d		2x2	2	14x14x6
Conv2d	16	5x5	1	10x10x16
MaxPool2d		2x2	2	5x5x16
Conv2d	256	1x1	4	2x2x256
MaxPool2d		2x2	2	1x1x256
Conv2d	120	1x1	1	1x1x120
Flatten	-	-	-	120
Fully Connected 1	-	-	-	84
Fully Connected 2	-	-	-	3

Table 4 *Modified LeNet-5 architecture used by model 1*

Layer	Filters /Neurons	Filter Size	Stride	Size of Feature Map
Input Image	-	-	-	32x32x3
Conv2d	6	5x5	1	28x28x6
MaxPool2d		2x2	2	14x14x6
Conv2d	16	5x5	1	10x10x16
MaxPool2d		2x2	2	5x5x16

Layer	Filters /Neurons	Filter Size	Stride	Size of Feature Map
Conv2d	120	5x5	1	2x2x120
MaxPool2d		2x2	2	1x1x120
Flatten	-	-	-	120
Fully Connected 1	-	-	-	84
Fully Connected 2	-	-	-	3

Table 5 *Architectural details of LeNet Model 2*

Layer	Neurons	Filter Size	Stride	Size of Feature Map
Input Image	-	-	-	32x32x3
Conv2d	6	5x5	1	28x28x6
MaxPool2d		2x2	2	14x14x6
Conv2d	16	5x5	1	10x10x16
MaxPool2d		2x2	2	5x5x16
TripleConv	32	3x3	1	5x5x32
MaxPool2d		2x2	2	3x3x32
Conv2d	128	1x1	4	1x1x128
Conv2d	256	1x1	1	1x1x256
Conv2d	120	1x1	1	1x1x120
Flatten	-	-	-	120
Fully Connected 1	-	-	-	84
Fully Connected 2	-	-	-	3

Table 6 *Architectural details of LeNet Model 3*

Layer	Neurons	Filter Size	Stride	Size of Feature Map
Input Image	-	-	-	32x32x3
Conv2d	6	5x5	1	28x28x6
MaxPool2d		2x2	2	14x14x6
Conv2d	64	5x5	1	10x10x64
MaxPool2d		2x2	2	5x5x64
Conv2d	128	1x1	4	2x2x128
MaxPool2d		2x2	2	1x1x128
TripleConv	32	3x3	1	1x1x32
MaxPool2d		2x2	2	1x1x32
TripleConv	64	3x3	1	1x1x64
MaxPool2d		2x2	2	1x1x64
Conv2d	120	1x1	1	1x1x120
Flatten	-	-	-	120
Fully Connected 1	-	-	-	84
Fully Connected 2	-	-	-	3

Table 7 *Architectural details of LeNet model 4*

Layer	Neurons	Filter Size	Stride	Size of Feature Map
Input Image	-	-	-	32x32x3
Conv2d	6	5x5	1	28x28x6
MaxPool2d		2x2	2	14x14x6
Conv2d	16	5x5	1	10x10x16
MaxPool2d		2x2	2	5x5x16
TripleConv	32	3x3	1	5x5x32
MaxPool2d		2x2	2	3x3x32
TripleConv	64	3x3	1	3x3x64
MaxPool2d		2x2	2	2x2x64
TripleConv	128	3x3	1	2x2x128
MaxPool2d		2x2	2	2x2x128
Conv2d	256	1x1	1	1x1x256
Conv2d	512	1x1	1	1x1x512
Conv2d	120	1x1	1	1x1x120
Flatten	-	-	-	120
Fully Connected 1	-	-	-	84
Fully Connected 2	-	-	-	3

5.1 Triple Convolution Architecture

Fig. 9 shows the architecture of triple convolution block used in handwriting feature extraction in the proposed ensemble model. Each block following the pattern Conv2d-BatchNorm2d-LeakyReLU, repeated 3 times. This design ensures hierarchical feature learning by combination spatial filtering, normalization and incorporated nonlinearity. The repeated pattern ensures that each stage produces normalized, non-saturated and discriminative feature maps. The Conv2d layers capture stroke-level and contour information, BatchNorm2d stabilizes activation distributions and LeakyReLU preserves gradient flow for all activation ranges in pen-pressure variations. This enables the model to efficiently learn fine-grained handwriting characteristics for improving classification accuracy.
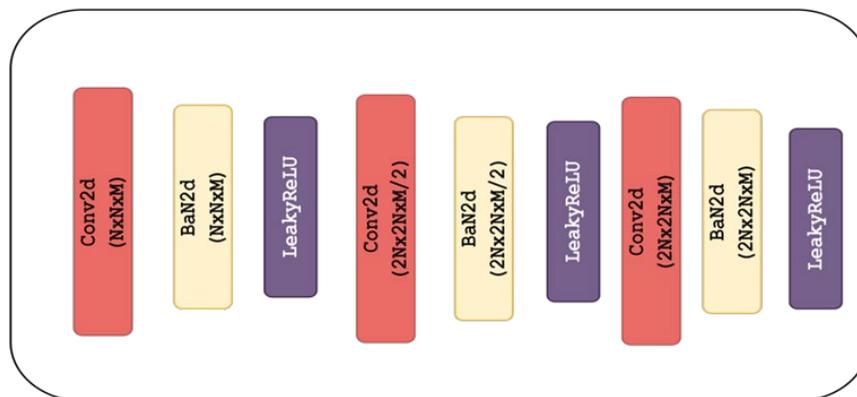
Fig. 9 *Triple convolution architecture*

6. Results and Discussions

The overview of the proposed system is shown in Fig. 1 depicts four variants of CNN models based on LeNet-5 architectures with varying layer configurations. The performance metrices for the models in terms of error rate and accuracy are presented in Fig. 10.

The performance of each individual architecture used in the study is depicted in Fig. 10, with consistent training parameters across all models. Among the architectures, Model 4 demonstrated the highest accuracy and the lowest error rate. The experimental results indicate that architectures with more layers dedicated to feature extraction tend to perform better, effectively classifying the three classes of handwriting.

As demonstrated in Fig. 10, the ensemble of the four CNN models in the proposed method clearly showed higher accuracy compared to individual models and the existing model. The accuracy of the proposed ensemble approach for both training and testing is summarized in Table 9. The hyperparameters for all models were fine-tuned to achieve higher accuracy than the existing approach.

The accuracy of existing approach was 85.26% ,88.73%with error rate of 14.73% [20][21] respectively. It is clear from Fig. 10 that the developed ensemble CNN model based on LeNet 5 architecture successfully attained a remarkable accuracy of 96.86% in classifying the three classes of handwriting. Furthermore, ensemble approach showed a good reduction in error rate as compared to all the individual models. The improved results of the proposed ensemble method can be attributed to its diverse architectural variations and the stacking strategy of combining the predictions. Hence reducing the overall prediction variance of individual models and enhancing the robustness of the ensemble model.

The comparative analysis between the individual CNN models and the ensemble model highlights the superior performance of the ensemble approach. The ensemble model not only outperformed individual models in terms of accuracy but also showed reduced error rates. Thus, validating the effectiveness of the ensembling technique.

The hyperparameters for all the four proposed Lenet-5 architectural variants were tuned extensively with varying ranges as described in Table 1, to achieve optimal configuration values for best classification results. The tuning process ensured that each model operated at its best possible configurations detailed in Table 1.

A detailed comparative analysis involved comparing all the four models and proposed ensemble model with the existing approaches Rosli et al. [20] and Isa et al. [21] which have attained an accuracy of 85.26% and 88.73% respectively. Both studies are based on LeNet-5 architecture without any enhancement as introduced by the present study (such as triple convolutions, batch normalization and optimized depth). The proposed approach built on these said enhancements showed improved accuracy, which can be credited to enhanced feature extraction strategy via triple convolution blocks. The ensemble model achieved an excellent accuracy of 96.86%, with a significantly lower error rate of 3.14%. These findings suggest that increasing the depth of feature extraction layers leads to better classification performance. Moreover, the ensemble approach provides a robust and accurate method for early identification and classification of dyslexic handwriting, which is crucial for timely intervention and remedial solution to the affected children.

Fig. 11 and Fig. 12 represent the confusion matrices for training and testing dataset. The proposed ensemble approach yielded exceptional performance metrics during training, achieving nearly perfect precision, recall, and F1-Score. This indicates that the model has effectively learned the patterns and characteristics of the training data well to distinguish between the classes. Likewise, the precision values for all three classes are exceptionally high, particularly for the Corrected and Reversed classes, both achieving 99.98%. This suggests the model is highly accurate in identifying positive instances of the three classes. The recall values are also high, with the Normal class achieving the highest recall of 99.96%, suggesting the model successfully identifies nearly all relevant instances of each class. This higher consistency in F1-Scores across all classes reflect the model's overall robustness and accuracy during training.

The enhanced performance values for the minority classes (Corrected and Reversed) can be attributed to the inclusion of class weighted loss formulation as has been described in Dataset Section 4. By using the approach proposed by Cui et al [38] to mitigate class imbalance, the proposed model effectively reduced the bias towards the Majority class (Normal), hence leading to balanced behaviour across all the three classes.

In addition to comparing the proposed ensemble with individual CNN models, we have also evaluated and compared the effectiveness of various ensembling strategies viz bagging, boosting and stacking. As shown in Table 8, proposed stacking ensemble approach outperformed both bagging and boosting across all performance metrices. These findings strengthen the option of choosing stacking ensemble, resulting in capturing inter-model relationships using meta-learner more effectively. Hence, reducing prediction variance and improving overall accuracy of the proposed ensemble.

Table 8 *Performance comparison of various ensemble approaches*

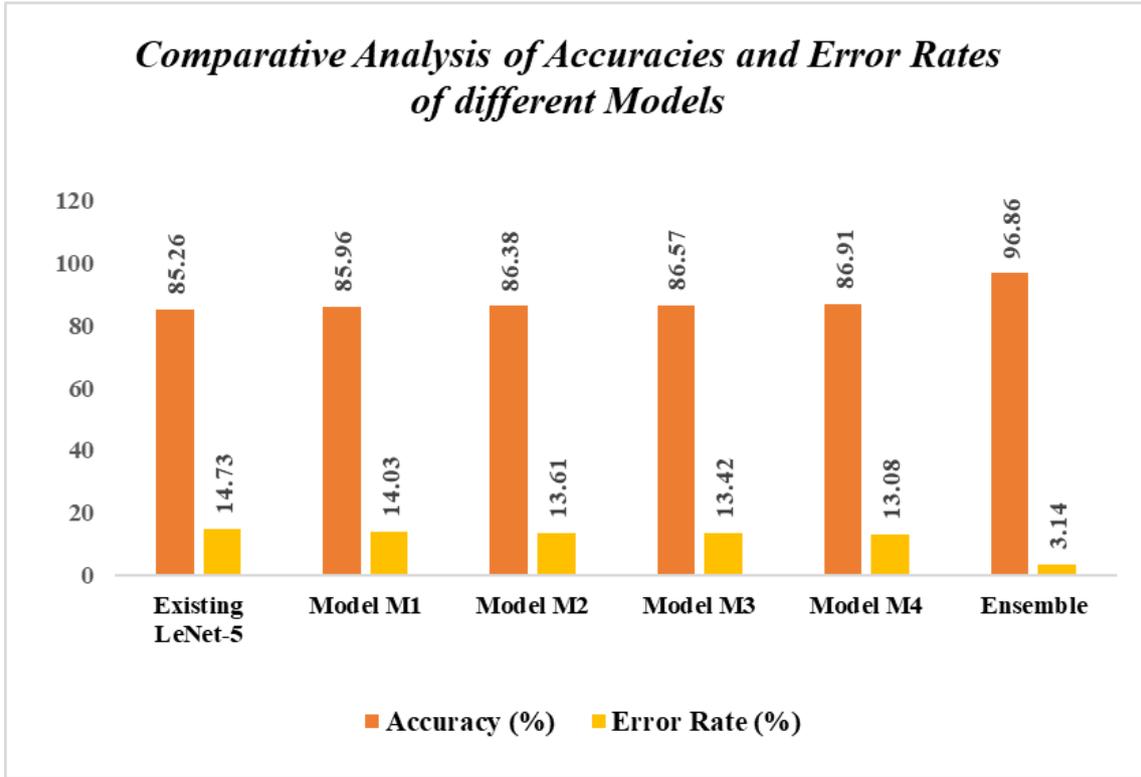| Ensemble Method	Accuracy	Precision	Recall	F1-Score
Bagging	87.34%	89.12%	87.34%	87.42%
Boosting	87.37%	89.13%	87.37%	87.45%
Stacking	98.96%	98.88%	98.96%	98.92%

Fig. 10 *Comparative analysis of accuracy and error rates of different models*
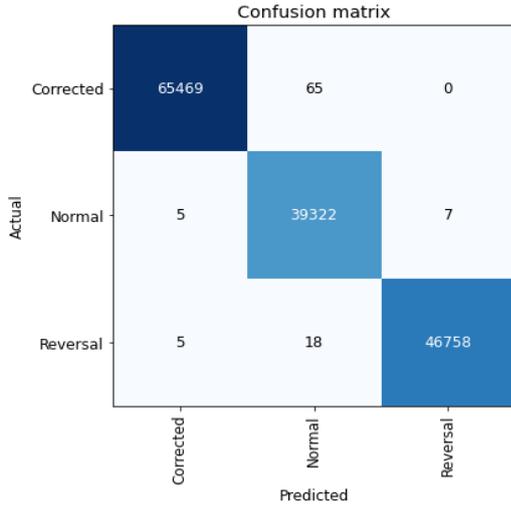

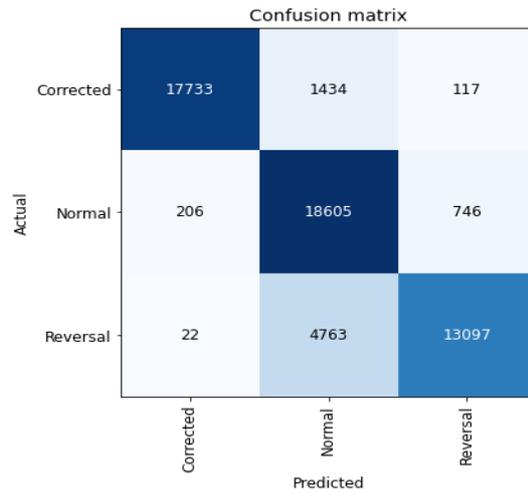**Fig. 11** *Confusion matrix of ensemble method for training*

Fig. 12 *Confusion matrix of ensemble method for testing*

While evaluating the testing dataset, the ensemble model maintained strong performance overall, with a slight drop in accuracy compared to the training dataset. The accuracy achieved was 96.86% for the Corrected class, 87.39% for the Normal class, and 90.04% for the Reversed class, indicating strong generalization capabilities across all classes. The precision values followed the trend Corrected > Reversed > Normal, with the Corrected class having the highest precision at 98.73%. The recall and F1-Score values followed a different trend: Corrected > Normal > Reversed, with the Corrected class leading in both metrics. These results validate the ensemble model's effectiveness across all classes and confirm its robustness on unseen test data.

Comparing the ensemble approach with the existing model [20], the ensemble model significantly outperformed across all performance metrices, attaining a training accuracy of 99.98% and a testing accuracy of 96.86%, significantly higher than the existing model (accuracy= 85.96%). The training error rate for the ensemble model

has been remarkably low at 0.6%, compared to the existing model's 14.03%, and the testing error rate was 3.14%, much lower than the existing model's rate. Additionally, the ensemble model demonstrated faster real-time prediction with an average time of 0.136 seconds, compared to the existing model's 0.196 seconds, hence making it more practical for real-time applications.
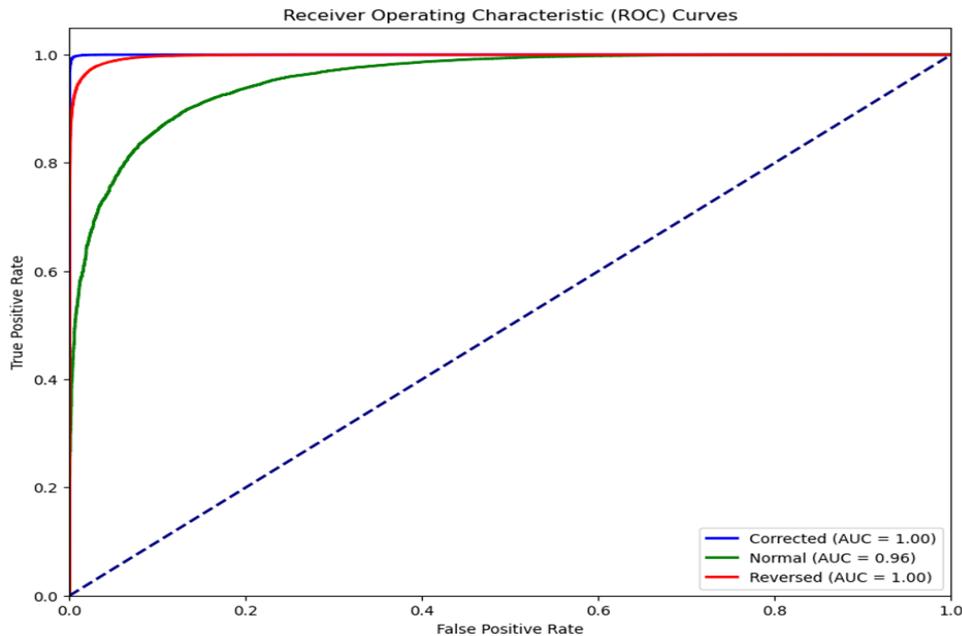

Fig. 13 *ROC curve*

The Fig. 13 above presents the Receiver Operating Characteristic (ROC) curves for the classification performance of the proposed ensemble across three handwriting classes: Corrected, Normal, and Reversed. The ROC curve illustrates the model's diagnostic ability, by plotting the True Positive Rate (TPR)(sensitivity) against the False Positive Rate (FPR) (1 - specificity) at various threshold settings. Analyzing the curves, the Corrected (blue line) and Reversed (red line) classes both exhibit perfect classification with AUC values of 1.00, indicating pure or perfect classification. The Normal class (green line) shows an AUC of 0.96, which signifies excellent classification performance with a high TPR and a low FPR. These experimental findings demonstrate that the proposed ensemble model performs exceptionally well, achieving perfect accuracy for the Corrected and Reversed classes and near-perfect accuracy for the Normal class, highlighting the model's robustness and reliability in dyslexia handwriting classification.

To better reflect the model's performance via the ROC curve shown in Fig. 14, we have presented the model prediction visualization on sample images from all the three classes. In Corrected and Reversed classes, all samples are correctly with high probability and lower loss, which is consistent with models near perfect separability for all the three categories. For the Normal class smaller number of misclassifications occur. This aligns according to lower values of AUC (0.96). This visualization provides a clearer understanding of model's strength, showcasing that most of the misclassifications occur in boundary cases within Normal class, while other classes remain robust.
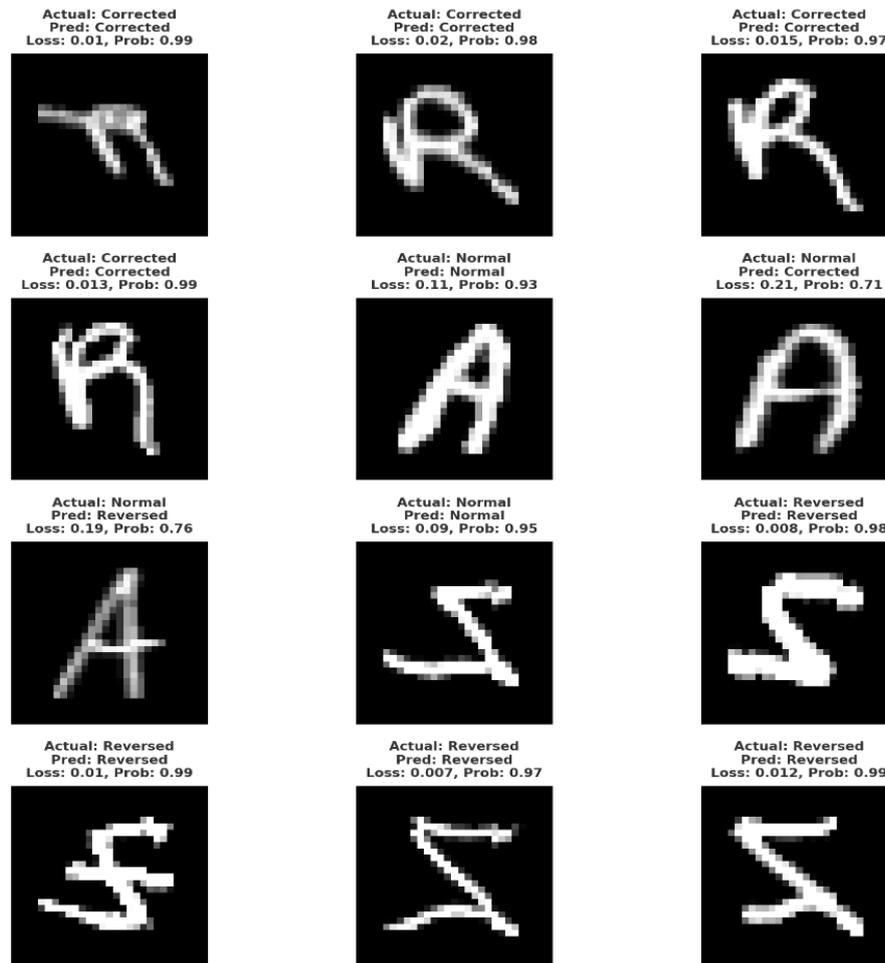
Fig. 14 *Proposed models predictions on sample images pertaining to all three classes. Each image displays the actual class, predicted class, loss value, and prediction probability*

The outperforming nature of the proposed ensemble model can be attributed to various factors such as increased depth and complexity of the individual models, incorporating multiple CNN architectures with varying triple convolutional layers, enabling the model to extract more intricate features from the data. The stacking ensemble combined the outputs of four CNN models, reducing variance and improving overall performance. Also, the careful tuning of hyperparameters ensured optimal performance, contributing to the high accuracy and low error rates. The near-perfect scores on the training dataset as shown in Table 9 indicate that the ensemble model has effectively learned the underlying patterns exhibited by the data, which is further validated by its strong performance on the testing dataset as shown in Table 10 hence justifying it as a robust learning without overfitting.

Overall, the ensemble approach demonstrates significant improvements over the existing model, with higher accuracy, lower error rates, and faster prediction times as shown in Table 11. To evaluate the performance improvements of the proposed model, an ablation study was conducted against the state of art models. The study compared proposed model with existing CNN based methods, Deep CNN methods, transformer-based methods and final stacking approach. As shown in Table 12, the baseline LeNet-5 model proposed by Rosli et al. [20] achieved an accuracy of 85.26%, with precision of 86.10% and F1-score of 85.53%. Likewise slight improvements were observed in the modified CNN architecture proposed by Isa et al. [21]. Also, transformer-based models like Vision Transformer (ViT) [28] and DeiT [30] performed significantly better, achieving accuracies of 94.8% and 95.7%, respectively. The proposed individual CNN Models 1 to 4, incorporating varying number of triple convolution layers, achieved incremental improvements up to 86.91%. However, the proposed stacking ensemble model outperformed all, achieving 96.86% accuracy, 98.88% precision, and 98.92% F1-score. These results confirm that while architectural enhancements also help, ensemble learning—particularly stacking—offers the most robust and accurate classification for dyslexic handwriting, highlighting the potential of the proposed method for accurate and efficient classification.

Table 9 *Performance metrices of ensemble method on training dataset*

Class	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Corrected	99.95	99.98	99.90	99.93
Normal	99.93	99.78	99.96	99.86
Reversed	99.97	99.98	99.95	99.96

Table 10 *Performance measures of Ensemble Method on Testing Dataset*

Class	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Corrected	96.86	98.73	91.95	95.22
Normal	87.39	75.01	95.13	83
Reversed	90.04	93.81	73.24	82.25

Table 11 *Comparison of Existing and Proposed Ensemble Method*

Model	Training Accuracy	Testing Accuracy	Training Error Rate	Testing Error Rate	Realtime Prediction Time (Seconds)
Existing Model	85.96%	85.96%	14.03%	14.03%	0.196 s
Proposed Model	99.98%	96.86%	0.6%	3.14%	0.136 s

Table 12 *Comparison of Proposed model with state-of-art methods*

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
LeNet-5 (Rosli et al. [20])	85.26	86.10	85	85.53
Modified CNN (Isa et al. [21])	88.73	89	88.6	88.74
Deep CNN (Ramlan et al. [18])	87.44	88.02	87.30	87.66
ViT (Dosovitskiy et al. [28])	94.8	92.9	91.6	92.3
DeiT (Touvron et al. [30])	95.7	93.8	92.5	93.1
Proposed Model 1	85.96	86.40	85.80	86.10
Proposed Model 2	86.38	86.90	86.30	86.59
Proposed Model 3	86.57	87.10	86.50	86.80
Proposed Model 4	86.91	87.40	86.90	87.14
Proposed Stacking Ensemble	96.86	98.88	98.96	98.92

7. Conclusion

This study addresses the most prominent learning disability "dyslexia" affecting millions of children worldwide, with a notable prevalence in India. In this paper, the key objective was to improve the accuracy and efficiency of classifying dyslexia handwriting images by exploring different LeNet-5 architectures. The proposed ensemble architecture, integrating four LeNet-5 variants with varying number of triple convolution layers, has demonstrated remarkable success. The ensemble model achieved an outstanding accuracy of 99.98% (training) and 96.86% (testing), significantly outperforming the existing approaches. These findings underscore the potential of advanced convolutional neural network structures in accurately identifying dyslexic handwriting patterns. The ROC curve analysis further validates the efficacy of the proposed approach, with perfect

classification for the Corrected and Reversed classes (AUC = 1.00) and excellent performance for the Normal class (AUC = 0.96). These findings underscore the robustness and efficiency of the proposed ensemble CNN model, making it a promising solution for accurate dyslexia handwriting classification.

Beyond the technical achievements, this study highlights the urgent need for timely and precise identification of dyslexic children. By providing a robust and efficient tool for screening dyslexia, our approach can play a key role by ensuring that affected dyslexic children receive the necessary support and remedial interventions as early as possible. This, in turn, contributes to protecting their educational journey and future potential. The promising results from our ensemble model assures that with the right technological innovations, we can be able to make a significant difference in the lives of those with learning disabilities.

8. Future Scope

While the proposed ensemble model demonstrates high accuracy and robustness on this dataset, its evaluation is limited it only. This restricts the generalizability of the model to handwritten images from other populations. Presently, there is lack of such publicly available dataset related to dyslexia with consistent labelling and structure, which poses a challenge for broader validation. So as a part of future work, we aim to evaluate the proposed model on external datasets from different demographic groups with varying ages and handwriting contexts to better understand its performance in real world applications.

Acknowledgement

Conflict of Interest

The authors declare that they have no conflict of interest in this paper.

Data Availability

Data will be made available on reasonable request.

Author Contributions

*Author Contributions: **Conceptualization**, T.G.J., S.M.K and S.Y.B; **methodology**, T.G.J., S.M.K and S.Y.B;.; **software**, S.M.K and S.Y.B;.; **validation**, T.G.J., S.M.K and S.Y.B;.; **formal analysis** T.G.J., S.M.K and S.Y.B; **investigation**, S.M.K and S.Y.B; **resources**, S.M.K and S.Y.B; **writing—original draft preparation**, T.G.J.; **writing—review and editing**, T.G.J., S.M.K and S.Y.B; **visualization** T.G.J., S.M.K and S.Y.B; **supervision**, S.M.K and S.Y.B; **project administration**, S.M.K and S.Y.B;. All authors have read and agreed to the published version of the manuscript.*

References

[1]   National Association of Special Education Teachers (NASET) (2008) *Characteristics of Children with Learning Disabilities*, NASET LD Report, (3), 1–26, https://media.naset.com/fileadmin/user_upload/LD_Report/Issue__3_LD_Report_Characteristic_of_LD.pdf

[2]   Masters in Special Education Degree Program Guide (2022) *5 Most Common Learning Disabilities*, Masters in Special Education Degree Program Guide, https://www.masters-in-special-education.com/lists/5-most-common-learning-disabilities/

[3]   Dyslexia Association of India (2022) *What is Dyslexia*, Dyslexia Association of India, https://www.dyslexiaindia.org.in/what-dyslexia2.html

[4]   "Different Types of Dyslexia & How to Identify Them | NeuroHealth," *Neurohealthah.com*, 2022, [Online]. Available: https://neurohealthah.com/blog/types-of-dyslexia/

[5]   Ribeiro, M. H. D. M., da Silva, R. G., Moreno, S. R., Mariani, V. C., & Coelho, L. dos S. (2022). Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting. *International Journal of Electrical Power & Energy Systems, 136*, 107712. https://doi.org/10.1016/j.ijepes.2021.107712

[6]   Książek, W., Hammad, M., Pławiak, P., Acharya, U. R., & Tadeusiewicz, R. (2020). Development of novel ensemble model using stacking learning and evolutionary computation techniques for automated hepatocellular carcinoma detection. *Biocybernetics and Biomedical Engineering, 40*(4), 1512–1524.https://doi.org/10.1016/j.bbe.2020.08.007

[7]   Djarum, D. H., Ahmad, Z., & Zhang, J. (2023). Reduced Bayesian Optimized Stacked Regressor (RBOSR): A highly efficient stacked approach for improved air pollution prediction. *Applied Soft Computing, 144*, 110466. https://doi.org/10.1016/J.ASOC.2023.110466

[8]   Frid, A., & Breznitz, Z. (2012). An SVM based algorithm for analysis and discrimination of dyslexic readers from regular readers using ERPs. In *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel (IEEEI 2012)*. IEEE. https://doi.org/10.1109/EEEI.2012.6377068

[9]   Perera, H., Shiratuddin, M. F., Wong, K. W., & Fullarton, K. (2018). EEG Signal Analysis of Writing and Typing between Adults with Dyslexia and Normal Controls. *International Journal of Interactive Multimedia and Artificial Intelligence, 5*(1), 62. https://doi.org/10.9781/ijimai.2018.04.005

[10]   Rello, L., & Ballesteros, M. (2015). Detecting readers with dyslexia using machine learning with eye tracking measures. In *Proceedings of the 12th Web for All Conference (W4A 2015).* https://doi.org/10.1145/2745555.2746644

[11]   A, J. P., Bhargavi, R., & Harish, B. (2019). Predictive Model for Dyslexia from Eye Fixation Events. *International Journal of Engineering and Advanced Technology, 9*(1S3), 235–240. https://doi.org/10.35940/ijeat.a1045.1291s319

[12]   Jothi Prabha, A., & Bhargavi, R. (2020). Predictive Model for Dyslexia from Fixations and Saccadic Eye Movement Events. *Computer Methods and Programs in Biomedicine, 195*, 105538. https://doi.org/10.1016/j.cmpb.2020.105538

[13]   Chakraborty, V., & Sundaram, M. (2020). Machine learning algorithms for prediction of dyslexia using eye movement. *Journal of Physics: Conference Series, 1427*(1), 012012. https://doi.org/10.1088/1742-6596/1427/1/012012

[14]   Radford, J., Richard, G., Richard, H., & Serrurier, M. (2021). Detecting dyslexia from audio records: An AI approach. In *HEALTHINF 2021: 14th International Conference on Health Informatics (BIOSTEC 2021), 5*,58–66. https://doi.org/10.5220/0010196000580066

[15]   Płoński, P., *et al.* (2017). Multi-parameter machine learning approach to the neuroanatomical basis of developmental dyslexia. *Human Brain Mapping, 38*(2), 900–908. https://doi.org/10.1002/hbm.23426

[16]   Richard, G., & Serrurier, M. (2020). *Dyslexia and Dysgraphia prediction: A new machine learning approach.* arXiv preprint arXiv:2005.06401. http://arxiv.org/abs/2005.06401

[17]   Spoon, K., Crandall, D., & Siek, K. (2019). Towards detecting dyslexia in children's handwriting using neural networks. *International Conference on Machine Learning, AI and Society of Good Workshop*, 1–5.

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

[19]   Jan, T. G., & Khan, S. M. (2023). A systematic review of research dimensions towards dyslexia screening using machine learning. *Journal of the Institution of Engineers (India): Series B, 104*(2), 511–522. https://doi.org/10.1007/s40031-023-00853-8

[20]   Bin Rosli, M. S. A., Isa, I. S., Ramlan, S. A., Sulaiman, S. N., & Maruzuki, M. I. F. (2021). Development of CNN transfer learning for dyslexia handwriting recognition. In *Proceedings of the 2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2021)*, (pp. 194–199). IEEE. https://doi.org/10.1109/ICCSCE52189.2021.9530971

[21]   Isa, I. S., Zahir, M. A., Ramlan, S. A., Li-Chih, W., & Sulaiman, S. N. (2021). CNN comparisons models on dyslexia handwriting classification. *ESTEEM Academic Journal, 17*, 12–25.

[22]   Sihvonen, A. J., Virtala, P., Thiede, A., Laasonen, M., & Kujala, T. (2021). Structural white matter connectometry of reading and dyslexia. *NeuroImage, 241*, 118411. https://doi.org/10.1016/j.neuroimage.2021.118411

[23]   Kuhl, U., *et al.* (2020). The emergence of dyslexia in the developing brain. *NeuroImage, 211*, 116633. https://doi.org/10.1016/j.neuroimage.2020.116633F

[24]   Ramus, F., Altarelli, I., Jednoróg, K., Zhao, J., & Scotto di Covella, L. (2018). Neuroanatomy of developmental dyslexia: Pitfalls and promise. *Neuroscience & Biobehavioral Reviews, 84*, 434–452. https://doi.org/10.1016/j.neubiorev.2017.08.001

[25]   Fraga González, G., *et al.* (2018). EEG resting state functional connectivity in adult dyslexics using phase lag index and graph analysis. *Frontiers in Human Neuroscience, 12*, 341. https://doi.org/10.3389/fnhum.2018.00341

[26] Oliaee, A., Mohebbi, M., Shirani, S., & Rostami, R. (2022). Extraction of discriminative features from EEG signals of dyslexic children; before and after the treatment. *Cognitive Neurodynamics, 2*. https://doi.org/10.1007/s11571-022-09794-2

[27] Jasira, K. T., Laila, V., & Ahmed, P. J. (2023). DyslexiScan: A dyslexia detection method from handwriting using CNN LSTM model. In *Proceedings of the 2023 International Conference on Innovation in Engineering and Technology (ICIET 2023)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICIET57285.2023.10220750

[28] Dosovitskiy, A., *et al.* (2020). *An image is worth 16×16 words: Transformers for image recognition at scale.* arXiv preprint arXiv:2010.11929

[29] Sait, A. R. W., & Alkhurayyif, Y. (2025). Lightweight hybrid transformers-based dyslexia detection using cross-modality data. *Scientific Reports, 15*, 17054. https://doi.org/10.1038/s41598-025-01235-4

[30] Touvron, H., Cord, M., & Jégou, H. (2022). DeiT III: Revenge of the ViT. In *Proceedings of the European Conference on Computer Vision (ECCV)*, (pp. 516–533). Springer Nature Switzerland.

[31] Sharma, S., & Jain, M. (2023). DxDetekt: A dyslexia detection method from handwriting using ensemble method. *GRENZE International Journal of Engineering and Technology (GIJET), 9*(2), 79–84.

[32] Ali, H., Ahmed, S., & Khan, A. (2024). *Multimodal ensemble learning with conditional feature fusion for dysgraphia detection.* arXiv preprint arXiv:2408.13754.

[33] Perera, S., Balasooriya, M., & Chandrasekara, A. (2023). Early dyslexia detection using ensemble machine learning applied to ERP components of EEG. *Biomedical Signal Processing and Control, 86*, 105274.

[34] Zhang, L., Wang, K., & Chen, Y. (2023). *EEG connectivity analysis using denoising autoencoder for dyslexia classification.* arXiv preprint arXiv:2311.13876.

[35] Gupta, A., & Singh, R. (2023). Ensemble feature selection for dyslexia detection using machine learning. In *Proceedings of the International Conference on Advances in Computing and Data Sciences (ICACDS)*, (pp. 389–398). Springer, Singapore.

[36] Kumar, P., & Bansal, S. (2023). Quiz-based dyslexia detection using XGBoost ensemble learning. *Indonesian Journal of Artificial Intelligence (IJAI), 2*(4), 221–229.

[37] Kunhoth, J., Al Maadeed, S., Saleh, M., & Akbari, Y. (2023). CNN feature and classifier fusion on novel transformed image dataset for dysgraphia diagnosis in children. *Expert Systems with Applications, 231*, 120740. https://doi.org/10.1016/j.eswa.2023.120740

[38] Cui, Y., Jia, M., Lin, T.-Y., Song, Y., & Belongie, S. (2019). Class-balanced loss based on effective number of samples. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 9260–9269). IEEE. https://doi.org/10.1109/CVPR.2019.00949.