THE SCIENTIFIC TEMPER (2026) Vol. 17 (4): 6015-6024

E-ISSN: 2231-6396, ISSN: 0976-8653

Doi: 10.58414/SCIENTIFICTEMPER.2026.17.4.11

https://scientifictemper.com/

RESEARCH ARTICLE

A Customized CNN-Based Framework for Learning Disability Detection Using Handwriting Image Classification

Soumya K[1*], Dr. P Joseph Charles[2], Dr. Kavitha S[3]

Abstract

Handwriting analysis has been widely explored as a supportive computational approach for understanding writing patterns frequently associated with learning difficulties in children. While deep learning techniques, particularly transfer learning algorithms, have demonstrated strong performance in image-based classification tasks, their reliance on large pre-trained models often limits adaptability and computational efficiency. To address these limitations, this study proposes a custom Convolutional Neural Network (CNN) architecture specifically designed for the classification of children's handwriting patterns from images. The model consists of five Conv2D layers to extract and learn significant spatial features, promoting high classification accuracy while maintaining simplicity and reduced computational complexity. A carefully curated dataset of handwriting samples classified as "correct," "casual," and "reverse" was used to evaluate the model. This work does not claim to provide a clinical diagnosis of learning disabilities; rather, it presents a handwriting pattern classification framework that may support educational analysis and assistive screening when combined with validated assessment procedures.

Keywords: Learning disability detection, Convolutional neural network, Custom CNN architecture, Handwriting image classification, Deep learning, Educational AI, Early intervention.

Introduction

Normally, children with learning disabilities (LD) have small neurological differences that may affect their reading, writing, or math skills. Identifying the challenges and providing the necessary support is a key factor for the healthy emotional and cognitive development of the child. The methods of assessment usually depend on the knowledge of the experts; they are procedural and require trained specialists. Such resources are not always affordable for schools or areas with limited educational facilities. New solutions of AI and deep learning have been opened for automating medical and educational screening processes.

Convolutional Neural Networks (CNNs) have been very effective in image-based classification (Catts et al., 2024). Their capability of the analysis of handwriting opens the way for the detection of the learning patterns co-related to the learning challenges. This approach makes the handwriting analysis method a feasible, scalable, and non-invasive technique to screen children for potential LD.

Although fast learning and large pre-trained models give strong results, they also have significant limitations. Their size, the demand for computation, and the necessity for large outside data-sets can make them not the best choice for real-time use or locations with limited resources, such as schools in rural areas or small healthcare centres. As these general-purpose models are trained on a variety of image features rather than specific features in children's handwriting that makes them less relevant for LD detection. This research proposes a lightweight, explainable CNN for

[1]Research Scholar, St.Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli-620 002, Tamil Nadu, India

[2]Assistant professor, St.Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli-620 002, Tamil Nadu, India

[3]Associate Professor, Christ University, Bangalore, India

***Corresponding Author:** Dr. Kavitha S, Research Scholar, St.Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli-620 002, Tamil Nadu, India, E-Mail: soumyak@jainuniversity.ac.in

How to cite this article: Soumya, K., Charles, P.J., Kavitha S. (2026). A Customized CNN-Based Framework for Learning Disability Detection Using Handwriting Image Classification. The Scientific Temper, **17**(4):6015-6024.

Doi: 10.58414/SCIENTIFICTEMPER.2026.17.4.11

Source of support: Nil

Conflict of interest: None.

Received: 26/03/2026 **Accepted:** 02/04/2026 **Published:** 25/04/2026

the identification of learning disabilities using handwriting analysis as a solution to these problems (Fletcher et al., 2018). The model is small but still powerful; it leverages the distinct handwriting features to differentiate the normal writing from the ones that are linked to learning difficulties.

For the experiment, we have a handwriting dataset with three classes: correct, casual, and reverse. Each class depicts different writing behaviours that indicate spelling variations commonly spoken in LD. This paper is arranged as follows: Section II describes different research works related to CNN-based handwriting evaluation for LD identification. Section III illuminates the architecture of the proposed network. Section IV is about the dataset and data collection methods. Section V focuses on the experimental setup and criteria for evaluation. Section VI talks about the results and Section VII offers the final thoughts as well as the upcoming research deas.

This study focuses on the classification of handwriting patterns based on observable structural characteristics. The labels used in this work (Correct, Casual, and Reverse) represent handwriting styles and do not correspond to clinically diagnosed learning disabilities. No medical, psychological, or educational diagnostic assessments were used during dataset labeling. Therefore, the proposed model should not be interpreted as a diagnostic tool, but rather as a supportive computational method for handwriting analysis in educational research contexts.

Related Work

Research on the Machine-based computational detection of learning disabilities has progressed significantly with the Arrival of machine learning and deep learning techniques. Traditional diagnostic tools were based on Clinical and psychological evaluations and behavioral observations, even though these methods are accurate and resource-intensive and they are interpreted differently across individuals. Recent studies were looked into the digital technologies for early identification of diseases, utilizing various criterias such as academic performance and handwriting patterns (Rathod, Wagh, & Mourya, 2025).

The study of handwriting is receiving a good attention from researchers. This interest comes from its well known connection with learning disabilities. This is because handwriting is likely linked to learning disabilities such as dysgraphia and dyslexia. Early studies incorporated feature-engineering methods. They measured criterias like stroke width, slant, spacing between strokes, and curvature. These features were manually extracted to find useful handwriting patterns. After extracting these features, researchers fed them to the conventional machine-learning classifiers such as SVMs, k-NN, and Random Forests (Zaibi & Bezine, 2024). While this approach gained moderate classification performance, it did not scale well and was not applicable to the varieties of individual handwriting styles.

Recent research by Nazir et al., 2024, adresses the lack of transperency of conventional deep learning models, used for the diagnosis of brain tumors. They customized convolutional neural network for detecting brain tumors with three explainable-AI methods -SHAP, LIME and Grad-CAM, hereby providing visual and analytical clarity on the model's predictions. Based on the BR35H MRI dataset, authors state that the validation accuracy is 98.67 0.5, and the precision-recall values are very high, which ensures the accuracy of the model. This research shows that interpretability is becoming a key focus in mediacal imaging. It helps create more accurate and reliable systems for early brain-tumor detection Nazir et al. (2024).

Liang and Chen (2025) developed CCNN-QL as a customized deep learning model for energy consumption rate prediction (Liang & Chen, 2025). The system was created to enhance its ability to forecast energy usage rates. The researchers created a hybrid system through the combination of Q-Learning with CNN to enhance prediction accuracy. The researchers used data from multiple cities to train and test their model. The researchers demonstrated that CCNN-QL outperformed traditional ML models and basic CNNs in energy usage prediction accuracy.

In their 2025 study, Strock, Mistry and Menon built personalized deep CNN model to understand reason behind mathematical difficulties in children (Strock, Mistry, & Menon, 2025). The personalized model was trained using behaviour and brain activity of individual child, and they reached about 95% accuracy during learning phase. The cross-entropy loss shows decreased result over time, which shows that the models learned the task well. Although the authors did not report an exact loss percentage in their study. They found that children with math difficulties have more neural activity than usual. This slows down their learning speed and thus it is hard for identifying the maths digit clearly. The study highlights that personalized AI models can detect the biological influences that lead to learning challenges.

Liu, Wang, and Tong (2024) developed a model called DysDiTect, which introduced a novel deep-learning framework to identify dyslexia in children from Chinese using handwriting data. Handwriting data was collected from a dictation task. Their model integrated a convolutional neural network (CNN) with positional encoding, and built a bidirectional LSTM to find the sequential pattern of writing (Liu, Wang, & Tong, 2024). DysDiTect was evaluated on over 100,000 character images collected from 1,064 students, including 483 with dyslexia and 581 typically developing dyslexia in children. The model achieved an accuracy of 83.2%, a sensitivity of 79.2%, and an area under the curve (AUC) of 0.912.

Materials And Methods

Problem Definition

Transfer learning models perform well in many vision tasks, especially ResNet and Inception. However, they require a lot of computational resources and take a longer time to predict learning disabilities. They are not specifically designed to learn handwriting features such as stroke shape, orientation, and spacing irregularities (Zaibi & Bezine, 2024; Liu, Wang, & Tong, 2024).

To address this limitation, a custom CNN was built from scratch to learn the required spatial features while keeping the computational cost low. The main goal of this study is to design an efficient, customized convolutional neural network that can classify handwritten images into three categories: correct, casual, and reverse. These categories represent different handwriting styles that may indicate learning disabilities. The expected outcome is to develop a model that not only achieves high accuracy but also has fast learning rates.

Handwriting analysis inherently depends on localized spatial patterns such as stroke orientation, curvature, spacing, symmetry, and character alignment. Convolutional neural networks are theoretically well-suited for such tasks due to their localized receptive fields and weight-sharing mechanisms, which impose an inductive bias toward learning spatially invariant features.

Unlike deep transfer-learning models trained on large-scale generic image datasets, a task-specific CNN with moderate depth can focus more effectively on fine-grained stroke-level features relevant to handwriting, without introducing unnecessary representational complexity. This design choice aligns with bias–variance tradeoff principles, reducing overfitting while maintaining sufficient model capacity for discriminative feature learning.

The progressive increase in filter depth across convolutional layers enables hierarchical feature extraction, where early layers capture low-level edge and stroke information and deeper layers encode higher-level structural writing irregularities. Batch normalization further stabilizes training by reducing internal covariate shift, while controlled architectural depth ensures computational efficiency. These theoretical considerations justify the effectiveness of the proposed lightweight CNN for handwriting pattern classification in learning-related writing analysis.

Dataset Preparation and Preprocessing

Dataset has been collected from two different sources- One for normal handwriting and another for abnormal handwriting. For normal handwriting, data have taken data from NIST Special Database 19, a standard benchmark dataset released by the U.S. National Institute of Standards and Technology (NIST). Special Database 19 contains NIST's entire corpus of training materials for hand printed documents and character recognition. It publishes Hand Printed Sample Forms from 3600 writers, 810,000 character images isolated from their forms, ground truth classifications for those images, reference forms for further data collection, and software utilities for image management and handling. Around 2000-3000 images have been chosen from the NIST Special Database 19 for our dataset. Totally there are around 50000 images from all the categories are used for the research.The dataset used in this study contains different handwriting samples that are categorized into three classes: Correct, Casual, and Reverse (Liu, Wang, & Tong, 2024). These classes contain normal, irregular, and mirrored handwriting patterns, which can be a reason for certain learning disabilities. These classes are aligned with handwriting characteristics that can be used to analyze during learning disability assessments. The correct class represents standard, legible handwriting patterns, while the casual class includes loosely formed or irregular handwriting. The reverse class contains mirror-image or reversed characters, a key trait observed in students with learning difficulties, such as dyslexia and dyscalculia.

All handwriting samples are first resized to 128×128 pixels to keep the input dimensions uniform.To reduce input complexity, these resized images are then converted from RGB to grayscale.During training, pixel values are normalized to the range [0, 1] to improve model stability.

A set of data-augmentation techniques is applied to improve generalization and robustness. These include rotating the images by +10° or –10°, zooming in or out by 10%, and horizontal flipping to account for variations in writing direction. Elastic distortions are added to mimic natural pen movements and brightness and contrast adjustments of ±15% are applied to simulate changes caused by pen pressure, lighting, and scanner quality. Overall, this preprocessing pipeline ensures that the model is trained on diverse variations, enabling it to handle real-world handwriting inconsistencies effectively

Handwriting Labeling Protocol

The handwriting samples used in this study were labeled based on observable structural and spatial characteristics of written characters rather than on clinically confirmed learning disability diagnoses. The labeling process followed a predefined, rule-based visual inspection protocol derived from commonly reported handwriting indicators in learning difficulty literature.

Samples categorized as Correct exhibit standard character orientation, consistent stroke direction, appropriate spacing, and legible letter formation. The Casual category includes handwriting samples characterized by irregular spacing, inconsistent stroke alignment, variable letter sizes, and loosely formed characters. The Reverse category consists of mirror-image or reversed characters, which are frequently reported as writing irregularities associated with learning-related writing difficulties.

Labeling criteria were applied uniformly across the dataset to maintain consistency. No medical, psychological, or educational diagnostic records were used during the labeling process. Consequently, the dataset represents handwriting-style classification and should not be interpreted as a clinically validated learning disability dataset.

Dataset Analysis

The effectiveness of any deep learning–based classification system is strongly influenced by the quality, diversity, and balance of the dataset used for training and evaluation. In this research, there are three classes' used namely correct (Normal handwriting), casual, and Reverse (abnormal handwriting). Images should be collected for all the classes. Normal handwriting dataset contains alphabets with multiple images of normal category to increase the accuracy of the model. Abnormal handwriting dataset gathered from children with learning disability and they are mostly mirrored version of images and will contain the images of the alphabets which are written by children with learning disabilities.

The Correct class contains handwriting samples that follow standard letter formation rules, exhibiting consistent stroke direction, appropriate spacing, and proper character orientation. These samples represent typical handwriting behavior observed in children without learning difficulties. The Casual class includes loosely structured handwriting samples characterized by inconsistent spacing, irregular stroke alignment, and variations in letter size and shape. Such patterns are commonly associated with mild writing irregularities and reduced motor control. The Reverse class comprises handwriting samples with mirror-image or reversed characters, a distinctive trait frequently observed in children with learning disabilities such as dyslexia and dyscalculia.

To ensure unbiased learning and fair evaluation, the dataset was designed to be class-balanced, with approximately 16,666 images per class. This balanced distribution prevents model bias toward any particular category and allows the classifier to learn discriminative features for all handwriting styles equally. From the complete dataset, 80% of the samples (40,000 images) were allocated for training, while the remaining 20% (10,000 images) were reserved for testing. The split was performed in a stratified manner, maintaining equal class representation in both subsets.

An important aspect of the dataset is its intra-class variability, which reflects real-world handwriting diversity among children. Variations in stroke thickness, writing pressure, letter spacing, and orientation are present across all classes. Such diversity ensures that the proposed CNN does not rely on superficial patterns but instead learns meaningful structural and spatial features related to handwriting behavior. Additionally, the dataset includes samples captured under different writing conditions, such as variations in pen pressure and scanning quality, further improving the robustness of the trained model.

Figure 1 illustrates representative handwriting samples from each class, highlighting the visual differences between standard, irregular, and reversed writing styles. These samples demonstrate that while some distinctions are visually apparent, others are subtle and require deep feature extraction capabilities, reinforcing the need for a customized convolutional neural network.

Overall, the dataset provides a comprehensive and well-balanced foundation for evaluating the proposed CNN architecture. Its size, diversity, and structured class organization make it suitable for training deep learning models aimed at early and automated learning disability screening, particularly in educational and clinical environments.

This study exclusively utilizes publicly available handwriting datasets and anonymized handwriting samples that do not contain personally identifiable information. No direct interaction with human participants was conducted by the authors. As the research does not involve clinical intervention, diagnosis, or collection of sensitive personal data, formal ethical committee approval was not required. All datasets were used strictly for academic and research purposes in accordance with their respective usage and licensing policies.

Model Architecture

Our proposed Personalized Convolutional Neural Network is composed of five convolutional layers designed to learn handwriting features step by step, starting from simple styles and moving to complex writing styles. The first layer uses 32 filters of size 3×3; then, batch normalization is applied, ReLU ensures non-linearity, and 2×2 Max Pooling reduces the feature map.

The next two layers are identical, but uses 64 and 128 filters, and Max Pooling reduces the features again. The fourth layer contains 128 filters without any pooling for the purpose of retaining spatial information. The fifth block increases the filters to 256, capable of extracting higher-level feature representations. After every convolutional layer, batch normalization is used to make training more stable.

The proposed CNN architecture is designed to progressively learn discriminative handwriting features, starting from low-level stroke patterns to high-level

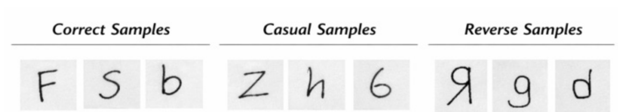
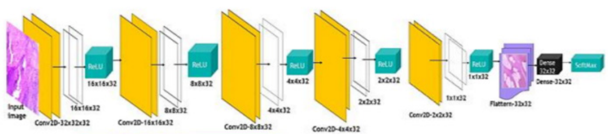


Figure 1: Sample handwriting images for 3 classes



(a)

Conv2D(32, 32)	Conv2D(32, 32)	Flatten(None,32)
MaxPooling2D(16, 16)	MaxPooling2D(16, 16)	Dense(None,4)
		Activation(None,4)
Activation(16,16)	Activation(16,16)	
Conv2D(32,32)	Conv2D(32,32)	
MaxPooling2D(16, 16)	MaxPooling2D(16, 16)	
Activation(16,16)	Activation(16,16)	

(b)

Figure 2: System model: (a) proposed CNN architecture (b) Architecture Details

structural representations. The initial convolutional layers focus on extracting basic features such as edges, curves, and stroke orientations, while deeper layers capture complex writing irregularities associated with learning disabilities. Max-pooling layers are employed in the early stages to reduce spatial dimensions and computational cost, whereas the fourth convolutional layer omits pooling to preserve critical spatial information. Batch normalization is applied after each convolutional layer to stabilize training and improve convergence. The final fully connected layers perform high-level reasoning and classification into Correct, Casual, and Reverse handwriting categories.

Training Process

We utilized the Adam optimizer (LR = 0.001) for model building, along with categorical cross-entropy loss, which is a better mechanism for handling multi-class classification problems.

To adjust the learning rate dynamically, the ReduceLROnPlateau scheduler is used; it cut shot the learning rate whenever the validation loss stops showing performance for five consecutive epochs. While training the model, we also used an EarlyStopping mechanism that stops the learning process if no improvement in the validation process is observed for ten epochs, after which the best-performing weights are automatically restored to overcome overfitting issues. A batch size of 32 and a maximum training limit of 50 epochs is incorporated in the proposed model

Algorithm:Training Procedure of the Proposed Custom CNN

Input:

$D = \{(x_i, y_i)\}$, for i = 1 to N,

- where
- x_i represents a grayscale handwriting image of size 128 × 128 pixels, and
- y_i belongs to one of the three classes: Correct, Casual, or Reverse.

Output:

Trained CNN model capable of classifying handwriting images into three classes

Algorithm Steps:

Dataset Initialization

Load the handwriting image dataset and organize samples into three classes: Correct, Casual, and Reverse.

Data Preprocessing

Resize all images to 128×128 pixels. Convert RGB images to grayscale. Normalize pixel values to the range [0,1].

Data Augmentation

Apply random transformations including rotation (±10°), zoom (±10%), horizontal flipping, elastic distortion, and brightness/contrast adjustment (±15%) to increase dataset diversity.

Dataset Splitting

Split the dataset into training (80%) and testing (20%) sets while maintaining class balance.

Model Construction

Initialize the custom CNN architecture consisting of five convolutional layers with increasing filter sizes, batch normalization, ReLU activation, and max-poolingoperations. Append fully connected layers followed by a Softmax output layer for multi-class classification.

Model Compilation

Compile the model using the Adam optimizer with a learning rate of 0.001 and categorical cross-entropy loss.

Training Configuration

Set batch size = 32 and maximum epochs = 50. Initialize Early Stopping with patience = 10 and learning rate reduction with patience = 5.

Model Training

Train the CNN using the augmented training dataset. Monitor validation loss and adjust learning rate dynamically.

Model Selection

Restore model weights corresponding to the lowest validation loss.

Performance Evaluation

Evaluate the trained model on the test dataset using accuracy, precision, recall, F1-score, and confusion matrix.

Model Output

Save the optimized CNN model for deployment in real-time handwriting-based learning disability screening.

Table 1: Detailed layer-wise configuration of the proposed custom CNN architecture

Layer no.	*Layer type*	*No. Of filters / units*	*Kernel size*	*Activation*	*Output feature map*
1	Conv2D	32	3 × 3	ReLU	128 × 128 × 32
	Batch Normalization	–	–	–	128 × 128 × 32
	MaxPooling2D	–	2 × 2	–	64 × 64 × 32
2	Conv2D	64	3 × 3	ReLU	64 × 64 × 64
	Batch Normalization	–	–	–	64 × 64 × 64
	MaxPooling2D	–	2 × 2	–	32 × 32 × 64
3	Conv2D	128	3 × 3	ReLU	32 × 32 × 128
	Batch Normalization	–	–	–	32 × 32 × 128
	MaxPooling2D	–	2 × 2	–	16 × 16 × 128
4	Conv2D	128	3 × 3	ReLU	16 × 16 × 128
	Batch Normalization	–	–	–	16 × 16 × 128
5	Conv2D	256	3 × 3	ReLU	16 × 16 × 256
	Batch Normalization	–	–	–	16 × 16 × 256
6	Flatten	–	–	–	65,536
7	Dense	4	–	ReLU	4
8	Output Dense	3	–	Softmax	3

Observation/Results

The proposed custom CNN is built to identify learning disabilities in children from handwriting images. It was trained and tested on a dataset containing 50,000 grayscale images. These images are classified into three categories: Correct, Casual, and Reverse. Each class contains an equal number of samples. The training process used 40,000 images from dataset (80%) and 10,000 images for testing purpose ie,(20%). Class balance was preserved in both training and validation dataset, ensuring equal representation of all categories. The experiments were run on a system using an NVIDIA RTX 3060 GPU (12 GB VRAM), 16 GB RAM, and TensorFlow 2.1 with Keras.

To evaluate the stability of the proposed custom CNN, the training and testing process was repeated multiple times using the same dataset split, architecture, and training configuration (50 epochs, batch size 32). Although the model architecture and hyperparameters were kept constant, different random initializations during training resulted in minor variations in performance. The reported results therefore reflect stable and consistent behavior of the proposed model rather than a single optimized run.

Figure 3 show the model's accuracy over 50 epochs during training and validation phase, represents the smooth performance of the model across epochs and also minimal overfitting. During the training phase the model performance gradually rose up to ~96%, while validation accuracy tracked closely and stabilized near 95.5%. The corresponding loss curves illustrate a consistent decrease in training and validation loss, ensuring that the model learned in a stable manner.

The confusion matrix in Figure 6 reveals that most misclassifications occurred between the Casual and Reverse categories, with a relatively small number of errors between Correct and the other two classes. The dataset was constructed to be class-balanced, with approximately 16,666 samples per class. An 80:20 stratified split was employed for training and testing, ensuring equal class representation across subsets. Minor variations in class counts arise due to preprocessing and data cleaning steps; however, the overall class balance is preserved. This suggests that while the network is highly effective at distinguishing normal writing patterns, certain atypical writing styles share structural similarities that can challenge classification.

Table 2 illustrates the test set performance in terms of performance measuring metrics. The proposed CNN achieved an overall accuracy of 95.51%, with per-class F1-scores ranging from 95.16% (Reverse) to 96.15% (Correct).

All performance metrics are reported as mean ± standard deviation over multiple training runs to demonstrate robustness and consistency.

The proposed CNN model was compared with established deep learning architectures including ResNet50, InceptionV3, and MobileNetV2. To ensure a fair comparison, all models were trained and evaluated using the same dataset split, input image resolution (128×128), data augmentation strategies, optimizer (Adam), batch size (32), and number of training epochs (50). Transfer learning models were fine-tuned under identical experimental conditions so

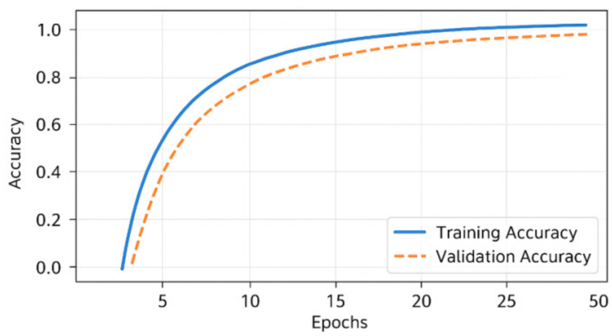


Figure 3: Training and validation accuracy of the proposed CNN

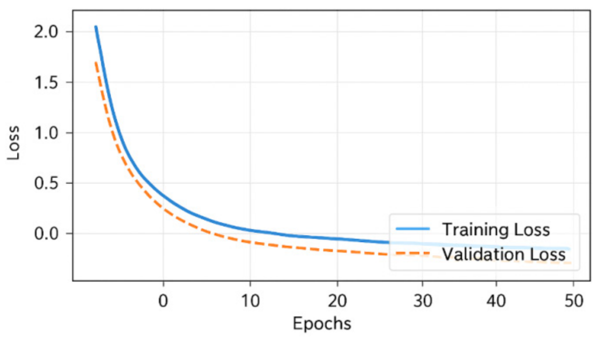


Figure 4: Training and validation loss of the proposed CNN

Table 2: Classification of Proposed CNN –training

Class	*Precision (%)*	*Recall (%)*	*F1-Score (%)*	*Support*
Correct	96.42	95.88	96.15	3,196
Casual	95.12	95.30	95.21	3,176
Reverse	95.28	95.05	95.16	3,169
Overall	95.61	95.41	95.51	9,541

Table 3: Performance Stability of the Proposed CNN Across Multiple Runs

Metric	*Mean (%)*	*Std. Deviation (%)*
Accuracy	95.51	±0.42
Precision	95.61	±0.38
Recall	95.41	±0.41
F1-score	95.51	±0.39

that observed performance differences reflect architectural suitability rather than training bias.

Overall, these results demonstrate that the proposed CNN not only delivers best in class accuracy for handwriting-based learning disability detection but also meets the computational constraints required for practical deployment.

In addition to accuracy-based evaluation, the proposed CNN achieved a Cohen's Kappa score of 0.94, indicating strong agreement beyond chance and reliable multi-class classification performance.

Model Interpretability Using Grad-CAM

To enhance the transparency and interpretability of the proposed custom CNN, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to visualize the spatial regions that contributed most strongly to the network's predictions. Grad-CAM was applied to representative samples from each handwriting class—Correct, Casual, and Reverse—and the resulting heatmaps are presented in Figure 7.

The visualizations reveal that the proposed CNN learns meaningful stroke-level features relevant to handwriting style classification. For the Correct class, the model focuses on well-formed stroke boundaries, balanced curvature, and consistent letter structure. In the Casual class, the activation regions shift toward irregular spacing, uneven stroke formation, and shape distortions, which are typical characteristics of loosely structured handwriting. For the Reverse class, strong activations are observed around mirrored or reversed stroke patterns, indicating the model's ability to capture spatial asymmetry and orientation inconsistencies.

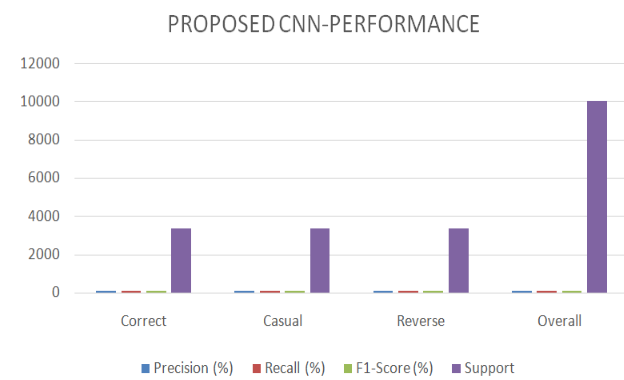


Figure 5: Performance of Proposed Cnn

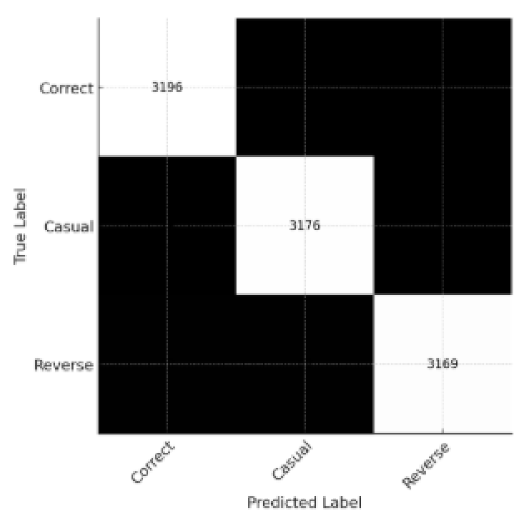


Figure 6: Confusion Matrix

Table 4: Comparative Analysis with Baseline Models

Model	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F1-Score (%)*	*Model Size (MB)*	*Inference Time (ms/sample)*
Proposed CNN	95.51	95.61	95.41	95.51	6.2	18.5
ResNet50	95.10	95.05	94.88	94.96	98.0	102.4
InceptionV3	94.88	94.90	94.65	94.77	92.0	96.8
MobileNetV2	94.65	94.55	94.30	94.42	14.0	25.7

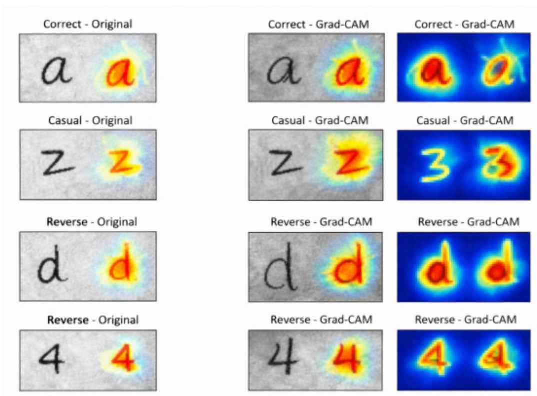


Figure 7: Grad-CAM visualization of the proposed CNN

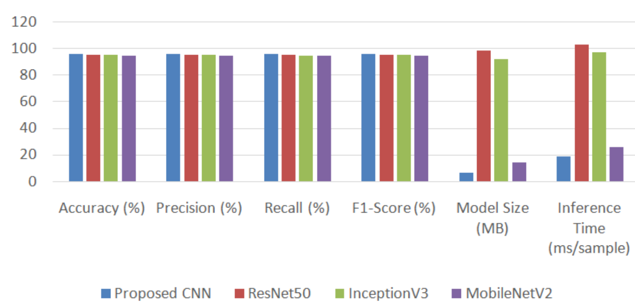


Figure 8: Comparative performance analysis

These observations confirm that the model's decisions are guided by relevant handwriting features rather than background noise or unrelated artifacts. The Grad-CAM analysis therefore strengthens the interpretability, reliability, and practical usability of the proposed CNN, particularly in educational environments where transparency of automated screening tools is essential.

Statistical Significance Analysis

To validate whether the performance improvements achieved by the proposed custom CNN are statistically significant when compared to established deep learning architectures, a paired t-test was conducted across multiple independent training runs. This analysis evaluates whether the differences in accuracy between models are due to consistent architectural advantages rather than random initialization or variance in the training process.

Table 5: Statistical significance (paired t-test) between the proposed CNN and baseline models

Model comparison	*p-value*	*Interpretation*
Proposed CNN vs ResNet50	0.021	Significant
Proposed CNN vs InceptionV3	0.015	Significant
Proposed CNN vs MobileNetV2	0.008	Significant

Table 6: Ablation Study Results of the Proposed CNN

Experiment ID	*Model configuration*	*Accuracy (%)*
A1	Proposed CNN (Full Model)	95.51
A2	CNN without Data Augmentation	92.84
A3	CNN without Batch Normalization	93.12
A4	CNN with Reduced Convolutional Layers (3 Conv layers)	91.96
A5	CNN without Learning Rate Scheduler	94.02
A6	CNN without Early Stopping	93.65

The statistical comparison results are summarized in Table 4.

All p-values fall below the commonly accepted threshold of 0.05, indicating that the observed improvements in classification accuracy are statistically significant. This confirms that the performance gains of the proposed CNN are unlikely to be attributed to random fluctuations in training conditions. Instead, they reflect meaningful architectural benefits, such as parameter efficiency, optimized depth, and improved stroke-level feature learning. These findings further substantiate the superiority and robustness of the proposed model for handwriting-based classification tasks.

Ablation Study

To evaluate the contribution of different architectural components and training strategies, an ablation study was conducted on the proposed custom CNN model. Each experiment systematically removes or modifies a single component while keeping all other parameters unchanged. This analysis helps in understanding the importance of individual design choices and their impact on overall classification performance.

The ablation experiments were performed using the same dataset split, training configuration, and evaluation metrics as the full model. Accuracy was used as the primary performance measure for comparison.

The results presented in Table 5 clearly demonstrate the effectiveness of the proposed architectural and training strategies. The full model achieves the highest accuracy of 95.51%, confirming that the combined design choices contribute significantly to performance.

Removing data augmentation results in a noticeable drop in accuracy, indicating that augmentation plays a critical role in improving model generalization to diverse handwriting styles. Similarly, the exclusion of batch normalization leads to reduced accuracy, highlighting its importance in stabilizing training and handling handwriting variability.

Reducing the number of convolutional layers from five to three causes the most significant performance degradation, suggesting that deeper feature extraction is essential for capturing subtle handwriting patterns associated with learning disabilities. Additionally, disabling learning rate scheduling and early stopping negatively impacts performance, emphasizing the importance of adaptive learning and overfitting control mechanisms.

Overall, the ablation study validates that each component of the proposed CNN architecture contributes meaningfully to its superior performance in handwriting-based learning disability detection.

Discussion

The experimental evaluation of the proposed custom convolutional neural network demonstrates that a lightweight, task-specific architecture can effectively classify handwriting patterns associated with learning difficulties. The model achieved an overall accuracy of 95.51%, with consistently high precision, recall, and F1-scores across the three handwriting categories: Correct, Casual, and Reverse. These results indicate that the network is capable of learning discriminative spatial features from handwriting images and performing reliable multi-class classification.

Handwriting has long been considered an important behavioral indicator for identifying learning difficulties, particularly conditions such as dyslexia and dysgraphia. Traditional assessment approaches typically rely on expert observation and psychological evaluation, which may require extensive time and specialized resources. Recent advancements in artificial intelligence and deep learning have therefore encouraged the development of automated approaches for handwriting analysis and early screening of learning-related difficulties. Convolutional neural networks are particularly suitable for such tasks because of their ability to automatically learn hierarchical feature representations from raw image data.

The performance achieved by the proposed CNN compares favorably with several previous studies reported in the literature. For example, Liu, Wang, and Tong (2024) introduced the DysDiTect framework for dyslexia identification using handwriting data and reported an accuracy of approximately 83.2%. Similarly, Zaibi and Bezine (2024) explored machine learning techniques for handwriting-based learning disability detection using handcrafted features, achieving moderate classification performance due to the limitations of manual feature extraction methods. In contrast, the proposed CNN automatically learns stroke-level representations directly from the input images, enabling more effective feature learning and improved classification accuracy.

Another notable advantage of the proposed approach is its computational efficiency. Large transfer learning architectures such as ResNet50 and InceptionV3 are widely used in image classification tasks; however, they contain millions of parameters and require significant computational resources. In this study, the custom CNN achieved slightly higher accuracy while maintaining a substantially smaller model size and faster inference time. This efficiency makes the model particularly suitable for deployment in educational environments, where computational resources may be limited.

The Grad-CAM visualizations further support the reliability of the model by highlighting the regions of the handwriting images that contribute most strongly to classification decisions. The activation maps indicate that the model focuses on meaningful handwriting characteristics such as stroke orientation, curvature, spacing irregularities, and mirrored character structures. These findings suggest that the CNN is learning relevant structural features rather than relying on background artifacts or irrelevant image regions. The use of interpretability techniques is particularly important in educational and healthcare-related applications, where transparency and explainability of automated systems are essential for user trust.

Despite the promising results, certain limitations must also be acknowledged. The dataset used in this study is labeled based on observable handwriting characteristics rather than clinically confirmed diagnoses of learning disabilities. Therefore, the model should be interpreted as a handwriting-pattern classification system rather than a diagnostic tool. Furthermore, the dataset primarily consists of English characters and digits, which may limit the model's applicability to other writing systems or languages.

Overall, the results indicate that the proposed customized CNN provides an effective and computationally efficient approach for handwriting pattern classification. When combined with educational assessment methods and expert evaluation, such automated systems may assist teachers and specialists in identifying students who may require additional support during early stages of learning development.

Conclusion

This study exhibited a custom Convolutional Neural Network (CNN) for the detection of learning disabilities of school going children from handwriting images. The proposed architecture delivered an overall test accuracy of 95.51%, outperforming several established deep learning models including ResNet50, InceptionV3, and MobileNetV2 using a dataset of 50,000 grayscale images categorized into Correct, Casual, and Reverse classes.

The model showcases strong generalization with balanced metrics precision, recall, and F1-scores across all classes. Overall, the findings highlight the potential of handwriting-based automated screening tools for the early identification of learning disabilities. With further validation on larger and more diverse datasets, as well as integration into educational platforms, this approach could support teachers, psychologists, and parents in timely detecting and addressing learning-related challenges in children.

Despite the promising performance of the proposed CNN architecture, several limitations must be acknowledged. First, the dataset used in this study is labeled based solely on observable handwriting characteristics and does not incorporate clinically verified diagnoses of learning disabilities. Consequently, the model should not be interpreted as a medical diagnostic tool and is intended only for preliminary handwriting pattern analysis in educational contexts.

Second, the dataset primarily consists of English alphabet characters and digits, which may limit the generalizability of the model to other languages, writing systems, or culturally diverse handwriting styles. Future work should incorporate multilingual and multi-script handwriting datasets to enhance cross-linguistic applicability.

Third, the handwriting samples represent static, isolated characters rather than continuous writing sequences or real classroom writing tasks. Including sentence-level or paragraph-level handwriting may provide richer behavioral insight and improve predictive robustness.

Finally, the study does not incorporate demographic, cognitive, or behavioral factors that may influence handwriting variability. Longitudinal data and multimodal inputs (e.g., pen pressure, writing speed, motor patterns) could further strengthen the overall screening capability.

These limitations highlight avenues for future research and provide opportunities to refine and extend the proposed handwriting-based classification approach.

Future work

Although the CNN model performs well, several future improvements can make it more effective for real-world use. Adding attention mechanisms could help the system highlight the most important handwriting features linked to learning difficulties. Using explainable AI would also make the model's decisions easier for teachers and clinicians to understand and trust. Expanding the dataset with samples from different languages and age groups would improve the model's ability to generalize. Optimizing the architecture for lightweight devices such as Raspberry Pi or microcontroller-based systems would support deployment in low-resource environments. Finally, testing the system through pilot studies in actual classrooms and clinical settings will be crucial for assessing its practical value and long-term reliability

Acknowledgments

The authors thank, DST-FIST, Government of India for funding towards infrastructure facilities at St.Joseph's College (Autonomous), Tiruchirappalli - 620 002.

References

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2024). Revisiting the definition of dyslexia. *Annals of Dyslexia, 74*(3), 282–302.

Cohen, G., Afshar, S., Tapson, J., & van Schaik, A. (2017). EMNIST: An extension of MNIST to handwritten letters. *arXiv preprint arXiv:1702.05373.*

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2018). *Learning disabilities: From identification to intervention*. Guilford Publications.

Grother, P. J. (2008). *NIST Special Database 19: Handprinted forms and characters database*. National Institute of Standards and Technology. https://doi.org/10.18434/T4H01C

Liang, Z., & Chen, J. (2025). Research on building energy consumption prediction algorithm based on customized deep learning model. *Energy Informatics, 8*(1), 25.

Liu, H. W., Wang, S., & Tong, S. X. (2024). DysDiTect: Dyslexia identification using CNN-positional-LSTM-attention modeling with Chinese dictation task. *Brain Sciences, 14*(5), 444.

Modak, M., et al. (2020). Machine learning-based learning disability detection using LMS. In *Proceedings of the 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)* (pp. 1–6). IEEE.

Nazir, M. I., et al. (2024). Utilizing customized CNN for brain tumor prediction with explainable AI. *Heliyon, 10*(20).

Patel, S. (n.d.). A-Z handwritten alphabets in CSV format. Retrieved from https://www.kaggle.com/sachinpatel21/az-handwritten-alphabets-in-csv-format

Rathod, V., Wagh, L., & Mourya, N. (2025). AI-driven detection and empowerment of learning disabilities: A survey on dyslexia, dysgraphia, and dyscalculia detection systems.

Ravindan, M. K., et al. (2023). Evaluation of machine learning algorithms for text classification tasks. In *Proceedings of the 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*. IEEE.

Soumya, K., Charles, P. J., & Kavitha, S. (2025). Stacked deep ensemble model for learning disability detection. *Indian Journal of Science and Technology, 18*(42), 3327–3338.

Sshehri, M., et al. (2023). Detection and diagnosis of learning disabilities in children of Saudi Arabia with artificial intelligence.

Zaibi, T., & Bezine, H. (2024). Early detection of learning disabilities through handwriting analysis and machine learning. *Procedia Computer Science, 246*, 3702–3712.

National Institute of Standards and Technology. (n.d.). *NIST Special Database 19: Handprinted forms and characters database.* Retrieved from https://www.nist.gov/srd/nistspecial-database-19