Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

DualAttentionNet: A Lightweight Dual-Attention Framework for Robust Multi-Class EEG Digit Decoding

Md Rashidul Islam
Albukhary International University
Md Shakil Hossain
Manarat International University
Md Saiful Arefin
Albukhary International University
Md Ridwan Ali
Jashore University of Science and Technology
Nick Rahimi
University of Southern Mississippi
Saydul Akbar Murad
smurad@emporia.edu

Emporia State University

Research Article

Keywords: MNIST Classification, Multi-Class Classification, EEG Decoding

Posted Date: September 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10796304/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

DualAttentionNet: A Lightweight Dual-Attention Framework for Robust Multi-Class EEG Digit Decoding

Md Rashidul Islam[1], Md Shakil Hossain[2], Md Saiful Arefin[1], Md Ridwan Ali[3], Nick Rahimi[4], Saydul Akbar Murad[5*]

[1]School of Computing and Informatics, Albukhary International University, Kedah, Malaysia.
[2]Department of Computer Science and Engineering, Manarat International University, Ashulia, Bangladesh.
[3]Department of Electrical and Electronic Engineering, Jashore University of Science and Technology, Khulna, Bangladesh.
[4]School of Computing Sciences and Computer Engineering, University of Southern Mississippi, Mississippi, USA.
[5]School of Business and Technology, Emporia State University, KS, USA.

*Corresponding author(s). E-mail(s): smurad@emporia.edu;

Abstract

The visual classification of a stimulus by electroencephalography (EEG) has been a significant line of investigation in Brain Computer Interface (BCI) systems. The current issue of converting EEG signals into useful cognitive representations, including digit recognition, is a difficult issue because of low signal to noise ratio, high inter trial variability and overlapping neural tracks. This paper presents a DualAttentionNet with a multi-class EEG digit classification system based on Visual MNIST. The offered model improves the traditional EEGNet model by incorporating two channel-attention blocks. The architectural design attains 92.36% classification accuracy, being superior to all the architectures compared and at the same time with 92.36 computational efficiency. Under the analysis of signal variability and explainability, statistical analysis indicates the strength of the model. It is a strong, interpretable and reproducible multi-class EEG-based classification framework. The outcomes emphasise the significance of organised attention systems and streamlined training interventions in enhancing the performance of EEG decoding. The code is available at: https://github.com/mrirashid/EEG-Based-MNIST-Digit-Classification.

Keywords: MNIST Classification, Multi-Class Classification, EEG Decoding.

1 Introduction

Electroencephalography (EEG) is a non-invasive neurophysiological technique that records the electrical activity of the brain through electrodes placed on the scalp [1]. It measures voltage fluctuations generated by the synchronous firing of large populations of cortical neurons and provides millisecond-level temporal resolution [2], making it highly suitable for studying rapid neural processes related to perception, attention, cognition, and motor activity [3]. Due to its portability, relatively low cost, and excellent temporal sensitivity, EEG has become one of the most widely used signal acquisition modalities in neuroscience and brain–computer interface (BCI) research [4]. In practical BCI systems, EEG signals are processed to decode user intentions, perceptual states, and enable human–computer interaction [5].

In EEG-based visual decoding tasks, a visual stimulus is presented to a subject, and the brain generates corresponding electrophysiological responses that are recorded as multichannel time-series signals [6, 7]. Classification models aim to map these complex neural patterns to predefined stimulus categories [8]. This process not only provides insights into how perceptual information is represented in brain dynamics but also supports the development of visual BCI systems [9]. However, transforming raw EEG signals into reliable class predictions remains a challenging task [10].

One of the primary challenges in EEG classification is the inherently low signal-to-noise ratio (SNR), where informative neural responses are often weak compared to background brain activity and recording noise [11]. EEG signals are highly non-stationary, with statistical properties varying across trials, subjects, sessions, and recording conditions [12]. Moreover, recordings are frequently contaminated by artifacts such as eye movements, muscle activity, electrode impedance fluctuations, and environmental interference [13, 14]. In addition, EEG data are high-dimensional, consisting of multiple channels and temporal samples, while discriminative features may only exist in specific temporal windows or subsets of channels [15]. These factors make EEG classification substantially more complex than conventional image or tabular data classification, requiring carefully designed preprocessing, feature extraction, and evaluation strategies.

The availability of publicly accessible EEG datasets has accelerated research progress by enabling systematic benchmarking of different methods. For instance, the MindBigData Visual MNIST dataset reformulates the classical handwritten digit recognition problem in the EEG domain by recording brain activity while subjects observe visual digits [16, 17]. However, variability across subjects, channel noise, and trial inconsistencies can significantly affect model performance. Therefore, this benchmark requires not only accurate models but also robust and reproducible pipelines capable of generalizing under realistic EEG conditions [4].

Traditional EEG classification approaches typically rely on handcrafted features, frequency-domain analysis, dimensionality reduction, or manual channel selection.

More recently, deep learning models have been introduced to automatically learn discriminative features from raw or minimally processed EEG signals [6, 18]. Among these, EEGNet has gained significant attention due to its computational efficiency and design tailored to EEG signal characteristics [3]. Nevertheless, existing methods still face several limitations. Many approaches depend heavily on dataset-specific preprocessing steps, selective channel usage, or partially documented experimental settings, which hinder reproducibility and fair comparison [19, 20]. While some methods improve performance by increasing model complexity, this often leads to higher computational costs and reduced interpretability [21]. Conversely, lightweight models may struggle to effectively suppress noise and emphasize informative spatial features, particularly in fine-grained multi-class EEG classification tasks [18].

These challenges highlight the need for models that are accurate, robust, lightweight, and interpretable. An effective EEG classification framework should capture meaningful temporal and spatial patterns, reduce the influence of noisy or irrelevant channels, and maintain computational efficiency suitable for real-world BCI applications [22]. Additionally, such models should be supported by clearly documented and reproducible training pipelines to ensure reliable evaluation and future extensibility [23]. This requirement becomes even more critical in EEG-based digit classification, where fine-grained discrimination among visually similar classes must be achieved under noisy neural recording conditions [24, 25].

In this paper, we propose *DualAttentionNet*, an attention-enhanced EEG classification framework designed for multi-class digit recognition. The proposed model extends the EEGNet backbone by integrating two Squeeze-and-Excitation (SE) channel-attention modules, enabling adaptive reweighting of feature channels and improved focus on informative neural representations. The main contributions of this study are summarized as follows:

- We propose DualAttentionNet, a lightweight EEG-based digit classification model that enhances the EEGNet architecture with dual SE channel-attention modules for improved feature refinement.
- We develop a structured and reproducible EEG classification pipeline incorporating band-pass filtering, channel-wise normalization, data augmentation, Mixup, OneCycle learning rate scheduling, gradient clipping, and test-time augmentation to improve robustness and generalization.
- We provide explainability analyses using temporal saliency mapping, channel-attention visualization, and input gradient magnitude heatmaps to better understand the neural patterns utilized by the model.

The remainder of this paper is organized as follows. Section 2 presents the literature review. Section 3 describes the proposed methodology, including dataset details, preprocessing, model architecture, training strategy, and evaluation framework. Section 4 discusses the experimental results. Finally, Section 5 concludes the paper and outlines future research directions.

2 Literature Review

In recent years, a significant amount of research has been done on the preprocessing and classification of EEG signals. It has also been demonstrated that proper preprocessing and channel selection contribute positively to the classification performance with the use of the MindBigData dataset. The number of studies have shown that such strategy leads to better results, regardless of the methodology employed to utilize the datasets and the type of task performed on them [8, 26–28]. Other papers have considered architectural innovation and the model's representation strategy to improve the generalization and accuracy of models used in architectural studies [5, 29–31]. Nonetheless, some small datasets and simplistic experimental conditions have led to limited performance in most EEG digit classification studies, casting doubt on their extrapolation to real-world BCI applications.

Mishra et al. proposed a pipeline using a convolutional neural network (CNN) for visual recognition of stimuli from short-duration EEG measurements, achieving 70% accuracy with spatially selective channels [26]. Tsamourgelis and Adamopoulos [27] concentrated on the distinction of simple visual stimuli and when it comes to digit/non-digit discrimination, the results are high precision, yet, fine-grained Discrimination of digits was not easy. Tanner and Cutsuridis [28] used intensive filtering, artifact rejection and the selective use of channels with an accuracy of 92 percent, and subsequently, expanded on their efforts with the AC-GAN models in the EEG to image reconstruction. To achieve a high accuracy rate, (95.4) and high quality image reconstruction that is based on the EEG, Khaleghi et al. refined CNN-GAN architecture and managed to improve performance [8].

Table 1 summarizes representative studies in EEG-based visual and digit decoding and compares them with the proposed DualAttentionNet framework.

In parallel with these developments, there are other studies, which have explored the architectural diversity and feature enrichment strategies. Singh et al. [5] proposed a hybrid CNN-RNN model to simultaneously learn the spatial and temporal EEG properties. Pratama et al. classic frequency-band and PCA-based feature extraction and multi-digit classification tasks showed low accuracy rates [29]. Chowdhury et al. [30] introduced the SSL-SE-EEG framework that integrates self-supervised learning and squeeze-and-excitation modules to enhance the resistance to noise and the imbalance of classes. Falciglia et al. paid much attention to the multi-domain feature fusion, introducing spatial, spectral, and temporal features into one representation, and this scheme performed better than some of the single-domain systems [31]. They achieved 86% accuracy.

Our research proposes a DualAttentionNet multi-class EEG digit decoder. The proposed method improves the model robustness by combining dual channel-attention (SE) modules and optimised training schemes in classification than the traditional methods based on the EEGNet and yet retains computational efficiency and reproducibility.

Table 1 Comparison of recent EEG-based visual and digit decoding approaches with the proposed method.

Study	Dataset	Channels	Model	ML	DL	Acc. (%)	Other metric
[26]	MNIST of Brain Digits MindWave	4	CNN	No	Yes	70	–
[28]	MNIST of Brain Digits EPOC	4	CNN, AC-GAN	No	Yes	92	SSIM (29%)
[30]	BigMindData Muse2-v0.17 + TUH-AB	21	SSL, SE-CNN	No	Yes	91.12	F1-score (85.18%)
[29]	MNIST of Brain Digits	14	PCA, KNN	Yes	No	31	–
[5]	MNIST of Brain Digits EPOC	14	CNN, RNN	No	Yes	73	–
[31]	MindBigData (VisMNIST Cap64)	64	CNN manifold	No	Yes	86	–
[8]	BigMindData Muse2-v0.17	14	CNN, GAN	No	Yes	95.4	SSIM (92.9%)
[27]	BigMindData Muse2-v0.17	14	Multiple classifiers	No	Yes	86	–
Proposed method	**MindBigData (VisMNIST Cap64)**	**16**	**DualAttentionNet**	No	Yes	**92.36**	ROC–AUC (99%)

3 Methodology

This section explains our full pipeline for MNIST digit classification using DualAttentionNet [17]. We design a structured and reproducible workflow that transforms raw EEG signals into reliable digit predictions to address these challenges. The complete pipeline is illustrated in Figure 1.

3.1 Dataset Overview

In this study, we employ the CustomCap64Morlet-v0.016 dataset [17], which is the latest version of the MindBigData Visual MNIST (VisMNIST_Cap64) database, updated on 27 December 2022 [17]. The dataset was collected using a home-built Custom Cap 64-channel EEG acquisition device while subjects were visually observing handwritten digits from the well-known Yann LeCun MNIST dataset. The detailed specifications of the recording device and dataset configuration are summarised in Table 2.

The dataset contains 1,166 digit stimulus events and 1,166 black screen events, resulting in a balanced distribution between stimulus and non-stimulus conditions. Furthermore, the digit classes (0–9) are presented. The number of EEG recordings per digit class (0–9) and black screen condition, demonstrating a balanced dataset suitable for supervised learning without strong class imbalance bias.

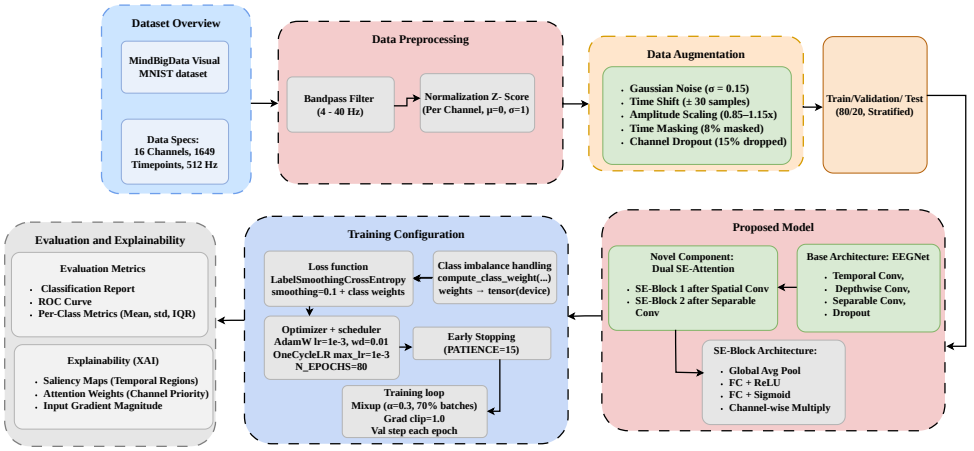


Fig. 1 Overview of the proposed DualAttentionNet workflow for EEG-based MNIST digit classification.

Table 2 Technical specifications of the 64-channel EEG recording system and dataset configuration.

Parameter	**Value**
Number of channels	64
Samples per channel	400
Total samples per recording	25,600 (64×400)
Recording duration	2 s
Sampling frequency	200 Hz
Number of EEG recordings	148,736
Channel order	FP1, FPz, FP2, AF3, AFz, AF4, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO5, PO3, POz, PO4, PO6, PO8, CB1, O1, Oz, O2, CB2

3.2 Data Preprocessing

EEG signals are often contaminated by noise, baseline drift, and channel-wise amplitude variations. If these effects are not addressed, deep learning models may learn irrelevant patterns instead of meaningful neural activity. Therefore, a preprocessing pipeline is applied before training the proposed model. The pipeline consists of two main steps: (1) band-pass filtering and (2) channel-wise normalization.

Step 1: Band-pass Filtering. First, the raw EEG signals are filtered using a fourth-order Butterworth band-pass filter in the range of 4–40 Hz. This frequency range preserves important cognitive EEG components while removing slow baseline

drift (below 4 Hz) and high-frequency noise such as muscle artifacts (above 40 Hz). Zero-phase forward–backward filtering is applied to avoid phase distortion and preserve the temporal structure of the signal. Figure 2 shows the filtering suppresses unwanted frequency components and produces a cleaner and smoother EEG signal.

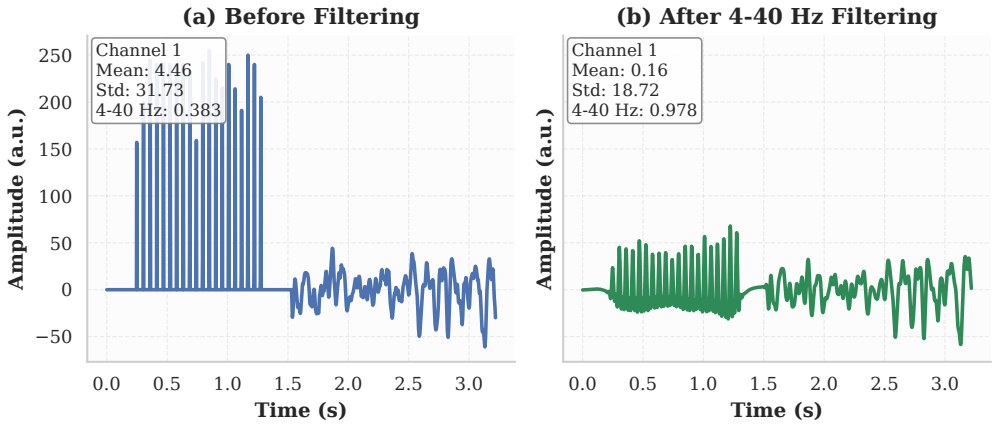


Fig. 2 Illustrating before and after band-pass filtering of one channel

Step 2: Channel-wise Normalization. Channel-wise Z-score normalization is the next step after the filtering to decrease the amplitude variability between channels and trials. The signal of every channel is normalized, i.e. the mean of the channel is removed and then divided by the standard deviation of the channel, such that the mean becomes zero and the standard deviation equals one. The reason behind this step is to keep the input on a similar scale, enhance numerical stability in training, and allow channels with high amplitude to dominate feature learning. The normalized EEG signal is presented in Figure 3. The preprocessing stage involves a representation of each EEG trial in the form of a matrix.

The size of the input is as follows: fewer than 64 EEG channels in a row, fewer than 400 samples per second in a column, which is why the input is in the form of a table-like structure, i.e. more details. It offers consistent and constant inputs to the proposed EEGNet-based model due to this standardized representation.

3.3 Data Augmentation

EEG signals remain noisy and highly variable due to physiological differences and recording conditions, despite the VisMNIST dataset being balanced. Several data augmentation techniques are applied during training to improve model generalisation and reduce overfitting. These augmentations simulate realistic variations in EEG recordings by preserving the original class labels. Validation and test data remain unchanged but augmentation is applied only to the training set. Figure 4 illustrates the EEG signals before and after each augmentation technique.

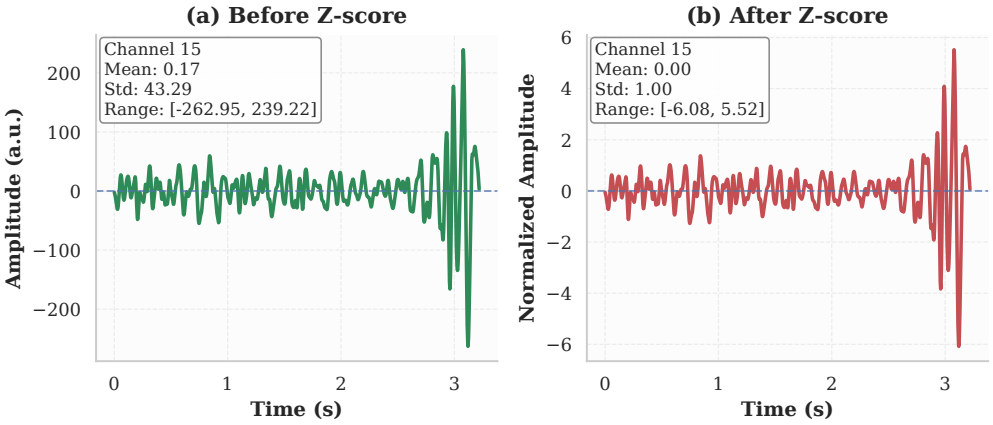


Fig. 3 Illustrating before and after channel-wise normalization of one channel

Step 1: Gaussian Noise Injection. Small Gaussian noise is added to the signal to simulate measurement disturbances. This encourages the model to learn stable spatial–temporal patterns rather than relying on minor amplitude fluctuations.

Step 2: Random Time Shifting. The EEG signal has been shifted along the time axis to replicate the temporal mismatch across the trials. It enhances the capacity of this model to identify various patterns of time positions that take place at slightly.

Step 3: Amplitude Scaling. Signal amplitude is also randomized to compensate for the difference due to electrode impedance or subject variations. This helps the model concentrate on signal form rather than absolute intensity.

Step 4: Time Masking. There is a random masking (setting to zero) of a short portion of the EEG signal. This causes the model to pick up distributed features of time and not a single peak or time window.

Step 5: Channel Dropout. Training randomly suppresses some channels of an EEG to represent an electrode failure or a localized noise. This promotes the acquisition of strong spatial representations within a number of channels in this model.

These augmentation strategies introduce variability of EEG signals across amplitude, temporal, and spatial dimensions. The network learns more invariant features and achieves improved generalization by exposing the model to such perturbations during training.

3.4 Proposed Model

We propose a DualAttentionNet for EEG-based MNIST digit classification. EEG signals are often noisy and non-stationary, so not all learned feature channels contribute equally to classification. Our proposed model extends the lightweight EEGNet backbone by integrating two Squeeze-and-Excitation (SE) attention blocks to improve feature selection. These blocks dynamically emphasize informative feature channels while suppressing less relevant ones, leading to more robust EEG representations. Figure 5 illustrates the overall architecture of the proposed model.

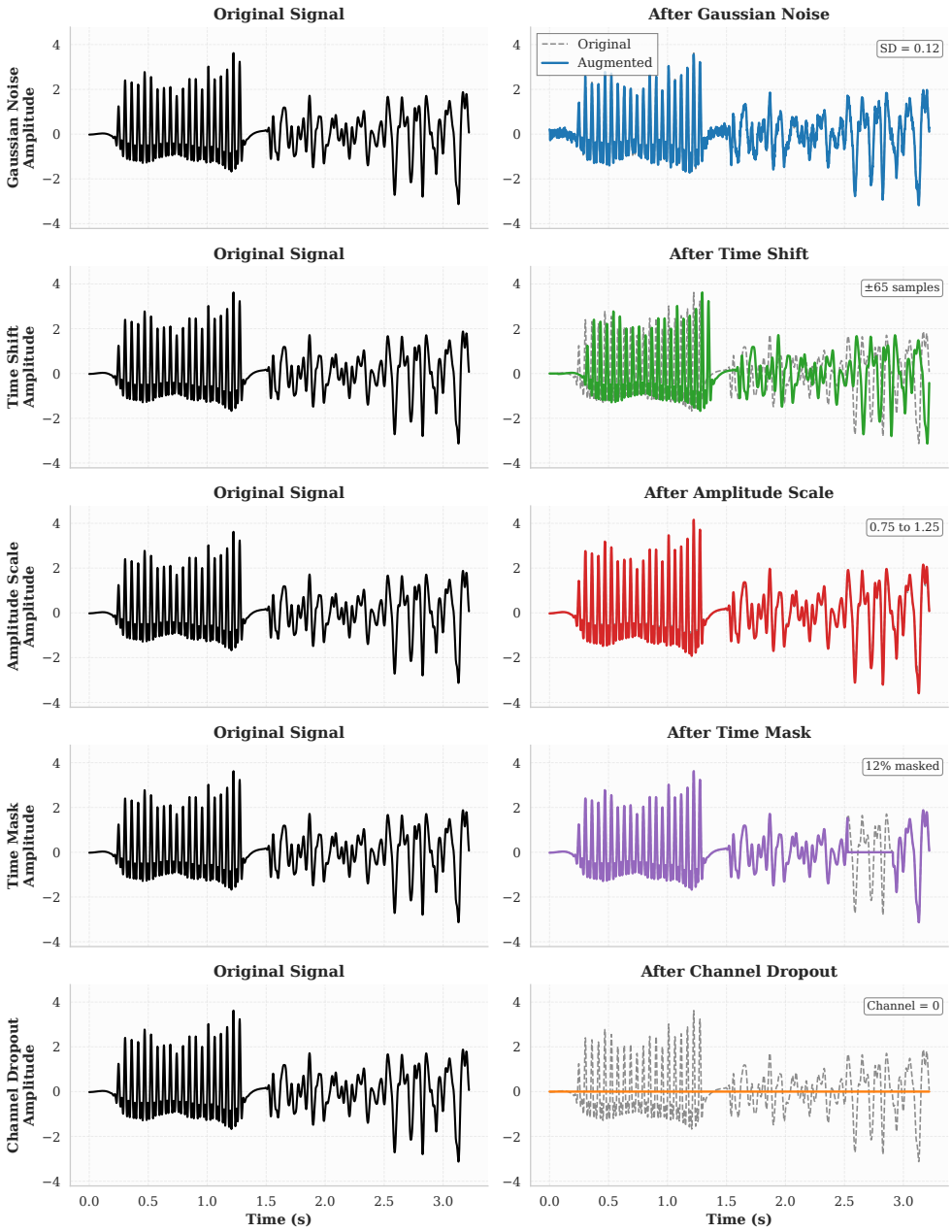


Fig. 4 EEG signals before and after applying different data augmentation techniques, including Gaussian noise injection, time shifting, amplitude scaling, time masking, and channel dropout.

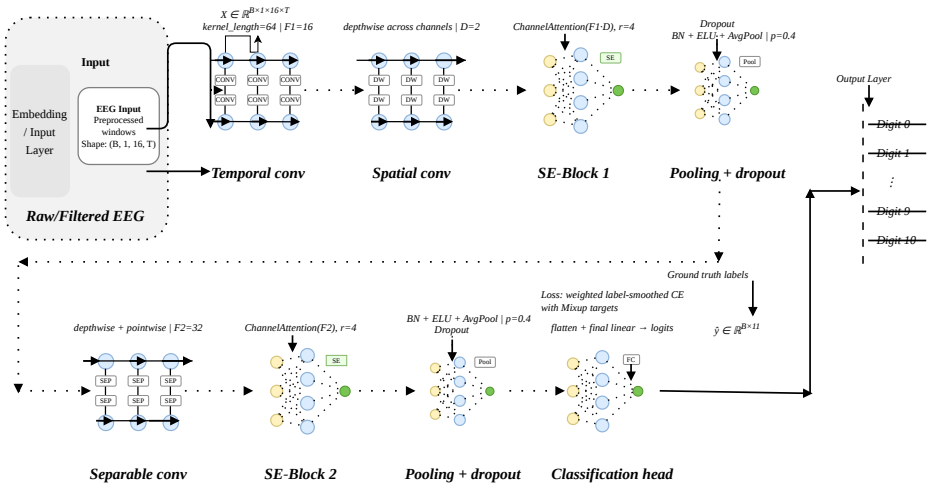


Fig. 5 Proposed Dual-Attention EEGNet architecture combining the EEGNet backbone with two SE-based channel attention modules.

3.4.1 Basic Architecture EEGNet

The model follows the EEGNet backbone for efficient EEG feature learning. Temporal convolution extracts time-dependent patterns from the EEG signals. Depthwise spatial convolution learns spatial relationships across EEG electrodes. After pooling and dropout, separable convolution captures higher-level EEG features while keeping the number of parameters small. Finally, the extracted features are flattened and passed to a linear classifier that predicts 11 classes (Digit 0–9 and Black Screen).

3.4.2 Dual SE-Attention

Two-channel attention modules are inserted into the EEGNet backbone to improve feature quality. The first SE block is placed after the spatial convolution stage to refine early spatial feature maps. The second SE block is applied after the separable convolution stage to enhance deeper and more discriminative representations. This dual-attention strategy allows the network to progressively refine channel importance at different depths.

3.4.3 SE-Block Architecture

Each SE block performs channel-wise attention through three operations: squeeze, excitation, and re-weighting. Global average pooling first summarizes each feature channel into a compact descriptor. A small gating network then learns channel importance weights, which are used to rescale the feature maps. As a result, informative EEG channels are amplified while less useful channels are suppressed, improving feature discrimination and model robustness.

3.5 Training Configuration

The proposed model is trained using a structured optimization strategy designed to ensure stable convergence and strong generalization for EEG signal classification. The complete configuration is summarized in Table 3. Class weights are computed using a balanced formulation to mitigate small class distribution differences in the VisMNIST_Cap64 dataset.

$$w_c = \frac{N}{K \cdot n_c} \tag{1}$$

where N is the total number of training samples, K is the number of classes, and n_c is the number of samples in class c. Model parameters are optimized using the AdamW optimizer, which decouples weight decay from gradient updates. The parameter update rule is

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t} + \epsilon} - \eta \lambda \boldsymbol{\theta}_t \tag{2}$$

A base learning rate of 5×10^{-4} with weight decay $\lambda = 0.02$ is used, together with a OneCycle learning rate schedule that increases the learning rate early in training and gradually decreases it through cosine annealing. To improve generalization and encourage smoother decision boundaries, mixup augmentation is applied. For two samples $(\mathbf{x}_i, y_i)$ and $(\mathbf{x}_j, y_j)$, synthetic examples are generated as

$$\tilde{\mathbf{x}} = \lambda \mathbf{x}_i + (1 - \lambda)\mathbf{x}_j \tag{3}$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j \tag{4}$$

where $\lambda \sim \text{Beta}(\alpha, \alpha)$ and $\alpha = 0.3$. The model is trained for up to 80 epochs using batch size 64. Early stopping with a patience of 15 epochs is employed to prevent overfitting, and the model with the best validation accuracy is selected for final evaluation.

3.6 Evaluation and Explainability

The DualAttentionNet model is evaluated on the assessment of generalisation, class-wise behavior, and interpretability. Test-Time Augmentation (TTA) and multiple explainability techniques are employed to enhance robustness and interpretability.

3.6.1 Evaluation Metrics

During inference, TTA generates $N = 7$ slightly perturbed versions of each test sample (Gaussian noise, mild amplitude scaling). Predictions are averaged to improve stability. Model performance is evaluated using overall accuracy:

$$\text{Accuracy} = \frac{\sum_{i=1}^{M} \mathbf{1}(\hat{y}_i = y_i)}{M} \times 100\% \tag{5}$$

where M is the number of test samples. Confusion matrices are computed to analyse per-class performance:

Table 3 Training configuration of the proposed DualAttentionNet.

Parameter	**Value**
Optimizer	AdamW
Base learning rate	5×10^{-4}
Maximum learning rate (OneCycleLR)	2×10^{-3}
Weight decay	0.02
Learning-rate schedule	OneCycleLR with cosine annealing; `pct_start`=0.2; updated after each mini-batch
Loss function	Label-smoothed cross-entropy with class weighting
Label-smoothing factor	0.1
Class weighting	Balanced formulation, $w_c = \frac{N}{Kn_c}$
Batch size	64
Maximum number of epochs	80
Early-stopping patience	15 epochs
Gradient clipping (maximum norm)	$\|\nabla_{\boldsymbol{\theta}} \mathcal{L}\|_2 \leq 1.0$
Data-augmentation probability	Warm-up from 0.25 to 0.60 during the first 12 epochs, followed by 0.60
Mixup	Enabled ($\alpha = 0.3$), with gradually increasing application probability
Augmentation operations	Gaussian noise, random time shifting, amplitude scaling, time masking, and channel dropout
Proposed-model test protocol	Standard inference with test-time augmentation (TTA)
Baseline test protocol	Standard inference without TTA

$$C_{i,j} = |\{n \mid y_n = i \text{ and } \hat{y}_n = j\}| \tag{6}$$

Precision, recall, and F1-score for each class c are calculated as

$$\begin{aligned} \text{Precision}_c = \frac{TP_c}{TP_c + FP_c}, \quad \text{Recall}_c = \frac{TP_c}{TP_c + FN_c}, \\ F1_c = \frac{2 \cdot \text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \end{aligned} \tag{7}$$

Additionally, descriptive statistics (mean, SD, median, IQR) of predicted probabilities are computed to assess calibration and prediction confidence.

3.6.2 Explainability (XAI)

Three techniques are applied for interpretability: gradient-based saliency mapping and channel-attention visualization and input gradient magnitude heatmpas.

Gradient-based Saliency Maps

For an input $\mathbf{x}$ and predicted class c, the saliency map is

$$\mathbf{S} = \left| \frac{\partial f_c(\mathbf{x})}{\partial \mathbf{x}} \right| \tag{8}$$

Averaging gradients across channels produces temporal importance curves, highlighting time intervals that strongly influence classification.

Channel-Attention Visualization

Channel-attention weights $\mathbf{a} \in \mathbb{R}^C$ are extracted via forward hooks, and averaged across test samples:

$$\bar{\mathbf{a}} = \frac{1}{M} \sum_{i=1}^{M} \mathbf{a}^{(i)} \tag{9}$$

Higher $\bar{\mathbf{a}}$ values indicate channels consistently contributing to decisions. Visualization confirms that classification relies on distributed neural patterns rather than single channels.

Input Gradient Magnitude Heatmaps

The heatmaps are generated by averaging absolute input gradients over the test set per class. High-intensity regions in the time-channel plane indicate elements that most influence predictions. This method complements attention-based spatial analysis, providing both temporal and spatial interpretability. The combination of TTA, class-wise metrics, saliency maps, attention analysis, and gradient heatmaps ensures that the DualAttentionNet model is not only accurate but interpretable and scientifically meaningful for EEG-based digit classification.

4 Results

This section explains the experimental results of the DualAttentionNet and discusses its effectiveness for EEG-based MNIST digit classification. The evaluation focuses on three main aspects: a) The classification performance, b) Explainability analysis, and c) Comparison with other models.

4.1 Classification and Performance Analysis

The classification performance of the DualAttentionNet model was evaluated on an 11-class EEG digit recognition of digits 0–9 and a black screen class. The proposed model achieves a test accuracy of 92.36%, which is showing strong performance for multi-class EEG classification (Table 4). The learning curves further demonstrate stable convergence behaviour during training (Figure 6). Both training and validation loss decrease consistently during the early epochs, indicating effective optimisation and stable learning dynamics. The close alignment between the training and validation curves also shows that the model does not suffer from significant overfitting

and maintains good generalisation performance. In addition to accuracy-based evaluation, the discriminative capability of the proposed model was further assessed using the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC). The multi-class ROC analysis demonstrates that the classifier achieves strong separability among the EEG digit classes, with a micro-average AUC score of 0.99 illustrated in Figure 7.

Table 4 Classification performance of the proposed model for each digit class and the Black Screen category.

Class	Precision	Recall	F1-score	Support
Digit 0	0.885	1.000	0.939	23
Digit 1	0.929	0.929	0.929	28
Digit 2	0.840	0.913	0.875	23
Digit 3	0.714	0.652	0.682	23
Digit 4	0.889	0.960	0.923	25
Digit 5	0.762	0.762	0.762	21
Digit 6	0.909	0.909	0.909	22
Digit 7	1.000	0.821	0.902	28
Digit 8	0.842	0.762	0.800	21
Digit 9	0.680	0.739	0.708	23
Black Screen	1.000	1.000	1.000	234
Accuracy			**0.924**	471



Fig. 6 Training and Validation loss and accuracy Curves of the Proposed Dual-Attention EEGNet.

The results were compared with several existing studies in EEG-based visual stimulus recognition to further validate the proposed method. Table 1 summarises the comparison between previous works. For example, Mishra et al. achieved approximately 70% accuracy using a CNN-based framework with spatially selected EEG

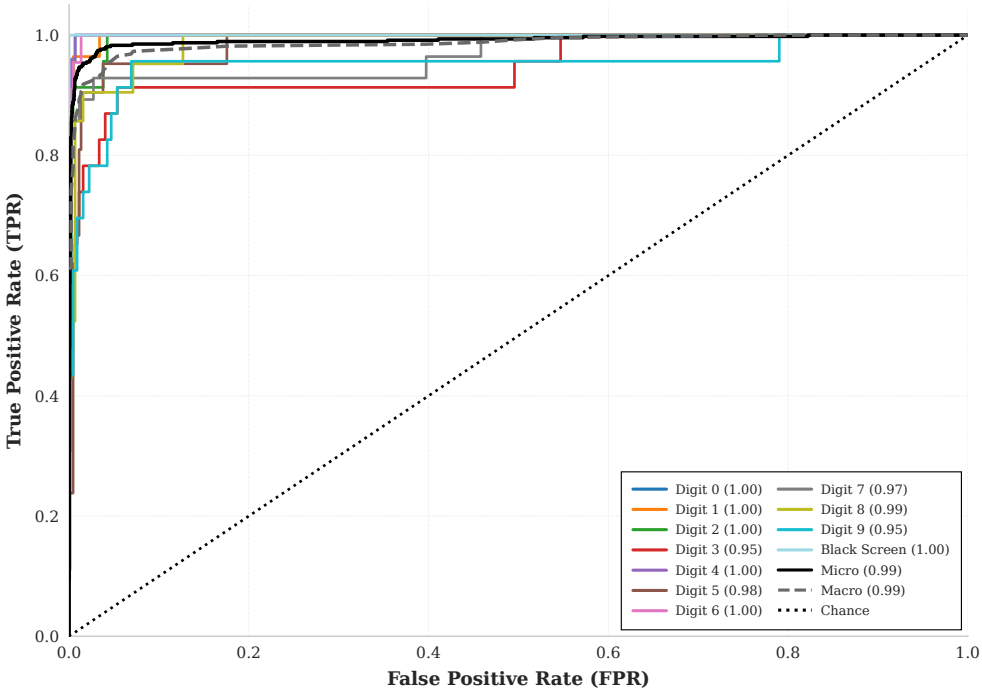


Fig. 7 Multi-class ROC curves for EEG digit classification showing strong separability across classes with a micro-average AUC.

channels. Although their results demonstrate the feasibility of convolutional models for EEG decoding, the reliance on manually selected electrodes may limit generalisation. The results demonstrate that the proposed model achieves higher classification accuracy while maintaining computational efficiency.

4.2 Explainability Analysis

An explainability analysis was conducted using three techniques to better understand how the proposed model makes its predictions. These methods provide insight into the temporal and spatial patterns that influence model decisions.

Temporal Saliency Analysis

Temporal saliency maps highlight the signal segments that contribute most to the model's predictions (Figure 8). For correctly classified samples such as Digit 9, the important regions are concentrated within specific time intervals rather than distributed across the entire signal. This indicates that the network captures structured temporal EEG patterns associated with the visual stimulus. In contrast, the Black Screen example shows lower and more uniformly distributed importance, which is expected since the signal contains minimal stimulus-related activity. Misclassified cases

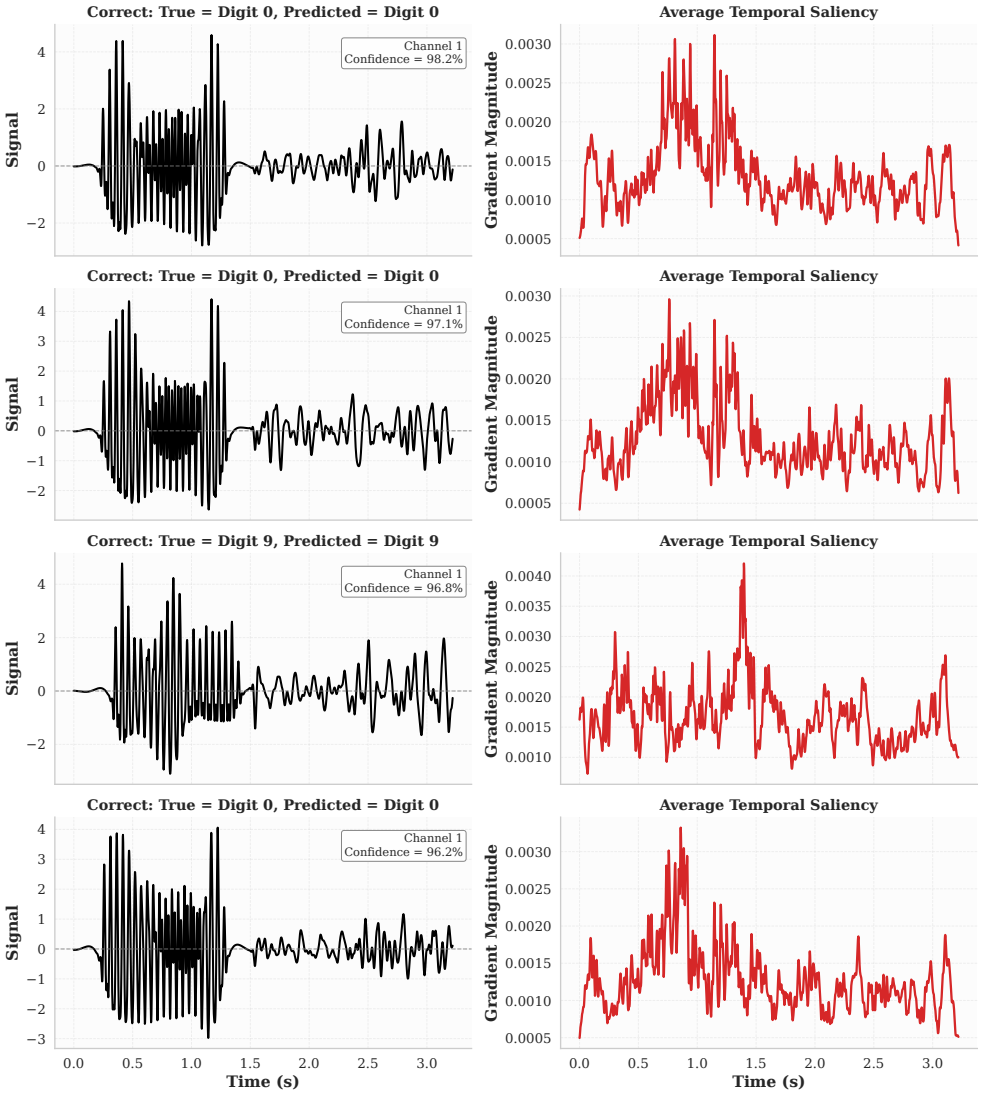


Fig. 8 Gradient-based temporal saliency analysis for correctly classified and misclassified EEG samples.

(Digit 8 predicted as Digit 3) exhibit more scattered saliency patterns without a clear dominant region, suggesting overlapping temporal features between classes.

Channel Attention Analysis

Channel attention weights extracted from the two SE blocks reveal how the network prioritises spatial EEG features (Figure 9). The attention weights remain relatively

uniform, indicating that most channels contribute similarly during early feature extraction in Channel Attention Block 1 (after spatial convolution), while the weights become more varied with certain channels receiving higher importance in Channel Attention Block 2 (after separable convolution). The deeper layers perform stronger feature selection by emphasising more discriminative spatial representations.

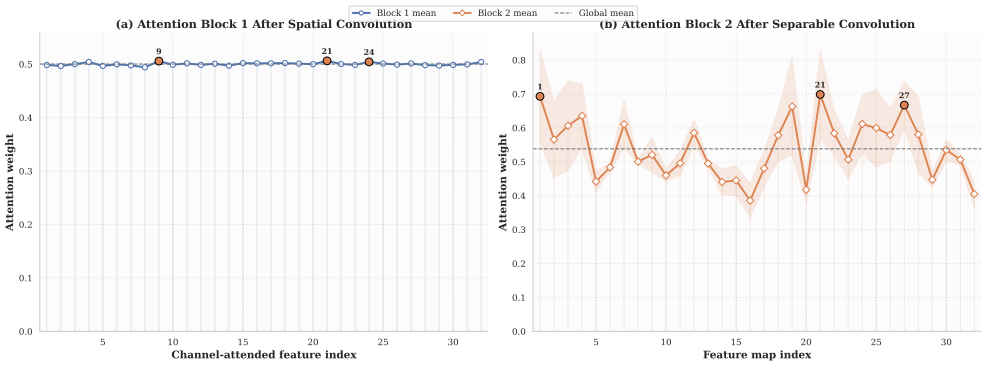


Fig. 9 Channel attention weight distribution after spatial and separable convolution stages.

Input Gradient Magnitude Heatmaps

Input gradient magnitude heatmaps were computed by averaging the absolute gradients of the loss with respect to the EEG input across the test set (Figure 10) to further analyze model sensitivity. These heatmaps reveal the time–channel regions that most influence the model's predictions. The results show that stronger gradients are typically concentrated in early temporal segments and lower EEG channels, indicating that these regions contribute most to the classification process. At the same time, weaker but distributed activations across other channels rely on multiple spatial patterns rather than a single dominant electrode.

4.3 Comparison with Other Models

In this section, we compare the proposed model with several previous studies and other models. Table 5 compares the statistical characteristics of the previous dataset and our dataset. While the earlier study reports Gaussian mean, standard deviation (SD), and interquartile range (IQR), our dataset includes the number of samples, mean prediction confidence, standard deviation, and median. The results show higher mean confidence values for most classes in our dataset, particularly Digit 1, Digit 6, Digit 7, and Black Screen. For instance, the mean confidence for Digit 1 increases from 0.4176 to 0.7268, and Black Screen from 0.5341 to 0.7174, indicating stronger and more confident predictions.

Table 6 shows the overall performance comparison across different models. The proposed DualAttentionNet achieves the highest classification accuracy of 92.36%

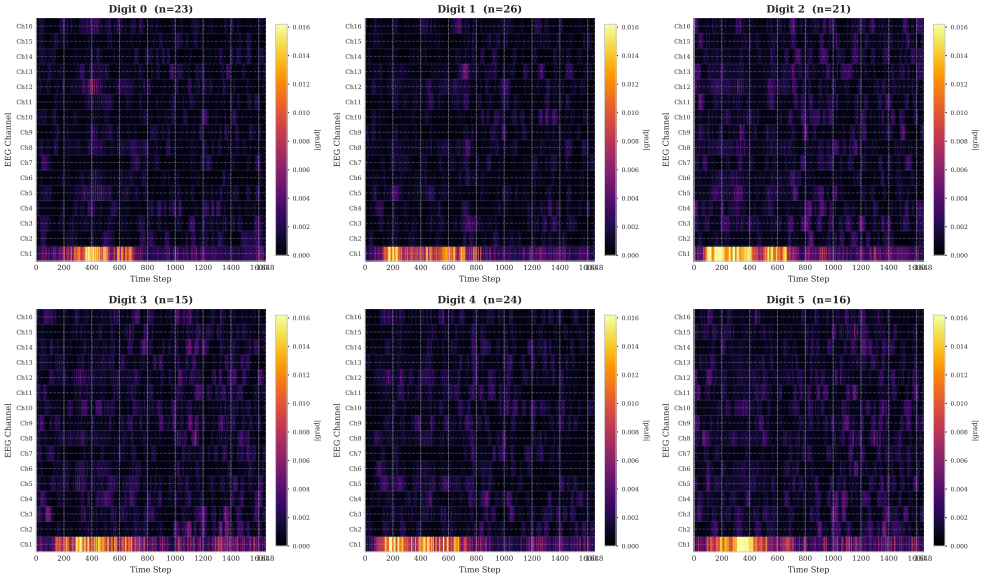


Fig. 10 Input gradient magnitude heatmaps averaged per class. High-intensity regions indicate time–channel combinations that strongly influence model predictions.

Table 5 Statistical distribution comparison between the previous study and the proposed DualAttentionNet.

Class	Samples	Falciglia et al. [31]			DualAttentionNet		
		Mean	SD	IQR	Mean	SD	IQR
Black screen	234	0.5341	0.0657	0.0817	0.7174	0.0870	0.1228
Digit 0	23	0.5972	0.0763	0.0954	0.7672	0.2048	0.2955
Digit 1	28	0.4176	0.0756	0.0944	0.7268	0.1742	0.1499
Digit 2	23	0.5455	0.0731	0.0907	0.7404	0.2323	0.3329
Digit 3	23	0.4707	0.0861	0.1066	0.5369	0.2980	0.5337
Digit 4	25	0.5152	0.1061	0.1327	0.6332	0.1915	0.2765
Digit 5	21	0.4966	0.0953	0.1179	0.4988	0.2586	0.3264
Digit 6	22	0.5111	0.1063	0.1328	0.7065	0.2406	0.3290
Digit 7	28	0.4700	0.0976	0.1216	0.6154	0.2740	0.3195
Digit 8	21	0.4048	0.0837	0.1040	0.4534	0.2190	0.2736
Digit 9	23	0.4814	0.0847	0.1038	0.5406	0.2714	0.4285

outperforming all baseline architectures. EEGConformer achieves the second-best performance with 84.93% accuracy, followed by the baseline EEGNet with 89.17%. ShallowFBCSPNet achieves 77.49%, while ATCNet obtains 58.17%. In addition to accuracy, the proposed model also achieves the highest weighted F1-score 0.9149 and Cohen's kappa score 0.8950, indicating strong agreement between predictions and ground truth labels.

Table 6 Overall performance comparison of the evaluated models.

Model	Accuracy (%)	F1-score	κ	ROC–AUC	Params	Time (ms)
DualAttentionNet	**92.36**	**0.9232**	**0.8950**	**0.9915**	22,219	1.758
EEGNet	89.17	0.8891	0.8512	0.9904	10,347	0.332
EEGConformer	84.93	0.8475	0.7928	0.9909	1,220,523	0.647
ShallowFBCSPNet	77.49	0.7688	0.6907	0.9719	72,491	7.169
ATCNet	58.17	0.5714	0.4222	0.9132	114,695	0.787

Table 7 Per-class F1-score comparison across the evaluated models.

Class	DualAttentionNet	EEGNet	ShallowFBCSPNet	EEGConformer	ATCNet
Black screen	1.0000	1.0000	1.0000	1.0000	0.9894
Digit 0	0.9388	0.8333	0.6818	0.7660	0.1714
Digit 1	0.9286	0.9286	0.6176	0.7541	0.0000
Digit 2	0.8750	0.7391	0.5333	0.6667	0.2105
Digit 3	0.6818	0.4889	0.4800	0.6182	0.2143
Digit 4	0.9231	0.9020	0.3784	0.7660	0.2000
Digit 5	0.7619	0.5714	0.2581	0.4375	0.1356
Digit 6	0.9091	0.8800	0.6800	0.8444	0.2545
Digit 7	0.9020	0.8627	0.7917	0.7600	0.1739
Digit 8	0.8000	0.7179	0.3077	0.6364	0.1463
Digit 9	0.7083	0.7925	0.5806	0.6667	0.1081

The results demonstrate strong discriminative capability across multiple metrics. The proposed model achieves an ROC-AUC of 0.9915 while maintaining a lightweight architecture with 22,219 parameters and fast inference speed (1.758 ms per sample), showing a good balance between accuracy, complexity, and efficiency. Table 7 presents the per-class F1-score comparison. The proposed Dual-Attention EEGNet achieves strong performance across most digit classes and maintains perfect classification for the Black Screen class, with notable improvements for classes such as Digit 4 and Digit 6, indicating effective learning of discriminative EEG patterns. Compared with baseline architectures and previous studies' approaches, the proposed DualAttentionNet demonstrates superior performance across multiple evaluation metrics and class-level

analyses. As a result, the combination of architectural enhancement, optimised training strategy, and structured attention mechanisms significantly improves EEG-based digit classification performance while maintaining computational efficiency.

5 Conclusion

This study represents an improved deep learning framework for multi-class EEG-based digit classification. It demonstrates that structured attention mechanisms significantly improve feature representation by allowing the network to emphasise informative EEG channels and suppress less relevant signals. In comparison with previous studies, which often focused on binary classification, lower accuracies, or simplified architectures, this work addresses a more challenging fine-grained multi-digit classification task and provides a reproducible and well-documented pipeline. The implementation of explainability analysis also enhances interpretability, offering insights into how the model makes its decisions.

Future research may extend this framework by incorporating cross-subject generalisation, larger datasets, multimodal fusion (such as combining EEG with eye-tracking or EMG), and transformer-based hybrid attention mechanisms. Further improvements in real-time inference and lightweight deployment could support practical applications in brain–computer interfaces, assistive communication systems, and cognitive state monitoring. This study demonstrates that integrating dual attention mechanisms with a carefully optimised training strategy leads to more reliable, robust, and interpretable EEG-based digit classification, contributing to the advancement of EEG-driven cognitive decoding systems.

Declarations

Funding. No funding was received for conducting this study or for the preparation of this manuscript.

References

[1] Ahuja, C., Sethia, D.: Harnessing few-shot learning for eeg signal classification: A survey of state-of-the-art techniques and future directions. Frontiers in Human Neuroscience **18** (2024) https://doi.org/10.3389/fnhum.2024.1421922

[2] Kantipudi, M.V.V.P., Kumar, N.S.P., Aluvalu, R., Selvarajan, S., Kotecha, K.: An improved gbso-taenn-based eeg signal classification model for epileptic seizure detection. Scientific Reports **14**(1) (2024) https://doi.org/10.1038/s41598-024-51337-8

[3] Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. Journal of Neural Engineering **15**(5), 056013 (2018) https://doi.org/10.1088/1741-2552/aace8c

[4] Das, A., Singh, S., Kim, J., Ahanger, T.A., Pise, A.A.: Enhanced eeg signal classification in brain computer interfaces using hybrid deep learning models. Scientific Reports **15**(1), 27161 (2025) https://doi.org/10.1038/s41598-025-07427-2

[5] Singh, G., Adira, P., Leema, A., Balakrishnan, P.: EEG-Based Brain Signals Analysis for Imagined Digit Recognition. Preprint (2024)

[6] Gao, J., Liu, Y., Li, Z., Huang, K., Wang, F., Xu, J., Zhao, L., Li, T., Fu, Y.: An eeg dataset for visual imagery-based brain–computer interface. Scientific Data **13**(1) (2026) https://doi.org/10.1038/s41597-025-06512-5

[7] Hakkak, H., Khalilzadeh, M.M., Azarnoosh, M., Kobravi, H.: Evaluation of cnn and rnn architectures for multiclass visual content classification from visually evoked eeg signals. Engineering Research Express **8**(4), 045211 (2026) https://doi.org/10.1088/2631-8695/ae43b0

[8] Khaleghi, N., Hashemi, S., Ardabili, S.Z., Sheykhivand, S., Danishvar, S.: Salient arithmetic data extraction from brain activity via an improved deep network. Sensors **23**(23), 9351 (2023)

[9] Zeynali, M., Seyedarabi, H., Afrouzian, R.: Classification of eeg signals using transformer-based deep learning and ensemble models. Biomedical Signal Processing and Control **86**, 105130 (2023)

[10] Kamil Gatfan, S.: A review on deep learning for electroencephalogram signal classification. Journal of Al-Qadisiyah for Computer Science and Mathematics **16**(1), 137–151 (2024) https://doi.org/10.29304/jqcsm.2024.16.11453

[11] Hamidi, A., Kiani, K.: Motor imagery eeg signals classification using a transformer-gcn approach. Applied Soft Computing **170**, 112686 (2025) https://doi.org/10.1016/j.asoc.2024.112686

[12] Mathiyazhagan, S., Devasena, M.S.G.: Motor imagery eeg signal classification using novel deep learning algorithm. Scientific Reports **15**(1) (2025) https://doi.org/10.1038/s41598-025-00824-7

[13] Padhmashree, V., Bhattacharyya, A.: Human emotion recognition based on time–frequency analysis of multivariate eeg signal. Knowledge-Based Systems **238**, 107867 (2022)

[14] N, J.N., Gore, D., G, V., E, N., V, N., K, K.: Efficient predefined time adaptive neural network for motor execution eeg signal classification based brain-computer interaction. Knowledge-Based Systems **303**, 112270 (2024) https://doi.org/10.1016/j.knosys.2024.112270

[15] Banovoth, R.S., K. V., K.: Roman domination-based spiking neural network for

optimized eeg signal classification of four-class motor imagery. Computers in Biology and Medicine **194**, 110397 (2025) https://doi.org/10.1016/j.compbiomed.2025.110397

[16] Vivancos, D.: MindBigData 2023 MNIST-8B: The 8 Billion Datapoints Multimodal Dataset of Brain Signals (2023)

[17] Vivancos, D., Cuesta, F.: Mindbigdata 2022: A large dataset of brain signals. arXiv preprint arXiv:2212.14746 (2022) arXiv:2212.14746

[18] Hwaidi, J.F., Chen, T.M.: Classification of motor imagery eeg signals based on deep autoencoder and convolutional neural network approach. IEEE Access **10**, 48071–48081 (2022)

[19] Jemal, I., Mezghani, N., Abou-Abbas, L., Mitiche, A.: An interpretable deep learning classifier for epileptic seizure prediction using eeg data. IEEE Access **10**, 60141–60150 (2022)

[20] Tiwari, S., Goel, S., Bhardwaj, A.: Eeg signals to digit classification using deep learning-based one-dimensional convolutional neural network. Arabian Journal for Science and Engineering **48**(8), 9675–9691 (2023)

[21] Wang, X., Liesaputra, V., Liu, Z., Wang, Y., Huang, Z.: An in-depth survey on deep learning-based motor imagery electroencephalogram (eeg) classification. Artificial Intelligence in Medicine **147**, 102738 (2024)

[22] Chen, R., Ma, X., Li, X., Sui, L., Maeda, T., Chen, Q., Cao, J.: Eeg signal classification with data augmentation for epileptic focus localization and deep sleep detection. Sensors **26**(2), 474 (2026) https://doi.org/10.3390/s26020474

[23] Pant, A., Kumar, A.: Exploring eeg signal processing for effective filtering and classification of epileptic seizures. Discover Electronics **3**(1) (2026) https://doi.org/10.1007/s44291-026-00174-2

[24] Li, H., Liu, J., Li, J.: Cross-subject motor imagery eeg signal classification based on meta-transfer learning. Computer Methods in Biomechanics and Biomedical Engineering, 1–12 (2026)

[25] Morvay, B.T., Torma, S., Pitrik, J., Szegletes, L.: Enhancing multi-paradigm eeg signal classification in cross-subject settings using optimal transport. Biomedical Signal Processing and Control **113**, 108892 (2026) https://doi.org/10.1016/j.bspc.2025.108892

[26] Mishra, R., Sharma, K., Bhavsar, A.: Visual brain decoding for short duration eeg signals. In: 2021 29th European Signal Processing Conference (EUSIPCO), pp. 1226–1230. IEEE, ??? (2021)

[27] Tsamourgelis, A., Adamopoulos, A.: Eeg signal analysis for numerical digit classification: Methodologies and challenges. Stats **8**(1), 14 (2025)

[28] Tanner, T., Cutsuridis, V.: Generative Adversarial Network for Image Reconstruction from Human Brain Activity. Unpublished manuscript (2023)

[29] Pratama, I.W.P., *et al.*: Frequency band and pca feature comparison for eeg signal classification. Lontar Komputer: Jurnal Ilmiah Teknologi Informasi **12**(1), 1 (2021)

[30] Chowdhury, M.R., Ding, Y., Sen, S.: SSL-SE-EEG: A Framework for Robust Learning from Unlabeled EEG Data with Self-Supervised Learning and Squeeze-Excitation Networks. Unpublished manuscript (2024)

[31] Falciglia, S., Betello, F., Russo, S., Napoli, C.: Learning visual stimulus-evoked eeg manifold for neural image classification. Neurocomputing **588**, 127654 (2024)