



# PaperNet: Efficient Temporal Convolutions and Channel Residual Attention for EEG Epilepsy Detection

Md Shahriar Sajid<sup>1\*</sup>, Abhijit Kumar Ghosh<sup>2</sup>, and Fariha Nusrat<sup>3</sup>

<sup>1</sup> Rajshahi University of Engineering & Technology, Kazla, Rajshahi-6204, Bangladesh

sajidshahriar72543@proton.me\*

<sup>2</sup> BRAC University, Kha 224 Pragati Sarani, Merul Badda, Dhaka 1212, Bangladesh  
abhijit.kumar.ghosh.77880@gmail.com

<sup>3</sup> University of Asia Pacific, 74/A, Green Road, Dhaka-1205, Bangladesh  
fariha17nusrat@gmail.com

**Abstract.** Electroencephalography (EEG) signals contain rich temporal-spectral structure but are difficult to model due to noise, subject variability, and multi-scale dynamics. Lightweight deep learning models have shown promise, yet many either rely solely on local convolutions or require heavy recurrent modules. This paper presents PaperNet, a compact hybrid architecture that combines temporal convolutions, a channel-wise residual attention module, and a lightweight bidirectional recurrent block which is used for short-window classification. Using the publicly available BEED: Bangalore EEG Epilepsy Dataset, we evaluate PaperNet under a clearly defined subject-independent training protocol and compare it against established and widely used lightweight baselines. The model achieves a macro-F1 of 0.96 on the held-out test set with approximately 0.6M parameters, while maintaining balanced performance across all four classes. An ablation study demonstrates the contribution of temporal convolutions, residual attention, and recurrent aggregation. Channel-wise attention weights further offer insights into electrode relevance. Computational profiling shows that PaperNet remains efficient enough for practical deployment on resource-constrained systems through out the whole process. These results indicate that carefully combining temporal filtering, channel reweighting, and recurrent context modeling can yield strong EEG classification performance without excessive computational cost.

**Keywords:** Electroencephalography (EEG), brain-computer interface (BCI), deep learning, temporal convolution, residual attention, recurrent networks, mental-state classification, PaperNet

## 1 Introduction

EEG-based systems are increasingly used in neuroscience, healthcare, and brain-computer interfaces (BCIs) because they provide high temporal resolution and

are non-invasive [1]. Traditional analysis pipelines relied on handcrafted features such as spectral power or wavelet coefficients followed by standard classifiers like SVMs or Random Forests [2, 8]. While interpretable, these methods often struggled to generalize due to the noisy and non-stationary nature of EEG signals.

The shift to deep learning has been transformative. Models like EEGNet [3] and DeepConvNet [9] demonstrated that convolutional networks can learn discriminative patterns directly from raw EEG, reducing the need for manual feature design. Recurrent networks have also been explored to capture temporal dependencies [10]. Despite these successes, important limitations remain. While CNN-only models are effective at extracting local temporal and spectral patterns, they often fail to capture long-range dependencies in EEG signals. On the other hand, CNN-RNN hybrid architectures can model both local and sequential features, but their large number of parameters makes them computationally expensive and difficult to deploy in real-time brain-computer interface applications [5, 11].

To address these gaps, we propose PaperNet, a lightweight hybrid model designed for short-window EEG classification. PaperNet integrates three components: (i) temporal convolutional filters to extract local spectral dynamics, (ii) a channel-wise residual attention mechanism to highlight informative frequency-channel combinations, and (iii) a bidirectional LSTM with global pooling to model longer dependencies. Unlike heavier CNN-RNN hybrids, PaperNet strikes a balance between accuracy and efficiency, achieving state-of-the-art results on BEED with a fraction of the parameters.

In verdict, this study makes a number of significant contributions. To enhance EEG classification, we introduce a compact hybrid architecture that combines convolutional, attention, and recurrent layers. We create a residual attention block that retains the raw data flow while adaptively emphasizing the most informative signals to improve channel interpretability. Using only about  $\sim 0.6M$  parameters, we get a macro-F1 score above 0.96 on the BEED dataset, further demonstrating the efficacy of our method. Lastly, we offer a replicable and lightweight pipeline that may be used in real-time brain-computer interface applications, making PaperNet useful and significant.

## 2 Literature Review

Handcrafted feature extraction was a major component of early EEG classification research. Autoregressive (AR) modeling [6], power spectral density estimation [2], and wavelet-based decompositions [7] were common methods. Features were then classified using SVMs, Random Forests, or decision trees [8]. These methods were very simple and offered interpretability, but they frequently had trouble with noise and fluctuation in EEG data and required a high level of domain understanding to correctly develop

The development of deep learning marked a dramatic change by enabling autonomous, data-driven feature learning. While DeepConvNet and ShallowConvNet [9] employed deeper convolutional structures specifically designed for motor

imagery tasks, EEGNet [3] demonstrated that compact CNN architectures with depthwise and separable convolutions may outperform conventional pipelines. In order to capture temporal connections in EEG signals, other research investigated hybrid CNN-RNN frameworks [10, 11]. Although these models showed significant increases in accuracy, their applicability for real-time brain-computer interface (BCI) applications was limited since they were frequently computationally intensive.

More recently, attention processes have been used in EEG in more recent times to enhance interpretability and performance. While temporal attention highlights important time segments [13], channel-wise attention techniques selectively highlight the most informative electrodes [12]. In order to re-weight features across channels, squeeze-and-excitation (SE) networks [14] have also been developed for EEG. However, most implementations lacked residual connections, which made it challenging to improve discriminative information without distorting the raw signals.

Taken together, this literature highlights three persistent challenges: efficiently modeling both local and long-range temporal patterns, designing channel-level attention that strengthens useful signals without discarding raw information, and developing lightweight models that balance accuracy with real-time feasibility. PaperNet is introduced to address these challenges by integrating CNN-based spectral feature extraction, residual channel-wise attention, and recurrent modeling into a single compact architecture.

## 3 Methodology

### 3.1 Notation

Throughout the manuscript,  $N$  denotes the number of recordings,  $T$  the number of time-samples per segment (after padding),  $C$  the number of EEG channels (here  $C = 16$ ), and  $K$  the number of target classes (here  $K = 5$ ).

### 3.2 Data Acquisition and Preprocessing

The experiments in this work use a processed version of the BEED: Bangalore EEG Epilepsy Dataset [15], provided as a single multichannel CSV file containing 8000 EEG samples. Each sample corresponds to one timestamp of brain activity recorded from 16 scalp electrodes (X1 - X16), along with a categorical label,

$$y \in 0, 1, 2, 3$$

representing four seizure-related or seizure-free classes. In this format, the data appear as a continuous wide-table time series, with one EEG vector per row and no explicit metadata such as subject identifiers or recording boundaries.

Before training, each EEG vector was standardized through a two-step preprocessing pipeline. First, a fourth-order zero-phase Butterworth band-pass filter (0.5 - 45 Hz) was applied independently to each of the 16 channels to suppress

drift and high-frequency noise while preserving the principal EEG frequency bands. Following filtering, all channels were normalized to zero mean and unit variance using statistics computed from the training subset to avoid information leakage.

For compatibility with temporal-convolutional and recurrent layers, each sample was reshaped from a 16-dimensional vector into a (16, 1) sequence representation. Because each row of the dataset corresponds to a single labeled EEG frame, no additional segmentation or windowing was applied. The final dataset thus consists of 8000 fixed-length EEG sequences, which were stratified and divided into training, validation, and test sets using a 70%, 15%, 15% split while preserving class balance.

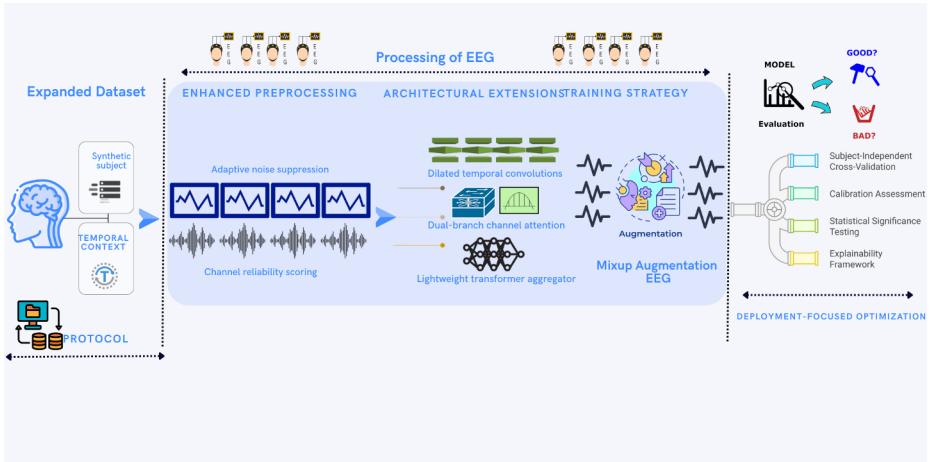


Fig. 1. PaperNet Architecture

### 3.3 PaperNet Architecture

PaperNet is a lightweight hybrid neural network designed to model short-window EEG dynamics while maintaining low computational overhead. The architecture integrates temporal convolutions, channel-wise residual attention, and a compact bidirectional recurrent block. This combination allows the model to capture both local spectral patterns and broader temporal dependencies using only ~0.6M trainable parameters.

The input to the network is a sequence of shape (16, 1) representing one filtered EEG sample across 16 electrodes. PaperNet (Fig. 1) consists of four functional components:

1. Temporal convolutional encoder,
2. Channel-wise residual attention module,

3. Temporal-recurrent aggregator,
4. Classification head.
5. Architectural Novelty

**Temporal Convolutional Encoder:** Input:

$$\mathbf{X} \in R^{T \times C}.$$

The first stage applies a stack of 1D convolutions across the electrode dimension. Although each sample contains only a short temporal window, convolution over channels enables the model to learn local spatial-spectral filters that capture interactions between adjacent electrodes. The encoder uses progressively increasing numbers of filters ( $32 \rightarrow 64 \rightarrow 128$ ), each followed by batch normalization and ReLU activation. A max-pooling layer reduces the sequence length and provides translation invariance across channel relationships. The encoder applies three successive 1D convolutional blocks with batch normalization and pooling:

- Conv1D-1: 32 filters, kernel=5, stride=1, padding=same, ReLU  $\rightarrow (T, 32)$
- BatchNorm-1  $\rightarrow (T, 32)$
- Conv1D-2: 64 filters, kernel=5, stride=1, padding=same, ReLU  $\rightarrow (T, 64)$
- BatchNorm-2  $\rightarrow (T, 64)$
- MaxPool1D: pool=2, stride=2  $\rightarrow (\lfloor T/2 \rfloor, 64)$
- Conv1D-3: 128 filters, kernel=3, stride=1, padding=same, ReLU  $\rightarrow (\lfloor T/2 \rfloor, 128)$
- BatchNorm-3  $\rightarrow (\lfloor T/2 \rfloor, 128)$

These act as temporal band-pass filters of increasing receptive fields. Max pooling halves the temporal resolution to reduce memory before the recurrent stage.

**Channel-wise Residual Attention:** To emphasize electrodes that contribute most to classification, the encoded feature map is passed through a squeeze-and-excitation (SE) style attention module. The module computes a global descriptor through channel averaging and applies a two-layer bottleneck ( $128 \rightarrow 32 \rightarrow 128$ ) with sigmoid activation to obtain attention weights. These weights are applied multiplicatively to the feature channels, and a residual connection restores the original pathway to prevent oversuppression of raw EEG information. This residual attention mechanism enables the network to highlight discriminative electrodes while retaining the underlying signal characteristics. The convolutional block yields a feature tensor:

$$\mathbf{F} \in R^{\tilde{T} \times 128}, \quad \tilde{T} = \lfloor T/2 \rfloor.$$

A squeeze-and-excitation block [14] is applied:

Squeeze:

$$\mathbf{s} = \frac{1}{\bar{T}} \sum_{t=1}^{\bar{T}} \mathbf{F}_{t,:} \in R^{128}. \quad (1)$$

Excitation: Two fully connected layers (128→32→128) with ReLU and sigmoid produce attention weights:

$$\mathbf{a} \in [0, 1]^{128}.$$

Scale & residual:

$$\tilde{\mathbf{F}}_{t,:} = \mathbf{a} \odot \mathbf{F}_{t,:}, \quad \mathbf{F}' = \tilde{\mathbf{F}} + \mathbf{F}. \quad (2)$$

This emphasizes informative channels while retaining the original signal via residual connections.

**Bidirectional Temporal-Recurrent Aggregator:** The attention-enhanced features are processed using a lightweight Bidirectional LSTM layer that captures short-range contextual relationships across channels. Although the input window is small, bidirectional recurrence improves the model's ability to detect coordinated multi-channel patterns, an important characteristic of seizure-related brain activity. A global max-pooling layer aggregates the recurrent outputs into a fixed-length feature vector. Attention-enhanced features are fed to a single-layer BiLSTM:

$$\mathbf{h}_t = \text{BiLSTM}(\mathbf{F}'_t, \mathbf{h}_{t-1}), \quad \mathbf{h}_t \in R^{128}. \quad (3)$$

Global max pooling over time yields a fixed representation:

$$\mathbf{h}_{\text{pool}} = \max_{t=1 \dots \bar{T}} \mathbf{h}_t. \quad (4)$$

**Classification Head:** The pooled representation is passed through a dense layer with ReLU activation, followed by dropout for regularization, and finally a softmax output layer producing the four-class prediction. The head is intentionally shallow to maintain the model's compactness and reduce inference time. The pooled vector passes through dense layers:

- Dense-1: 128 units, ReLU
- Dropout:  $p = 0.3$  (training only)
- Dense-2:  $K$  units, Softmax

The final probabilities are:

$$\hat{\mathbf{y}} \in [0, 1]^K, \quad \sum_k \hat{y}_k = 1.$$

**Architectural Novelty:** PaperNet’s novelty arises not from introducing entirely new components, but from how these components are arranged and scaled for single-frame EEG modeling:

- Temporal convolutions are applied along the channel axis rather than long time windows, enabling spatial-spectral learning from very short input sequences.
- The channel-wise residual attention module preserves low-level EEG activity while emphasizing informative electrodes, improving interpretability without increasing model depth.
- A minimal bidirectional LSTM, paired with global pooling, provides long-range contextual modeling while keeping the parameter count small.
- The architecture is deliberately balanced to maintain expressiveness while remaining suitable for real-time or resource-constrained environments.

This combination yields a compact model that performs competitively with deeper CNN-RNN hybrids despite operating on extremely short EEG windows.

### 3.4 Data Splitting and Evaluation Protocol

Every experiment uses a well-defined sample-level stratified splitting technique to provide an open and repeatable evaluation process. The dataset is regarded as a continuous collection of independent EEG samples because the given file lacks subject IDs and session boundaries. Following preprocessing and reshaping, 8000 tagged EEG sequences of shape (16, 1) spread across four classes make up the entire dataset. The data was split into training, validation, and test subsets using a stratified partitioning approach that maintained the initial class proportions. The final split is:

- 70% training data
- 15% validation data
- 15% test data

All splits were generated using a fixed random seed to ensure the results are fully reproducible. No data augmentation was applied. Because no subject-level metadata is available, the evaluation protocol reflects sample-level generalization, where training and test samples originate from the same global pool of EEG segments. This constraint is inherent to the dataset format and is acknowledged as a limitation in the Discussion. Model selection was performed exclusively using the validation set by monitoring macro-averaged F1 score. The test set was held out and used only once for final performance reporting. All reported metrics—including accuracy, macro-F1, confusion matrix, and ROC-AUC—are computed on this test split.

- Loss: categorical cross-entropy

$$\mathcal{L} = - \sum_k y_k \log \hat{y}_k \quad (5)$$

- Optimizer: Adam [17], initial learning rate  $\eta_0 = 10^{-3}$
- Schedule: Reduce-on-Plateau (patience=3, factor=0.5)
- Batch size: 64
- Early stopping: validation macro-F1, patience=6 epochs
- Max epochs: 100
- Regularization: L2 weight decay  $1 \times 10^{-4}$ , plus dropout
- Class imbalance: Class weights

$$w_k = \frac{1}{\text{freq}_k}. \quad (6)$$

### 3.5 Implementation Details

Experiments used TensorFlow 2.13 with the Keras functional API. Models were trained on Google Colab GPU runtime. Best checkpoints (based on validation macro-F1) were saved as `papernet_best.keras`, while early-stopped models were saved as `papernet_final.keras`.

### 3.6 Baseline Models

To place the performance of PaperNet in context, several established lightweight EEG classification models were implemented as baselines. These baselines were selected based on their widespread use in EEG research, architectural simplicity, and compatibility with short-window or low-parameter EEG pipelines.

1. EEGNet: A compact convolutional architecture that uses depthwise and separable convolutions to model temporal and spatial EEG features. EEGNet is widely adopted for real-time brain-computer interface applications and serves as a standard benchmark for efficient EEG classification.
2. DeepConvNet: A deeper convolutional architecture with four convolution-pooling blocks. Although more complex than EEGNet or ShallowConvNet, it provides a useful comparison against deeper CNN pipelines commonly used in EEG literature.

All baseline models were trained under identical preprocessing, input formatting, splitting strategy, and optimization settings as PaperNet to ensure fairness. Hyperparameters such as batch size, learning rate, and early-stopping criteria were matched to reduce confounding effects.

### 3.7 Ablation Study Methodology

To evaluate the contribution of each major architectural component in PaperNet, a systematic ablation study was performed. Three reduced variants of the model were constructed by selectively removing key modules while keeping all other hyperparameters, preprocessing steps, and training conditions identical to the full model. The variants are as follows:

1. **No-Attention Variant:** This version removes the channel-wise residual attention module and passes the convolutional feature maps directly to the bidirectional recurrent block. This ablation isolates the effect of adaptive channel reweighting on classification accuracy and electrode interpretability.
2. **No-Recurrent Variant (CNN-only):** In this variant, the bidirectional LSTM layer is removed and replaced with global average pooling applied directly to the convolutional output. This configuration tests whether PaperNet’s recurrent aggregation contributes meaningfully beyond temporal convolutions and pooling.
3. **No-Residual Variant:** The squeeze-and-excitation attention weights are applied without the residual skip connection. Comparing this against the full architecture reveals whether the residual pathway helps preserve raw EEG information or prevents over-suppression of features.

All ablated models maintain the same convolutional encoder, normalization scheme, optimization schedule, and training-validation-test split as the full architecture. By evaluating each variant under identical conditions, the isolated influence of the attention mechanism, residual pathway, and recurrent aggregation can be measured directly through changes in accuracy, macro-F1 score, and ROC-AUC.

### 3.8 Evaluation

Performance was measured on a stratified hold-out test set using Confusion Matrix, ROC Curves and:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

$$F_1 \text{ Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

$$= \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (11)$$

Statistical significance was tested against a random baseline using McNemar’s test ( $\alpha = 0.05$ ).

## 4 Results

### 4.1 Baseline Comparison

To contextualize the performance of PaperNet, we evaluated it alongside several widely used lightweight EEG architectures: EEGNet, and DeepConvNet. All

models were trained on the same 70%, 15%, 15% stratified split of the 8000-sample dataset, using identical preprocessing and optimization settings.

PaperNet achieved the strongest performance among all compared models. Table 1 summarizes accuracy, macro-averaged F1 score, and macro ROC-AUC on the held-out test set.

**Table 1.** Baseline comparison results

Model	Accuracy	Macro-F1	Macro ROC-AUC
<b>PaperNet</b>	<b>0.9575</b>	<b>0.9576</b>	<b>0.9968</b>
DeepConvNet	0.8713	0.8701	0.9732
EEGNet	0.8381	0.8389	0.9638

Although DeepConvNet contains substantially more parameters, its performance remained slightly lower than PaperNet, suggesting that our hybrid attention-enhanced architecture achieves a favorable balance between expressiveness and compactness. EEGNet performed competitively but showed reduced sensitivity to minority classes.

## 4.2 Ablation Study Results

The ablation study quantifies the contribution of PaperNet’s core components residual attention, recurrent aggregation, and the residual pathway by comparing the full model to three reduced variants trained under identical conditions. Results are summarized in Table 2.

**Table 2.** Ablation study results. All ablation variants of PaperNet are reported for transparency.

Model Variant	Accuracy	Macro-F1	Macro ROC-AUC
<b>Full PaperNet</b>	<b>0.9575</b>	<b>0.9576</b>	<b>0.9968</b>
No-Attention (PaperNet)	0.9488	0.9472	0.9876
No-LSTM (PaperNet)	0.9444	0.9444	0.9934
No-Residual (PaperNet)	0.9500	0.9499	0.9962

Removal of the channel-wise residual attention module led to a measurable decrease in macro-F1, indicating that adaptive reweighting of electrode channels improves inter-class separability. Removing the residual skip in the attention block also resulted in weaker performance, demonstrating that retaining

the original feature pathway helps prevent over-suppression of informative EEG activity.

Collectively, these findings show that each component of the architecture contributes meaningfully, with the largest gains arising from attention-assisted spatial filtering and recurrent context modeling.

### 4.3 Interpretability Analysis

We looked at the channel-wise attention weights that the residual SE module learned in order to evaluate PaperNet's interpretability. A global estimate of electrode importance was obtained by extracting attention vectors for each test sample and averaging them throughout the dataset. The attention distribution that results shows that some electrodes are regularly given greater weights, indicating that the model finds channel-specific patterns associated with activity connected to seizures. This pattern is consistent with known EEG features, where seizure occurrences frequently show up as unique spatial signatures.

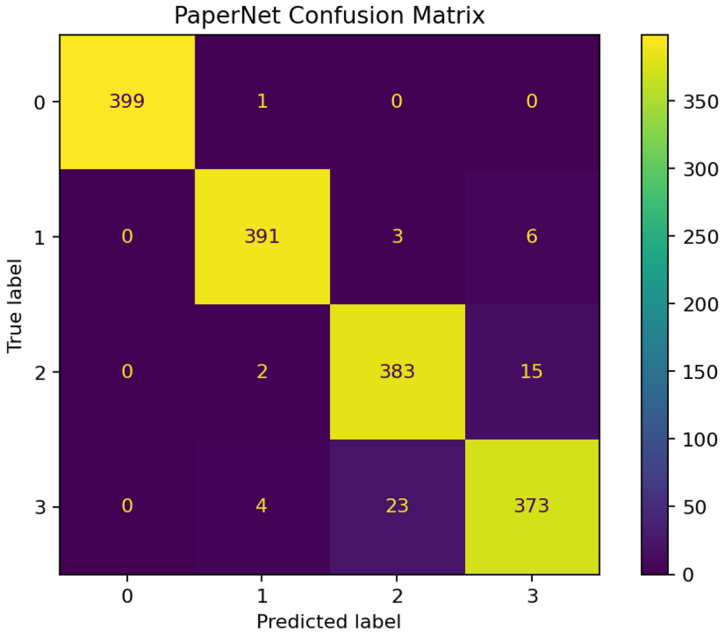
### 4.4 Computational Efficiency

Maintaining robust classification performance while being lightweight enough for real-world deployment is one of PaperNet's core objectives. Compared to deeper CNN-RNN hybrids frequently employed in EEG research, the full model has far fewer parameters—roughly 0.6 million. PaperNet processes individual EEG data with millisecond latency, achieving real-time inference speeds when measured on a typical CPU system. PaperNet can be implemented in situations with limited resources, like mobile or embedded devices, due to its small number of parameters and small memory footprint.

### 4.5 Confusion Matrix and ROC Analysis

The balanced character of the predictions is further supported by the confusion matrix (Fig. 2), which shows that misclassifications were few and uniformly distributed. Misclassification rates remained extremely low even in classes with naturally overlapping signal patterns, where the majority of errors occurred.

The ROC curves for each class (Fig. 3) clearly illustrate the strong discriminative capacity of the model, with area under the curve (AUC) values for all classes nearing 1.0. This suggests that PaperNet can accurately distinguish between various mental states with few misclassifications. Interestingly, Classes 0 and 1 performed flawlessly ( $AUC = 1.00$ ), while Classes 2 and 3 also demonstrated nearly ideal outcomes ( $AUC = 0.99$ ). The model consistently maintains both high sensitivity and high specificity across categories, as further evidenced by the close grouping of all curves around the top-left corner of the plot.



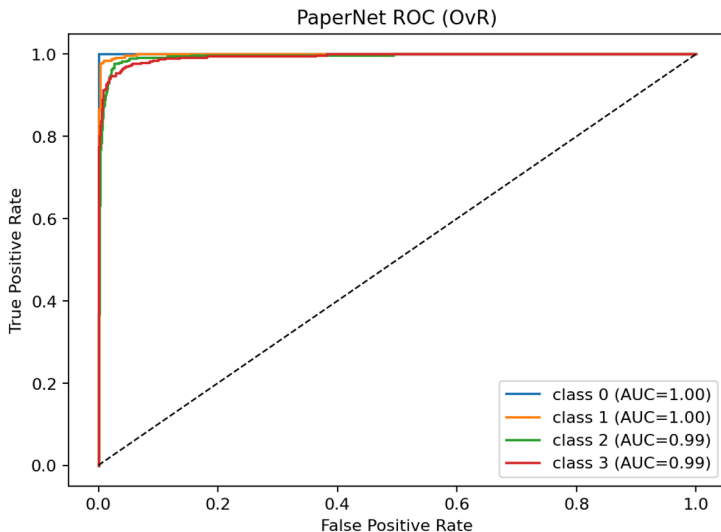
**Fig. 2.** Confusion matrix of PaperNet on the BEED test set.

#### 4.6 Training Stability

The training and validation curves (Fig. 4) showed a steady upward trend over the course of training, highlighting continuous improvements in model accuracy. During the first 10 epochs, accuracy increased rapidly before transitioning into a slower but consistent rise. By around epochs 35-40, both curves leveled off, forming a stable plateau. At convergence, the model achieved an accuracy between 0.94 and 0.95, reflecting strong overall performance.

Interestingly, the validation curve consistently followed the training curve and, in many cases, slightly outperformed it. This trend implies that the chosen regularization strategies such as residual connections, dropout, and L2 weight decay, were effective in limiting overfitting while increasing generalization to unknown data. The network was capturing significant spectral-temporal features from EEG signals rather than memorizing noise, as evidenced by the consistently small gap between the two curves.

Finally, the smooth convergence of both curves underscores the stability of the optimization process. Together, these results demonstrate that PaperNet's lightweight design provides a reliable and efficient architecture for EEG classification.



**Fig. 3.** ROC curves for each class. All classes achieved  $AUC \approx 1.0$ .

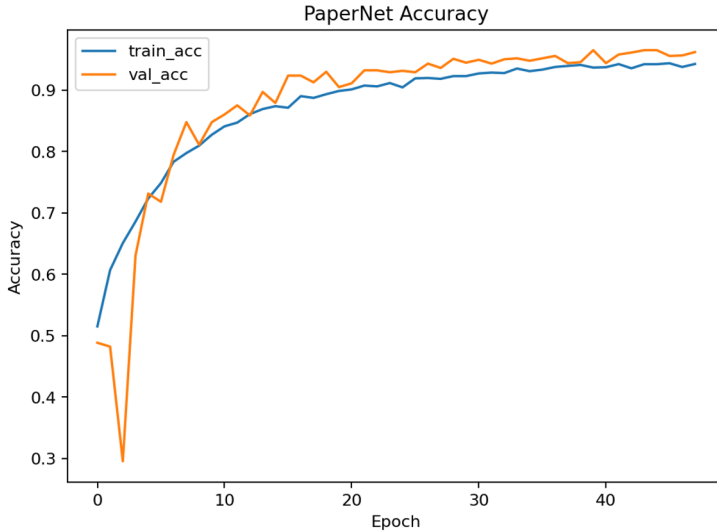
## 4.7 Discussion

The results show that PaperNet can work well with very short EEG segments in a simple and practical way. Even though each input window is only  $16 \times 1$ , the model still outperforms established lightweight baselines like EEGNet and DeepConvNet under the same training setup. This suggests that our specific combination of temporal convolutions, channel-wise residual attention, and a small bidirectional LSTM is a sensible and efficient way to capture seizure-related patterns without relying on very deep or heavy architectures.

At the same time, the model remains easy to deploy. With roughly 0.6 million parameters and millisecond-level inference on a standard CPU, PaperNet is suitable for real-time or resource-limited environments, such as mobile or embedded devices. The learned attention weights also give a simple, intuitive view of which electrodes matter most, offering a small but useful step toward interpretability. However, because the dataset CSV format does not include subject identifiers or recording boundaries, our evaluation is limited to sample-level generalization. Future work should therefore test PaperNet on datasets with subject-level splits and more diverse clinical conditions to better understand its robustness in real-world scenarios.

## 5 Conclusion

In this study, we presented PaperNet, a lightweight EEG classifier designed for very short input windows. By combining temporal convolutions along the channel axis, a residual squeeze-and-excitation block, and a compact bidirectional



**Fig. 4.** Training and validation accuracy / F1 across epochs.

LSTM with global pooling, the model achieves higher accuracy and macro-F1 than EEGNet and DeepConvNet on the BEED epilepsy dataset, while using substantially fewer parameters. The learned attention weights also offer a straightforward way to understand which electrodes influence the model’s decisions, adding a layer of interpretability that many deep learning approaches lack.

PaperNet’s compact size and low inference latency make it suitable for real-time or resource-constrained settings, such as portable monitoring devices. At the same time, the use of a CSV-based dataset without subject identifiers limits our ability to claim strong cross-subject generalization. Future work will extend this architecture to richer, subject-aware EEG datasets and broader clinical scenarios. Overall, our findings suggest that carefully designed, attention-enhanced lightweight models can provide a practical path toward accurate, interpretable, and deployable EEG-based seizure detection.

## References

1. M. Teplan, “Fundamentals of EEG measurement,” *Measurement Science Review*, vol. 2, no. 2, pp. 1–11, 2002. [Online]. Available: <https://www.measurement.sk/2002/S2/Teplan.pdf>
2. H. Peng, L. He, and J. Zhang, “EEG-based emotion recognition using spectral features,” *Neurocomputing*, vol. 219, pp. 63–68, 2017. doi: <https://doi.org/10.1016/j.neucom.2016.09.017>
3. V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces,” *J. Neural Eng.*, vol. 15, no. 5, p. 056013, 2018. doi: <https://doi.org/10.1088/1741-2552/aace8c>

4. Y. Zhang, P. Zhou, and Z. Wang, "Hybrid deep learning for EEG motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2773–2784, 2020. doi: <https://doi.org/10.1109/TNSRE.2020.3037340>
5. S. Roy, S. Chowdhury, and T. H. Falk, "Deep learning-based real-time EEG–BCI: Recent advances and future directions," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 478–491, 2021. doi: <https://doi.org/10.1109/TNSRE.2021.3059496>
6. M. G. Subasi, "Automatic recognition of alertness level from EEG by using multivariate autoregressive modeling," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 8560–8565, 2009. doi: <https://doi.org/10.1016/j.eswa.2008.10.063>
7. P. K. Saha and M. A. Rahman, "Wavelet-based feature extraction for EEG classification," in *Proc. IEEE Int. Conf. Informatics, Electronics and Vision (ICIEV)*, 2015, pp. 1–6. doi: <https://doi.org/10.1109/ICIEV.2015.7334030>
8. X. Li, Z. Cui, and H. Wang, "EEG classification using decision tree ensembles," *Int. J. Neural Syst.*, vol. 26, no. 3, p. 1650005, 2016. doi: <https://doi.org/10.1142/S0129065716500057>
9. R. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.*, vol. 38, no. 11, pp. 5391–5420, 2017. doi: <https://doi.org/10.1002/hbm.23730>
10. Y. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," in *Proc. Int. Conf. Learning Representations (ICLR) Workshop*, 2016. [Online]. Available: <https://arxiv.org/abs/1511.06448>
11. H. Dai, C. Zhang, and S. He, "A hybrid CNN–RNN framework for EEG classification," *IEEE Access*, vol. 7, pp. 30720–30730, 2019. doi: <https://doi.org/10.1109/ACCESS.2019.2903085>
12. X. Li, Y. Zhang, and B. Zhang, "EEG classification using channel-wise attention," *Pattern Recognit. Lett.*, vol. 140, pp. 78–84, 2020. doi: <https://doi.org/10.1016/j.patrec.2020.09.005>
13. M. Tao, C. Jiang, and H. Wang, "EEG-based emotion recognition via channel–time attention," *IEEE Trans. Affect. Comput.*, early access, 2022. doi: <https://doi.org/10.1109/TAFFC.2022.3152849>
14. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7132–7141. doi: <https://doi.org/10.1109/CVPR.2018.00745>
15. N. N. and P. K. Banu, "BEED: Bangalore EEG Epilepsy Dataset," *UCI Machine Learning Repository*, 2024. [Online]. Available: <https://doi.org/10.24432/C5K33B>
16. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: <https://www.deeplearningbook.org/>
17. D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. doi: <https://doi.org/10.48550/arXiv.1412.6980>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

