

Systematic Evaluation of Deep Learning Architectures for Epilepsy Detection with Patient-Based Validation

Mustafa Erginli*

¹ Simav Vocational School, Kutahya Dumlupınar University,

*(mustafa.erginli@dpu.edu.tr)

(Received: 19 October 2025, Accepted: 21 October 2025)

(5th International Conference on Trends in Advanced Research ICTAR 2025, October 16-17, 2025)

ATIF/REFERENCE: Erginli, M. (2025). Systematic Evaluation of Deep Learning Architectures for Epilepsy Detection with Patient-Based Validation, *International Journal of Advanced Natural Sciences and Engineering Researches*, 9(10), 347-359.

Abstract – EEG signal processing is crucial for diagnosing neurological diseases, such as epilepsy. This study systematically compares the performance of time-domain and time-frequency domain representations in detecting epilepsy seizures. The data were divided into four-second windows at a sampling frequency of 256 Hz, and data leakage was prevented by applying a patient-based discrimination protocol. The training, validation, and test sets were separated by patient IDs, and windows from the same patient were not included in more than one partition. Six deep learning models, including BiLSTM, 1D-CNN, STFT-CNN, STFT-CNN with attention mechanism, ResSTFT-CNN, and ResSTFT-CNN with attention mechanism, were evaluated under the same training conditions. AdamW optimization, label smoothing, and early stopping algorithms were applied to all models. The models were compared in terms of accuracy, F1-Macro, AUROC, AUPRC, calibration metrics, and inference rate. Results showed that the 1D-CNN model exhibited significant superiority. The model offered the most balanced performance with an accuracy of 0.957, an AUROC of 0.993, and an inference time per sample of 0.09 milliseconds. BiLSTM came in second with 0.936 accuracy. For STFT-based models, the addition of an attention mechanism significantly improved performance, with STFT-CNN accuracy increasing from 0.78 to 0.91. Model complexity analyses revealed that 1D-CNN strikes a balance between high performance and a low parameter count. Working in the time-domain representation, 1D-CNN offers an optimal solution for epilepsy detection, striking a balance between accuracy, speed, and clinical applicability. The patient-based discrimination strategy allows for the realistic generalization capacity of the models to be evaluated.

Keywords – EEG, Epileptic Seizure Detection, Deep Learning, 1D-CNN, Attention Mechanism, Model Calibration, STFT Representation

I. INTRODUCTION

Electroencephalogram (EEG) signal processing plays a critical role in the diagnosis and monitoring of neurological diseases, such as epilepsy. EEG signals are multichannel biological data that record the brain's electrical activity in a time-dependent manner. These signals are difficult to analyze directly due to their high dimensionality, low signal-to-noise ratio, and nonlinear structure. Especially in diseases such as epilepsy, the distinction between pre-ictal, ictal, and interictal periods requires capturing the complex

time and frequency variations of the signal. In Almurshedi's study, it was emphasized that time-dependent frequency changes in these different phases of EEG signals are significant and play a critical role in diagnostic success [1].

Time-frequency analysis enables the simultaneous analysis of both temporal and frequency components of EEG signals. One of the most commonly used methods in this analysis is the Short-Time Fourier Transform (STFT). The STFT takes the Fourier transform of the signal at short time intervals, revealing how the frequency content of the signal changes over time. Che Wan Fadzal et al. demonstrated that STFT is effective in distinguishing between writing and resting states [2]. Similarly, Zabidi et al. reported that STFT can successfully distinguish between real and imaginary writing states at the frequency level [3]. This method has been widely used in the diagnosis of epilepsy. Almurshedi clearly observed the change in frequency components during different phases of epileptic seizures using STFT [1]. Mandhouj et al. showed that spectrograms obtained with STFT provide high accuracy when given as input to 2D-CNN models [4].

In the field of EEG signal processing, deep learning-based approaches have made significant progress in automatic feature extraction and classification. 1D-CNN (one-dimensional convolutional neural network) is an effective method for direct processing of time-series EEG signals. Nagabushanam et al. demonstrated that a 1D-CNN architecture is suitable and efficient for processing time series EEG signals [5]. Flores-Palermo et al. reported that the combination of 1D-CNN and LSTM yielded significant improvements in sensitivity, specificity, accuracy, and F1 score for epilepsy classification in pediatric patients using multichannel EEG data [6].

2D-CNN (two-dimensional convolutional neural network) approaches are prominent in processing two-dimensional representations obtained through time-frequency transformations. Mandhouj et al. provided spectrograms obtained from EEG signals as input to a 2D-CNN and achieved an accuracy of 98.22% in epilepsy detection [4]. Similarly, Chaudhary and Tyagi demonstrated that processing EEG scalograms with 2D-CNN can separate attention and meditation states with 87.48% accuracy [7]. These findings show that time-frequency analysis can achieve high success with 2D deep learning architectures.

With the evolution of deep learning models, residual connectionist networks, such as ResNet, have become increasingly common in EEG classification. ResNet was developed to mitigate the problem of gradient fading as layer depth increases. Liu and Zeng achieved 97.86% accuracy in classifying motor imagery EEG signals by applying features derived from Morlet wavelets, Stockwell transforms, and multi-window transforms to three separate ResNet networks. They reported a 39.65% improvement over classical methods [8].

BiLSTM (Bidirectional Long Short-Term Memory) networks have gained prominence due to their ability to learn long-term dependencies in time series, such as EEG. BiLSTM can process data in both forward and backward directions, simultaneously evaluating the past and future contexts of the signal. In Ghosh and Dey's study, a combination of CNN-BiLSTM-Attention was proposed for EEG seizure prediction, achieving an accuracy of 99.70% [9]. These results show that BiLSTM is a powerful complement for modeling temporal dependencies.

Metrics such as accuracy, sensitivity, specificity, precision, F1 score, and area under the ROC curve (AUC) are widely used in performance evaluations. Flores-Palermo et al. reported 94.05% sensitivity, 85.90% specificity, 90.12% accuracy, and 90.79% F1 score for a 1D-CNN+LSTM model trained with multichannel EEG data [6]. Ru et al. achieved high success in all of these metrics in a multi-head self-attention-based model [10].

Many studies create training and test sets by randomizing or stratifying the dataset. This approach can lead to data leakage when the same patient's data is included in both the training and test sets. In real clinical scenarios, the model should be tested on data from patients it has not seen before. Studies with patient-specific data partitioning are limited, and this leads to an unrealistic assessment of the generalization ability of the models.

In recent years, studies using deep learning architectures, such as BiLSTM, 1D-CNN, 2D-CNN, and ResNet, have increased in number, either individually or in comparison, in EEG classification. While STFT representations obtained by time-frequency analysis are particularly successful in 2D-CNN-based

models, 1D-CNN structures provide strong results with low computational cost on raw time series [11], [12], [13].

This study compares time-domain (1D) and STFT-based time-frequency domain (2D) representations on the same pipeline in the context of epilepsy seizure detection. BiLSTM, 1D-CNN, plain 2D-CNN, and ResNet-based 2D-CNN, as well as attention-added variants, are systematically evaluated. The dataset is in a combined CSV format generated from CHB-MIT EEG. The signals are divided into 1024 sample windows of 4 seconds at a sampling frequency of 256 Hz and labeled with a majority rule. In the study, the windows were organized under 2622 patient IDs, and patient-based discrimination was applied. For the 1D track, channel-wise z-score normalization with training statistics was employed. For the 2D track, within-sample STFT normalization was utilized. AMP (Automatic Mixed Precision) and early stopping protocols were applied in all models.

This study compares time-domain (1D) and STFT-based time-frequency domain (2D) representations on the same pipeline in the context of epilepsy seizure detection. BiLSTM, 1D-CNN, plain 2D-CNN, and ResNet-based 2D-CNN, as well as attention-added variants, are systematically evaluated. The dataset is in a combined CSV format generated from CHB-MIT EEG. The signals are divided into 1024 sample windows of 4 seconds at a sampling frequency of 256 Hz and labeled with a majority rule. In the study, the windows were organized under 2622 patient IDs, and patient-based discrimination was applied. For the 1D track, we employed channel-wise z-score normalization using training statistics, and for the 2D track, we utilized within-sample STFT normalization. The training process was conducted using Automatic Mixed Precision (AMP), which combines 32-bit and 16-bit floating-point operations to enhance computational speed and reduce memory usage. The experimental results showed the significant superiority of 1D-CNN under this setup. Applying the test data to the model yielded the best results, with an Accuracy of 0.957, an F1 score of 0.957, an AUROC of 0.993, an AUPRC of 0.993, and an inference time of 0.09 ms/sample. BiLSTM ranked second. The 2D-CNN and ResNet-based variants maintained an accuracy range of 78-85% and demonstrated a disadvantage in calibration metrics compared to the 1D line. These findings suggest that one-dimensional convolutions on the same pipeline can adequately capture epileptic patterns without requiring an additional transformation cost and are therefore better suited to real-time inference conditions. [13], [14], [15].

This paper presents a comprehensive analysis of epilepsy seizure detection by comparing six different deep learning models, including both time-domain and time-frequency domain representations, under the same conditions. Patient-based separation is applied to avoid data leakage, and the generalization capabilities of the models are realistically evaluated. In addition, metrics such as accuracy, calibration, and inference speed are considered together to examine the trade-off between performance and efficiency. In these aspects, the study makes an important contribution in terms of methodological integrity and clinical applicability.

II. MATERIALS AND METHOD

A. Dataset Description

CHB-MIT scalp EEG data were used in this study. The source of the data is the CHB-MIT Scalp EEG Database, version 1.0.0 [16], published on the PhysioNet platform. This dataset contains multi-day EEG recordings collected from children with pharmacoresistant epilepsy at Boston Children's Hospital. The dataset contains a total of 22 patients, 182 epileptic seizure annotations, and start-end time labels for each seizure. The clinical underpinning of the data is based on Ali Shoeb's MIT PhD Thesis titled "Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment" [17]. PhysioNet's infrastructure comprises PhysioBank, PhysioToolkit, and PhysioNet components, providing open research resources for complex physiological signals [18]. All data disclosures are publicly available on the PhysioNet platform.

For model training and evaluation, we utilized a derived file, the Preprocessed CHB-MIT Scalp EEG Database, which was published on IEEE Dataport. This file provides a combined, preprocessed version of the CHB-MIT data in CSV format, containing 24 columns and a total of 2,097,150 rows. Twenty-three of the columns correspond to EEG channels, and one column corresponds to a binary label column. The

sampling frequency is 256 Hz. Dataset description, access conditions, and citation information are publicly available on IEEE Dataport and PhysioNet[19].

B. Preprocessing

The raw EEG signals underwent various preprocessing steps before being transformed into model input. The signals were divided into 4-second windows, with each window containing 1024 samples, and the shift amount was set to 1024 samples per window. The labels of the windows were generated by a majority rule based on the majority of the samples within each window. The dataset yielded 1024 negative and 1023 positive windows.

To generalize the models, patient-based data separation was applied. Windows for each patient were only included in the training, validation, or testing part. During data transfer, each window was assigned an automatic patient identification number based on the data source and sequence information. A total of 2,622 patient IDs were generated, comprising 1,304 for training, 231 for validation, and 512 for testing. In this way, windows belonging to the same patient were prevented from being mixed into more than one partition, thereby preventing data leakage and providing a reliable basis for evaluating outpatient generalization.

In the time-domain pipeline, channel-head z-score normalization was applied with training statistics. In the time-frequency pipeline, STFT was calculated for each channel. The STFT parameters were set as follows: number of points in the fast Fourier transform (nfft) = 512, window length = 256, shift = 64 samples, and upper frequency limit = 128 Hz. The power spectrum obtained after STFT was logarithmically transformed and normalized using the mean and standard deviation of each sample to compensate for amplitude differences and to facilitate learning. This process yielded two-dimensional tensors along the channel, frequency, and time axes for each window.

In the training process, low-level data augmentation was applied on the 1D-CNN pipeline, a one-dimensional convolutional neural network. In this context, low-amplitude white noise was randomly added to the signals, and small shifts were made on the time axis. These operations were performed with a probability value of 0.1 for each sample. These operations were preferred to prevent overfitting of the model in noise-sensitive time series such as EEG. In the 2D-CNN pipeline, which is a two-dimensional convolutional neural network, data augmentation was disabled. This is because STFT-based time-frequency representations already contain enough variation, and additional distortions can negatively affect model performance. The labels used in the training were adjusted by applying a smoothing factor of 0.05 to prevent the model from producing overconfident predictions. The training process was conducted entirely on real data.

In the time-domain pipeline, channel-head z-score normalization was applied using the training statistics. On the time-frequency line, the STFT was calculated for each channel, with nfft = 512, window length = 256, offset = 64 samples, and an upper frequency limit of 128 Hz. The power spectrum was log-transformed and normalized using the within-sample mean and standard deviation to obtain 2D tensors on the channel, frequency, and time axes for each window.

C. Model Architectures

In this study, a total of six deep learning models for epileptic seizure detection working on both time-domain (1D) and frequency-time (2D) representations are developed and systematically compared. Each model aims to capture the unique dynamics of EEG signals, with LSTM-based models emphasizing temporal dependencies and CNN-based models emphasizing spatial and frequency patterns. All models are trained under the same training protocol, loss function, and optimization strategy.

The BiLSTM model processes the temporal succession of EEG signals in a bidirectional manner. Each EEG window is given to the model as a tensor with 23 channels and 1024 sample lengths (4 seconds $\times$ 256 Hz). The model contains two layers of bidirectional LSTM blocks, each with 128 hidden neurons. The bidirectional structure provides access to information at both past and future time steps, allowing for the capture of transient yet distinct patterns, such as seizure onset. The output layer employs temporal average pooling, followed by a fully connected classifier (a two-layer MLP with ReLU activation).

BiLSTM was chosen as a strong base model, particularly because it accurately represents the temporal dynamics of signals.

The 1D-CNN learns the temporal patterns of EEG signals through a filtering approach. The network consists of three convolutional blocks, each with 64, 128, and 256 filters and kernel sizes of 7, 5, and 3, respectively. Each block is followed by batch normalization, ReLU, and dropout layers. The width multiplier of the model is automatically mapped to the number of parameters of the 2D CNN, which provides a fair comparison in terms of capacity. After adaptive mean pooling, the fully connected layer separates the signal into two classes. This model gave the most balanced results in terms of efficiency, with a high accuracy of 0.957 and a very low inference time of 0.09 ms/sample.

STFT-CNN takes the short-time Fourier transform of each EEG channel and processes the resulting 2D time-frequency maps. Each input sample consists of as many frequency-time images as there are channels. The model consists of four convolution blocks. Layers with 32, 64, 128, and 256 filters learn the spectral and temporal patterns, respectively. ReLU activation, dropout, and 2×2 pooling are applied in each block. This structure is effective in capturing energy distributions in the EEG spectrum; however, the results are not as accurate as those obtained from 1D models. This difference is because STFT representations capture some transient features with a short window length. The STFT-CNN + Attention model is built by adding the Multi-Head Self-Attention (MHSA) module after the classical convolution blocks. This structure converts the spectral-temporal map at the output of the CNN layers into token sequences and learns the correlations between different regions with the eight-head attention mechanism. This allows the model to selectively focus on energy clusters or frequency bands that become salient during seizure moments. Layer normalization and dropout help mitigate the risk of overfitting in the attention mechanism. The results show that the attention extension improves the accuracy from 78% to 90% and makes the model more stable.

The ResSTFT-CNN (Residual STFT CNN) model is a deeper CNN architecture that operates on the STFT representation. The structure uses ResNet-like residual blocks. This facilitates gradient flow, allowing deeper layers to be trained. The four residual blocks contain 32, 64, 128, and 256 channels, respectively. Each block consists of two 3×3 convolutions and a shortcut connection. The network results in a fully connected layer with 128 neurons after adaptive mean pooling. Although this structure increased the number of parameters, it did not provide the expected increase in accuracy. The ResSTFT-CNN + Attention model is now a version of a block CNN with an MHSA layer added. The aim is to capture contextual interaction between different spatial regions in multichannel EEG spectrograms. In particular, it contributes to the detection of synchronous patterns occurring on the frequency axis (e.g., low-frequency spikes before seizures). The model achieved significantly higher accuracy and G-Mean than the baseline ResSTFT-CNN. This improvement suggests that the attentional mechanism can utilize the information density in 2D representations more efficiently.

D. Training Setup

Model training was conducted with a carefully balanced uniform training protocol to ensure that performance differences between different architectures are not only due to learning capacity, but also to structural aspects such as representation style and attention mechanism. The entire training process was conducted using PyTorch, and the learning parameters, regularization coefficients, and early stopping strategies were determined with a focus on achieving balance and generalization. One of the most critical components of the study is the subject-wise split of the data into three parts. Thus, the windows obtained from a patient's EEG recordings were included in only one subset (train, validation, or test), preventing data leakage between different patients. The training, validation, and test rates were set to 60%, 15% and 25% respectively. This approach prevented the model from memorizing patient-dependent patterns, providing a level of generalization closer to the clinical setting. All models were trained with the AdamW optimization algorithm due to its advantage of high parameter stability and fast convergence. The learning rate was set to 3×10^{-4} , and the weight decay to 5×10^{-4} . The learning rate was dynamically adjusted using a plateau-based reduction planner, which halved when the verification loss remained constant for a specified period. This mechanism reduced the overall optimization time by avoiding

unnecessary updates at points where training becomes stationary. The label-smoothing cross-entropy loss function was used, with a smoothing coefficient of 0.05. Thus, by taking into account the uncertainties inherent in EEG data, the model is prevented from producing overly sharp decisions, and the prediction probabilities become more balanced and reliable. Throughout the training cycle, validation accuracy was used as a monitoring metric. If the validation accuracy did not improve by the patience threshold, the training was automatically terminated. The training process was carried out on the GPU with AMP support. This technique utilizes both 32-bit and 16-bit floating-point operations to reduce memory consumption and increase computational speed. Especially in 2D-CNN models, this method accelerated the training. Data augmentation was used to a limited extent to create diversity without disturbing the structure of the EEG signals. Training was conducted on an NVIDIA GPU, and all models were initialized with the same random seed of 42. For parameter equality, the width coefficient of the 1D-CNN was automatically adjusted according to the total number of parameters of the ResSTFT-CNN. The learning capacity of each model was set to be equal to that of the others. Thus, the differences in performance between the models are not due to a model being larger or more powerful, but to differences in the data representation and architectural design used.

E. Evaluation Metrics

In this study, evaluation was conducted using both threshold-dependent classification metrics and threshold-independent discrimination power and calibration metrics, while also monitoring computational efficiency. Accuracy, F1-macro, AUROC, AUPRC, MCC, G-Mean, ECE, Brier, number of parameters, training time, and inference time per sample were used as evaluation metrics.

III. RESULTS

Table 1 summarizes the accuracy, F1-Macro, AUROC, AUPRC, G-Mean, and MCC metrics of the six models on the test set. The table shows that the 1D-CNN model is superior in almost all metrics, while BiLSTM gives similarly high but slightly slower results. In 2D-STFT-based models, the attention effect significantly improves the performance.

Table 1. Model Performance Comparison

Model	Accuracy	F1	AUROC	AUPRC	MCC	G-Mean	ECE	Brier	Param (M)	Infer (ms)
BiLSTM	0.9355	0.9355	0.9835	0.9856	0.8710	0.9355	0.0198	0.0953	0.59	0.73
1D-CNN	0.9570	0.9569	0.9926	0.9929	0.9162	0.9562	0.0132	0.0732	0.71	0.09
STFT-CNN	0.7852	0.7789	0.9506	0.9483	0.6203	0.7895	0.1095	0.3218	0.43	0.04
STFT-CNN-Attention	0.9082	0.9080	0.9775	0.9779	0.8246	0.9098	0.0486	0.1507	0.49	0.04
ResSTFT-CNN	0.7812	0.7743	0.9471	0.9300	0.6164	0.7858	0.1120	0.3394	2.82	0.16
ResSTFT-CNN-Attention	0.8496	0.8484	0.9512	0.9521	0.7185	0.8521	0.0556	0.2285	3.08	0.18

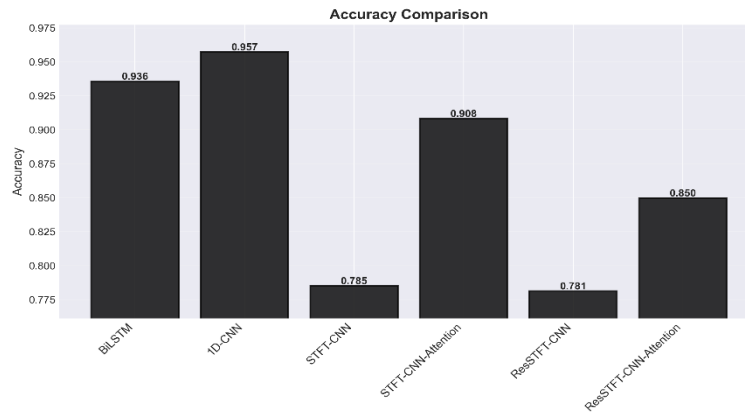


Fig. 1. Model Accuracy Overall Classification Performance Comparison

Fig. 1 shows the accuracy of the models in epilepsy detection. The 1D-CNN model has the highest accuracy with 95.7%. BiLSTM ranked second with 93.6%. STFT-CNN performed poorly, but its accuracy increased to 90.8% when attention was added. The ResSTFT-CNN model similarly improved from 78.1 percent to 85 percent. Overall, the 1D-CNN was the most balanced and robust model, while the addition of attention led to a significant improvement in time-frequency-based models.

As seen in Fig. 2, the F1-Macro curve maintains the same trend. BiLSTM and 1D-CNN stand out in balanced class performance.

Fig. 3 and Fig. 4 show the AUROC and AUPRC metrics, respectively, where the discrimination power of 1D-CNN reaches its highest value, with AUROC and AUPRC both at 0.993. AUROC visually supports the threshold-independent discrimination power, showing that 1D-CNN can discriminate more sharply than other models. AUPRC reflects the sensitivity-accuracy balance of the models in positive class detection, demonstrating that STFT-CNN-Attention, in particular, provides a significant improvement.

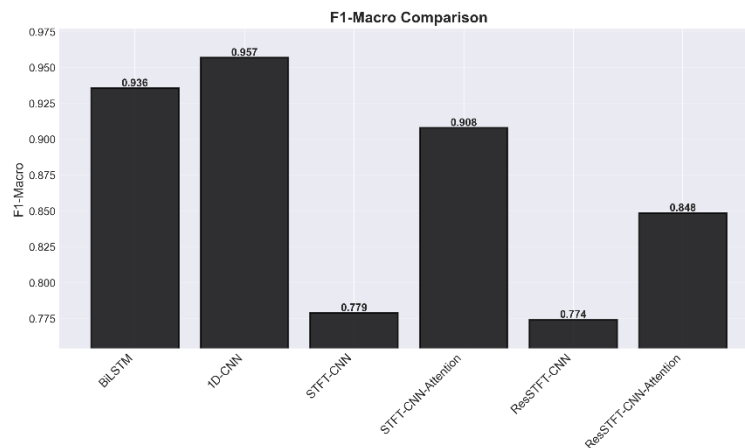


Fig. 2 F1-Macro score performance distribution

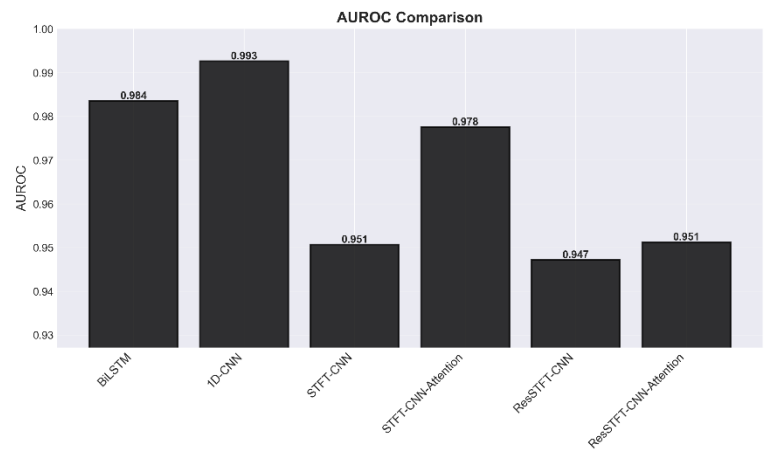


Fig. 1 AUROC comparison

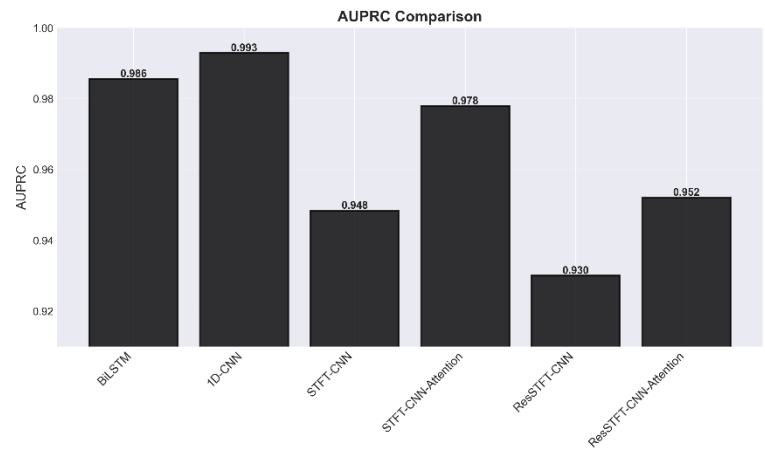


Fig. 4. AUPRC graph

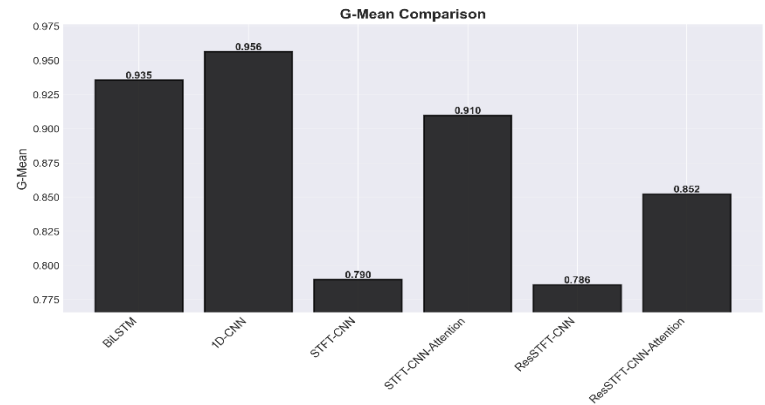


Fig. 5. Geometric Mean (G-Mean) performance graph

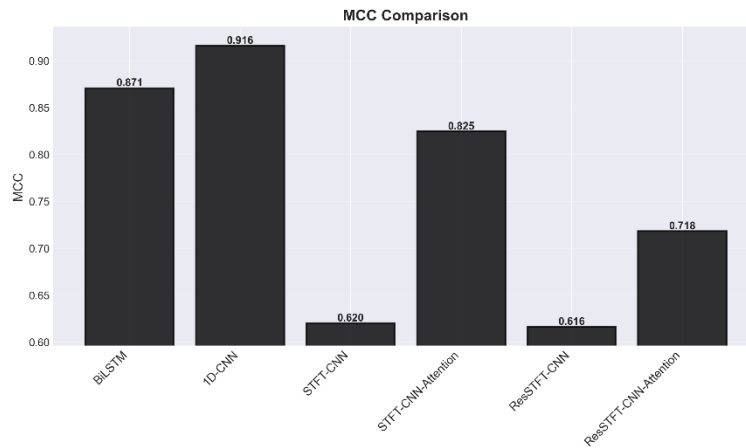


Fig. 6 Matthews Correlation Coefficient (MCC) graph

The G-Mean plots in Fig. 5 and the MCC plots in Fig. 6 show that the addition of attention improves the false positive-negative balance. In particular, the MCC value of STFT-CNN increased from 0.620 to 0.825. This suggests that the attention mechanism can capture the relationships between temporal-frequency features more efficiently. G-Mean emphasizes that the addition of attention enhances sensitivity in both classes and improves overall performance on imbalanced data. MCC visually indicates that the MCC of STFT-CNN increases from 0.62 to 0.82 with the addition of attention. This increase indicates that the error profile has stabilized.

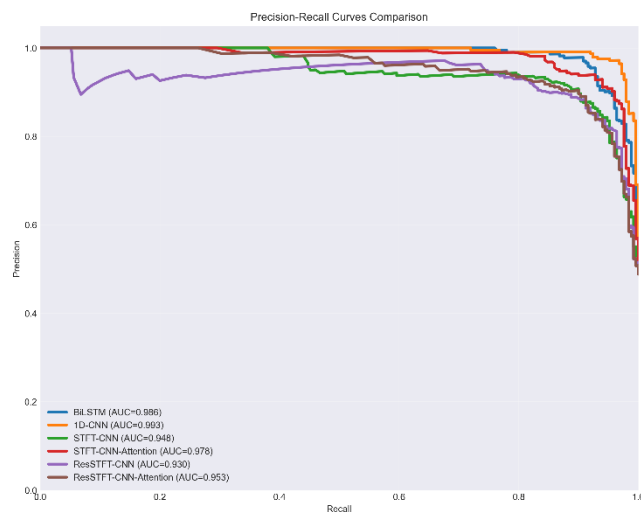


Fig. 7 Precision-recall curves comparison

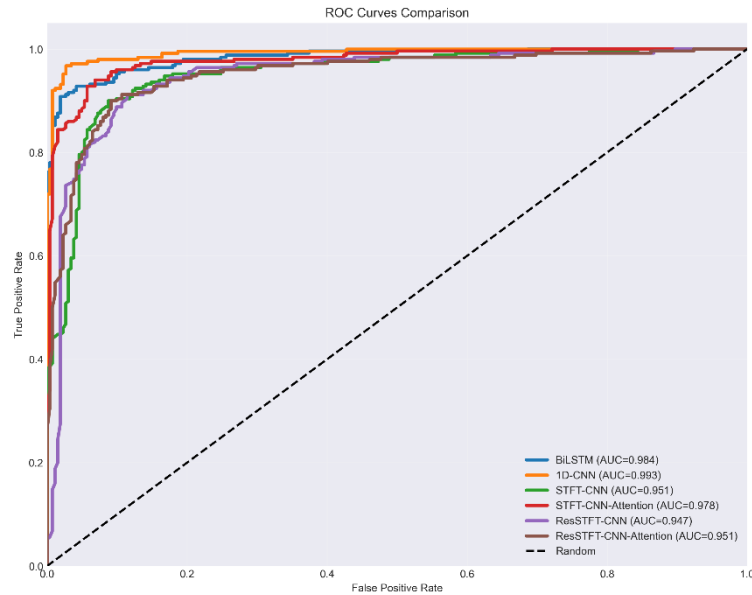


Fig. 2 ROC curves comparison

The areas under AUPRC in Fig. 7 show that 1D-CNN and BiLSTM maintain their high sensitivity even at low false alarm tolerance. STFT-CNN-Attention gained a significant slope compared to STFT-CNN, while ResSTFT variants showed a moderate improvement. These results support that the attention module highlights important temporal patterns in noisy EEG segments.

Figure 8 shows the ROC curves in a comparative format. The curves illustrate the change in the true positive rate (TPR) of each model as the false positive rate (FPR) increases. The 1D-CNN curve at the top shows good discrimination power with a value of 0.993. The BiLSTM model ranked second with an AUC of 0.984, while the STFT-CNN-Attention model performed remarkably well with 0.978. The pure STFT-CNN and ResSTFT-CNN curves saturate earlier, indicating more overlap in the decision boundaries of these models.

Taken together, AUPRC and ROC analyses confirm that the 1D-CNN is most consistently able to discriminate epileptic patterns. Attention-added 2D-STFT models showed significant gains, especially in low false alarm regions. This demonstrated that the attention mechanism can effectively highlight critical information regions in complex EEG signals. Not only is the accuracy of deep learning models important, but also the balance of computational load and speed.

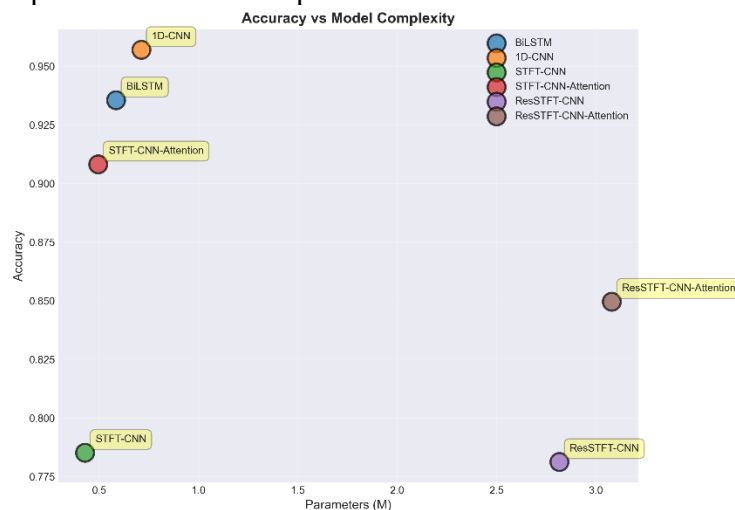


Fig. 9 Model complexity and performance relationship graph

Fig. 9 shows the relationship between accuracy and model complexity (number of parameters). 1D-CNN achieves the highest accuracy with around 0.7 million parameters, while ResSTFT-CNN-Attention achieves lower accuracy with over three million parameters. It demonstrates the impact of model depth on efficiency, showing that 1D-CNN achieves high performance with a minimal number of parameters.

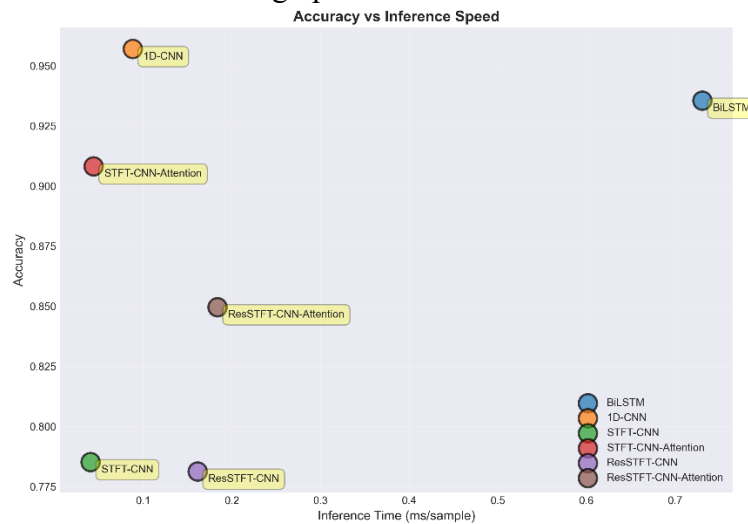


Fig. 3 Accuracy vs Inference Speed

Figure 10 presents the relationship between accuracy and inference time. 1D-CNN demonstrates a clear advantage in terms of efficiency, thanks to both its high accuracy and low inference time of 0.09 ms/sample. BiLSTM achieves similar accuracy, but the inference time is about eight times longer. This difference makes 1D-CNN more suitable for embedded systems or real-time epilepsy detection applications.

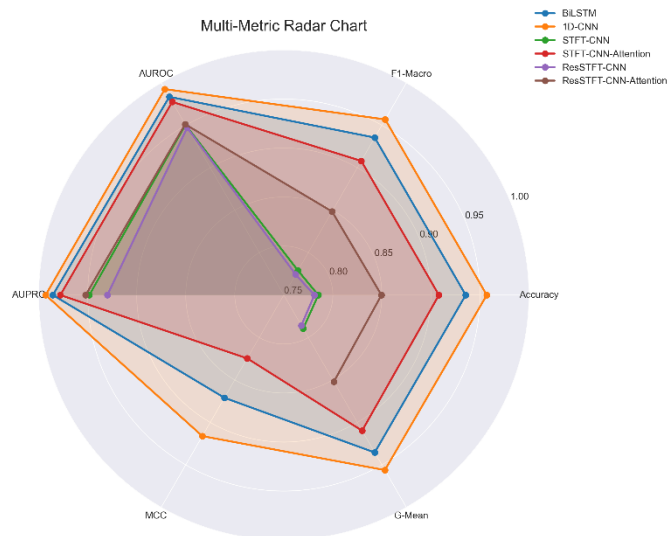


Fig. 4. Six-Metric Radar Chart Comparison

Fig. 11 is a multi-metric radar plot that displays the six metrics in the same plane. The graph visually summarizes the overall superiority of 1D-CNN, with the STFT-CNN-Attention model coming second in terms of multi-metric balance. BiLSTM's balanced performance is also noteworthy, but it is not as efficient as 1D-CNN in terms of parameter-speed cost.

All metrics indicate that 2D-STFT models with attention mechanisms enhance interpretability; however, in the efficiency-accuracy trade-off, 1D-CNN is the optimal solution.

IV. DISCUSSION

The 1D-CNN model demonstrated high sensitivity in capturing short-term preictal and interictal transitions, as it can directly learn the temporal succession of EEG signals. Accuracy and F1-Macro results indicate that this architecture outperforms other models in both class balance and overall performance. Moreover, the AUROC and AUPRC values confirm that the model has high discrimination power, regardless of the threshold. This performance is due to its ability to learn neural representations of 1D convolutions directly from the waveform.

Although STFT-based models provide information in the frequency dimension, they tend to propagate some temporal patterns due to the short windowing time. This is evident in the G-Mean and MCC metrics. The frequency representation especially blurs the localization of transient epileptic bursts.

The addition of Attention led to a significant performance jump in 2D-STFT representations. Precision-Recall curves reveal that the Attention mechanism leads to a significant increase in precision in the preictal segments representing the positive class. As shown in the radar plot, the STFT-CNN-Attention model outperformed the 2D-STFT model in the multi-metric balance. This gain is related to the fact that the attention module emphasizes critical regions in the time-frequency maps, thereby reducing the noise effect and increasing the model's generalization power. The ResSTFT-CNN and ResSTFT-CNN-Attention models, although theoretically capable of exploiting depth, were parameter overloaded due to their small window size of 256 samples. Fig. 9 shows that the marginal accuracy gain is lost when the number of parameters exceeds 3 million. This limits the efficiency of deep residual networks on small EEG segments. When the ECE and Brier values are analyzed, it is observed that 1D-CNN optimally calibrates the prediction probabilities. This means that the probabilities predicted by the model are close to the actual realization rates. In other words, instead of overconfident but inaccurate predictions, a measured and reliable structure emerges. The calibration-oriented metrics in the radar graph also support this.

V. CONCLUSION

This study systematically compared six deep learning architectures under a common training and evaluation protocol for EEG-based seizure detection on CHB-MIT data while preserving patient-based discrimination. Time-domain and frequency-time representations were analyzed on the same data processing pipeline. Metrics were reported, including discrimination power, calibration, and computational efficiency. Results showed that the 1D-CNN trained with the 1D time-domain representation achieved a balanced superiority in accuracy, F1-macro, AUROC, and inference time per sample. This result suggests that learning transient EEG patterns directly from the waveform is effective in short-term windows. The addition of attentional layers in frequency-time representations produced significant gains, especially in the STFT-CNN family, thereby enhancing the ability to focus on critical time-frequency regions. However, parameterization in deep residual structures offered limited gains in small windows. Model complexity had to be weighed more carefully against the gain in accuracy. Calibration metrics revealed that 1D-CNN distributed confidence scores more consistently and produced more reliable probabilities to support clinical threshold selection. Considering the speed and model size, the 1D-CNN architecture is suitable for real-time applications.

This study clearly demonstrated the trade-off between EEG representation, architectural design, and resource cost. Two directions for future work stand out: i) 1D and 2D representations in a common backbone, ii) hybrid approaches combining temporal and spectral cues, and iii) lossless modeling of long-range addictions by moving attentional mechanisms to transducer-based structures. In addition, generalization analyses with longer-term recordings, different patient cohorts, and prospective protocols, as well as further improvement of calibration through temperature scaling and boundary adaptation methods, and threshold selection strategies that directly optimize false alarm cost, will strengthen the clinical relevance of this field.

REFERENCES

- [1] A. F. Almurshedi, "Envelope analysis and short time Fourier transform of EEG for epilepsy diagnosis enhancement," *Romanian Journal of Neurology Revista Romana De Neurologie*, 2025, doi: <https://doi.org/10.37897/RJN.2025.1.12>.
- [2] Y. K. L. C. W. N. F. C. W. F. A. Zabidi W. Mansor, "Short-time Fourier Transform analysis of EEG signal generated during imagined writing," 2012 International Conference on System Engineering and Technology (ICSET), 2012, doi: <https://doi.org/10.1109/icsengt.2012.6339284>.
- [3] M. S. Badreddine Mandhouj Mohamed Ali Cherni, "An automated classification of EEG signals based on spectrogram and CNN for epilepsy diagnosis," *Analog Integr Circuits Signal Process*, vol. 108, no. 1, pp. 101–110, 2021, doi: <https://doi.org/10.1007/s10470-021-01805-2>.
- [4] P. D. P. B. M. N. S. R. P. Nagabushanam S. Thomas George, "Artifact Removal using Elliptic Filter and Classification using 1D-CNN for EEG signals," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 551–556, 2020, doi: <https://doi.org/10.1109/icaccs48705.2020.9074287>.
- [5] W. U. Oscar Flores-Palermo Christian Espiritu-Cueva, "Development of 1D-CNN Methods for Classifying Pediatric Epilepsy Through EEG Signals," *Communications in Computer and Information Science*, pp. 401–415, 2025, doi: https://doi.org/10.1007/978-3-031-91428-7_28.
- [6] A. T. Anuradha Chaudhary, "A 2D-CNN-Scalogram based Approach to Classify the Mental Tasks using EEG Signals," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 106–112, 2022, doi: <https://doi.org/10.1109/iciccs53718.2022.9787994>.
- [7] Y. Z. Wuqiao Liu, "Motor Imagery Tasks EEG Signals Classification Using ResNet with Multi-Time-Frequency Representation," 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), pp. 2026–2029, 2022, doi: <https://doi.org/10.1109/icsp54964.2022.9778786>.
- [8] D. D. ANTARA GHOSH, "Digital Twin for EEG seizure prediction using time reassigned Multisynchrosqueezing transform-based CNN-BiLSTM-Attention mechanism model," *Biomed Phys Eng Express*, 2025, doi: <https://doi.org/10.1088/2057-1976/ad992c>.
- [9] Z. W. H. C. Yandong Ru Gaoyang An, "Epilepsy detection based on multi-head self-attention mechanism," *PLoS One*, vol. 19, no. 6, pp. e0305166–e0305166, 2024, doi: <https://doi.org/10.1371/journal.pone.0305166>.
- [10] H. W. Boyang Lu Sheng Ge, "EEG-Based Classification of Lower Limb Motor Imagery with STFT and CNN," *Communications in Computer and Information Science*, pp. 397–404, 2021, doi: https://doi.org/10.1007/978-3-030-92310-5_46.
- [11] A. C. Jiacheng Wu Yi Zhong, "Radio Modulation Classification Using STFT Spectrogram and CNN," 2021 7th International Conference on Computer and Communications (ICCC), pp. 178–182, 2021, doi: <https://doi.org/10.1109/iccc54389.2021.9674714>.
- [12] [12] R. K. M. G. Dhousha Sagga Amira Ehtioui, "Epileptic Seizure Detection using EEG Signals based on 1D-CNN Approach," 2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), pp. 51–56, 2020, doi: <https://doi.org/10.1109/sta50679.2020.9329321>.
- [13] S. S. T. K. Anurag Singh Nalini Pusrarla, "CNN-based Epilepsy detection using image like features of EEG signals," 2020 International Conference on Electrical and Electronics Engineering (ICE3), pp. 280–284, 2020, doi: <https://doi.org/10.1109/ice348803.2020.9122874>.
- [14] S. C. Tabassum Islam Toma, "A Comparative Analysis of 2D Deep CNN Models for Arrhythmia Detection Using STFT-Based Long Duration ECG Spectrogram," 2022 Thirteenth International Conference on Ubiquitous and Future Networks (ICUFN), pp. 483–488, 2022, doi: <https://doi.org/10.1109/icufn55119.2022.9829574>.
- [15] J. Guttag, "CHB-MIT Scalp EEG Database (version 1.0.0)," 2010, PhysioNet. doi: 10.13026/C2K01R.
- [16] A. Shoenb, "Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment," Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.
- [17] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000, doi: 10.1161/01.CIR.101.23.e215.
- [18] R. Saemalahr and M. Ilyas, "Patient-Specific Preictal Pattern-Aware Epileptic Seizure Prediction with Federated Learning," *Sensors*, vol. 23, no. 14, 2023, doi: 10.3390/s23146578.