Advances in Artificial Intelligence and Machine Learning; Research 5 (4) 4575-4593 Received 04-08-2025; Accepted 07-11-2025; Published online 14-11-2025

Epileptic Seizure Prediction on CHB-MIT EEG Using Soft Fusion Post-Processing of Top-K Predictions with CNN Architectures

Asmaa Mohammad Balamas abalamash0003@stu.kau.edu.sa
Faculty of Computing and Information Technology
King Abdulaziz University Jeddah,
Saudi Arabia

Ghadah Abdullah Aldabbagh galdabbagh@kau.edu.sa
Faculty of Computing and Information Technology
King Abdulaziz University Jeddah,
Saudi Arabia

Samar Mohammed Alkhuraiji salkhuraiji@kau.edu.sa
Faculty of Computing and Information Technology
King Abdulaziz University Jeddah,
Saudi Arabia

Corresponding Author: Asmaa Mohammad Balamash

Abstract

Predicting epileptic seizures requires not only accurate detection but also the presence of a reliable alarm to make the predictions clinically meaningful. In this study, we introduce a Soft Fusion approach that applies Top-K averaging to preictal probabilities as a post-processing step to stabilize and improve the model's overall predictive performance. Using the CHB-MIT scalp EEG dataset, we evaluated three models—a baseline CNN, a CNN with Convolutional Block Attention Module (CBAM), and a CNN with Long Short-Term Memory (LSTM)—with STFT spectrogram inputs under a five-minute Seizure Prediction Horizon (SPH) and a 30-minute Seizure Occurrence Period (SOP). Soft Fusion improved the sensitivity of all models (macro-average up to 92.8%) while reducing false prediction rates to about 0.8 events per hour (≈20/day), consistently outperforming the conventional k-of-n (Hard Fusion) rule. Although the results remain higher than the ideal clinical tolerance, they show that the post-processing step is as essential as the model design for seizure prediction. Although the study was limited to one dataset and used patient-specific thresholds, the Soft Fusion method is a promising step toward making seizure prediction systems more reliable and clinically practical.

Keywords: Epilepsy, Seizure prediction, Electroencephalography (EEG), Post-processing, Convolutional Neural Networks (CNN), Soft fusion, CHB-MIT

Citation: Asmaa Mohammad Balamas, et al. Epileptic Seizure Prediction on CHB-MIT EEG Using Soft Fusion Post-Processing of Top-K Predictions with CNN Architectures. **Advances in Artificial Intelligence and Machine Learning.** 2025;5(4):254. https://dx.doi.org/10.54364/AAIML.2025.54254

1. INTRODUCTION

Epilepsy is a chronic brain disorder that affects millions of people worldwide and is characterized by seizures caused by abnormal electrical activity in the brain [1]. Since seizures are unpredictable, they can severely impact daily life, restrict social activities, and increase the risk of injury [2]. Therefore, developing reliable systems to predict seizures has become a key focus of current research. Electroencephalography (EEG), a noninvasive method for recording brain activity, is widely used to study epilepsy and identify early signs of an impending seizure [3].

Seizure prediction models fall into two main categories: threshold-based and classification-based approaches [4]. Threshold-based methods detect changes in EEG activity, such as energy, frequency, or non-linearity, that cross a defined threshold and trigger a seizure alert [3]. Although these techniques are computationally efficient, threshold-based methods are sensitive to noise and may not generalize well in different patients [5]. Classification-based approaches use machine learning to distinguish between preictal and interictal states [6]. They rely on extracted EEG features and can be either patient-specific or patient-independent. Patient-specific models train on data from one person and adapt to that individual's EEG patterns. Patient-independent models use data from multiple subjects to learn general patterns across people. Patient-specific models usually achieve higher accuracy, while patient-independent models are easier to scale but less precise due to variations in EEG signals. Research shows that patient-specific models are more dependable for clinical use, though they require more data and customization [6].

Earlier work focused on traditional machine learning with hand-crafted features. These studies extracted time, frequency, or time–frequency features from EEG signals, then trained classifiers such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), or decision trees to differentiate between preictal and interictal states. Cho et al. (2017) [7], combined multivariate empirical mode decomposition with an SVM, achieving 82.44% sensitivity and 82.76% specificity on the CHB-MIT dataset. Zhang and Parhi (2016) [8], used spectral analysis with SVMs, reaching 98.68% sensitivity and a false prediction rate of 0.05 per hour. These methods performed well but required expert feature design and often failed to capture the nonlinear nature of EEG signals [3].

The emergence of deep learning has shifted the focus toward end-to-end models that automatically extract features. Convolutional Neural Networks (CNNs) now play a significant role in learning spatial features from EEG representations without manual feature engineering. Truong et al. (2018) [9], employed STFT spectrograms as CNN input, achieving 81.2% sensitivity with an FPR of 0.16/h on the CHB-MIT dataset, while Khan et al. (2018) [2], used wavelet-based tensors to CNNs and reached 87.8% sensitivity and a 0.142/h FPR. Chen et al. (2019) [10], designed a patient-specific CNN based on STFT spectrograms, yielding 91.4% sensitivity and a 0.09/h FPR on the CHB-MIT data.

Researchers have integrated Long Short-Term Memory (LSTM) networks into seizure prediction pipelines to capture the temporal dependencies present in EEG sequences. Tsiouris et al. (2018) [3], combined time–frequency and connectivity features with a two-layer LSTM, reaching 100% event-based sensitivity and an FPR ranging from 0.11/h to 0.02/h depending on the duration of the preictal window. Baghdadi et al. (2020) [11], trained LSTMs directly on raw EEG signals, achieving 84.6% sensitivity and 0.27/h FPR, demonstrating that temporal modeling can improve robustness.

Hybrid models that combine CNNs for spatial feature extraction and LSTMs for temporal modeling have also shown strong performance. Ryu and Joe (2021) [12], achieved 93.28% accuracy and a 0.063/h FPR by integrating DenseNet with LSTM on wavelet-transformed EEG signals. At the same time, Halawa et al. (2022) [13], applied DWT preprocessing with 1D-CNNs in a patient-independent setting, achieving up to 97.87% accuracy. Ibrahim et al. (2023) [14], introduced a multiresolution CNN using MODWT decomposition, obtaining 82–85% sensitivity with post-processing to lower the FPR.

Recent work has incorporated attention mechanisms and Transformers to enhance spectral and inter-channel dependency modeling. Yang et al. (2021) [15], developed RDANet, a model that uses residual connections and dual self-attention, achieving 89.33% sensitivity and 93.02% specificity. Hu et al. (2023) [16], introduced a hybrid Transformer with rhythm-specific embeddings that demonstrated 91.7% sensitivity and zero false alarms with patient-specific fine-tuning.

Recent studies on patient-independent models have focused on addressing the variation in EEG patterns across individuals. Dissanayake et al. (2021) [6], used Mel Frequency Cepstral Coefficients (MFCCs) with multi-task learning and achieved 91.54% accuracy without patient-specific calibration, making seizure prediction models more scalable for clinical use. Researchers have also explored post-processing methods with the goal of improving their clinical usefulness. Batista et al. (2024) [17], proposed chronological and cumulative firing power techniques that considered seizure temporal patterns. The cumulative technique improved sensitivity, but it also raised the false prediction rate (FPR).

Daoud and Bayoumi (2019) [18], proposed a hybrid deep learning model that combines CNNs, Bi-LSTMs, and autoencoders to process raw EEG signals. Their approach reached 99.6% accuracy with a false prediction rate of 0.004 per hour, showing the benefit of combining spatial and temporal features. Usman et al. (2020) [19], used CNN-based feature extraction with SVM classification on STFT-preprocessed EEG signals, achieving 92.7% sensitivity and a 21-minute prediction horizon. Das et al. (2020) [20], applied a handcrafted waveform detection method with 92.66% accuracy.

Despite these advances, several challenges remain, including variability in EEG patterns between patients, limited labeled data, and maintaining high accuracy with few false alarms in real-time use [2]. Future research should explore transfer learning to adapt models to new patients [16], develop stronger CNN–LSTM hybrids, integrate attention-based features, and design real-time clinical prediction systems.

TABLE 1 summarizes the state-of-the-art in seizure prediction approaches on the CHB-MIT scalp EEG dataset and lists both sensitivity and FPR.

The key contributions of this study are as follows:

- We designed and tested three models for seizure prediction: CNN, CNN–CBAM, and CNN–LSTM.
- The Short-Time Fourier Transform (STFT) was employed to extract features from scalp EEG signals.
- We proposed a Soft Fusion method using Top-K predictions as a new post-processing step.

- We evaluated the models on the CHB-MIT scalp EEG dataset and compared Soft Fusion with Hard Fusion (the k-of-n method).

Table 1: Comparison of state-of-the-art seizure prediction approaches on the CHB-MIT scalp EEG dataset. This Work reports results for CNN, CNN+LSTM, and CNN+CBAM backbones under SPH=5 min and SOP=30 min.

Author(s) & Year	Dataset	Features	Classifier	Preictal Length	Sensitivity (%)	FPR (/h)
Khan et al. (2018) [2] (2018)	CHB-MIT	Wavelet coeff.	CNN	10 min	87.8%	0.142 /h
Tsiouris et al. (2018) [3] (2018)	CHB-MIT	Stat. + Wavelet + PSD	LSTM	15–120 min	99.28–99.86%	0.11-0.02 /h
Truong et al. (2018) [9] (2018)	CHB-MIT	STFT spectral images	CNN	30 min	81.2%	0.16 /h
Chen et al. (2019) [10] (2019)	CHB-MIT	STFT	CNN	1 h	91.4%	0.09 /h
Daoud & Bayoumi (2019) [18]	CHB-MIT	Raw EEG	DCAE + Bi-LSTM	1 h	99.72%	0.004 /h
Baghdadi et al. (2020) [11]	CHB-MIT	Raw EEG	LSTM	15 min	84.6%	0.27 /h
Usman et al. (2020) [19]	CHB-MIT	STFT + CNN	SVM	21 min	92.7%	≈ 0.27 /h
Ryu & Joe (2021) [12]	CHB-MIT	DWT	DenseNet + LSTM	5 min	92.92%	0.063 /h
Ibrahim et al. (2023) [14]	CHB-MIT	MODWT	Multiresolution CNN	30 min	82%	0.058 /h
Hu et al. (2023) [16] (2023)	CHB-MIT	Wavelet Transform	Hybrid Transformer	25 min	91.7%	0.00 /h
This Work – k-of-n	CHB-MIT	STFT	CNN	30 min	66.15%	0.53 /h
			CNN+LSTM		67.44%	0.57 /h
			CNN+CBAM		85.64%	0.82 /h
This Work – Top-K	CHB-MIT	STFT	CNN	30 min	92.82%	0.82 /h
			CNN+LSTM		82.95%	0.67 /h
			CNN+CBAM		90.26%	0.80 /h

2. MATERIALS AND METHODS

This section presents the methodology used to predict epileptic seizures from scalp EEG signals. FIGURE 1 illustrates the overall workflow of the proposed framework. The process starts with acquiring raw EEG recordings and then applies several preprocessing steps, including segmentation, the short-time Fourier transform (STFT), denoising, and oversampling. The processed spectrograms are then used as input to one of three model architectures—Baseline CNN, CNN+LSTM, or CNN+CBAM—to classify each segment as preictal or interictal. Post-processing is applied using either the Soft Fusion (Top-K) or Hard Fusion (k-of-n) approach to refine the model's predictions and improve their reliability. The figure provides a high-level overview of the complete pipeline, while the following subsections explain each stage in detail—from dataset preparation and preprocessing to model architecture and evaluation strategy.

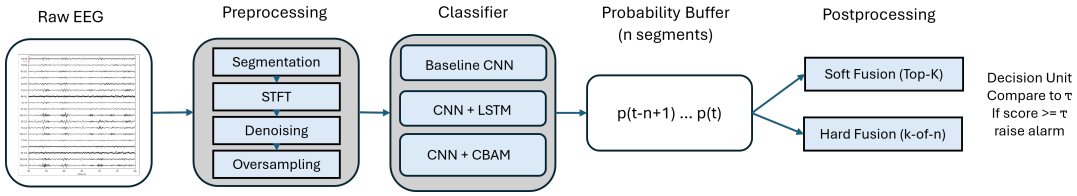


Figure 1: High-level pipeline for EEG seizure prediction

2.1 Dataset

This study used the publicly available **CHB-MIT Scalp EEG Database** [21], initially described by Shoeb (1981) [22], and hosted on **PhysioNet** (`https://physionet.org/content/chbmit/1.0.0/`) [23]. The dataset contains scalp EEG recordings from 23 pediatric patients with intractable epilepsy, collected at *Boston Children's Hospital*. In total, it provides 969 hours of EEG data and 198 annotated seizures. The dataset offers EEG recordings collected with 22 electrodes placed according to the international 10–20 system, sampled at 256 Hz with 16-bit resolution. Each subject contributed between 9 and 42 continuous EEG files in `(.edf)` format.

Following Truong et al. (2018) [9], the same preictal definition and filtering settings were applied. EEG recordings were divided into preictal and interictal categories. The preictal segments covered the 30 minutes before each seizure, with a five-minute gap before seizure onset (known as the Seizure Prediction Horizon, SPH). Interictal segments were taken at least four hours away from any seizure to represent normal brain activity. Seizures less than 30 minutes apart were merged into one episode. Patients with more than ten seizures in a single day were excluded to avoid overlapping preictal periods. Only patients with at least three distinct seizures and at least three hours of usable interictal data were included. After these criteria were applied, 13 patients remained in the study.

A key characteristic of the dataset was class imbalance. The interictal to preictal ratio varied by patient from roughly 3:1 to 23:1 (overall mean ≈6.5:1). This means that for every hour of preictal activity, there were more than six hours of seizure-free recordings.

A detailed summary of patient-wise seizure counts and interictal durations is provided in TABLE 2.

Table 2: Summary of the CHB-MIT Scalp EEG dataset (22 channels, sampling rate 256 Hz) used in this study.

Patient	No. of seizures	Interictal hours
chb01	7	17
chb02	3	22.9
chb03	6	21.9
chb05	5	13
chb09	4	12.3
chb10	6	11.1
chb13	5	14
chb14	5	5
chb18	6	23
chb19	3	24.9
chb20	5	20
chb21	4	20.9
chb23	5	3
Total	**64**	**209**

2.2 Data Preprocessing

The raw EEG data were converted to image-like inputs by computing log-scaled short-time Fourier transform (STFT) spectrograms on 30 s segments, producing time–frequency matrices suitable for CNNs. STFT used a Hann window (length 256 samples; 50% overlap; sampling rate 256 Hz) and was applied independently per channel. To suppress power-line interference common in CHB-MIT, we removed the DC bin and excluded frequency bands around the 60 Hz and 120 Hz harmonics (57–63 Hz and 117–123 Hz). No normalization or data-driven scaling was performed. The STFT spectrograms were selected, as they have been widely used in seizure prediction studies on CHB-MIT, thereby ensuring comparability with prior work.

To address the class imbalance, we applied an overlapping sampling strategy *during training only*. Additional preictal segments were generated by using a 30 s sliding window across preictal periods with a subject-specific stride. This increased the number of preictal samples to better match the interictal counts, improving learning stability and reducing bias toward the majority class (FIGURE 2). The required overlap ratio was computed per subject as

$$\text{overlap\_ratio} = 1 - \left(\frac{\text{base\_windows}}{\text{target\_windows}} \right),$$

where *base_windows* is the number of non-overlapping 30 s preictal segments, and *target_windows* is the number of interictal segments. The stride was then set to

$$\text{stride} = 30\ \text{s} \times \text{overlap\_ratio},$$

clipped to the range $[1, 30\text{s} - 1]$ to ensure valid sampling. For evaluation, test sets always used non-overlapping 30 s windows.

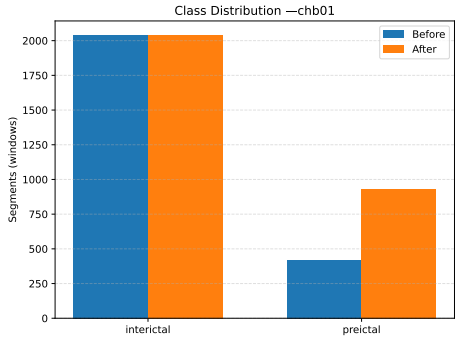
2.3 The Proposed Model

This work presents a novel *Top-K Soft Fusion* strategy as a post-processing method for seizure prediction, along with comparisons with the conventional k-of-n rule. To ensure fair benchmarking, all experiments shared a typical architecture, being progressively extended with additional modules.

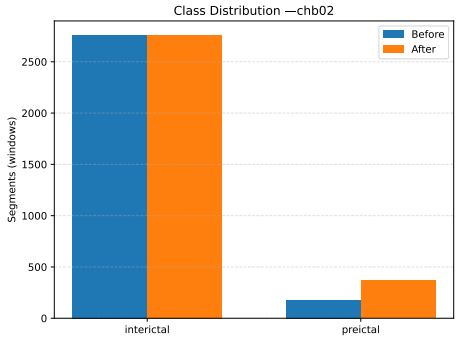
2.3.1 Baseline CNN

We adopt the patient-specific CNN proposed by Truong et al. (2018) [9], initially designed for seizure prediction on CHB-MIT. The model consists of three convolutional blocks followed by two fully connected (FC) layers. Each block applies Batch Normalization, a convolution with ReLU activation, and max pooling:

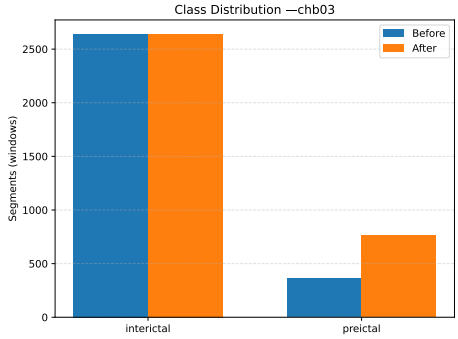
- **Block 1**: 16 kernels of size n×5×5 (n EEG channels), stride 1×2×2, max pooling 1×2×2.
- **Block 2**: 32 kernels of size 3×3, stride 1×1, max pooling 2×2.
- **Block 3**: 64 kernels of size 3×3, stride 1×1, max pooling 2×2.



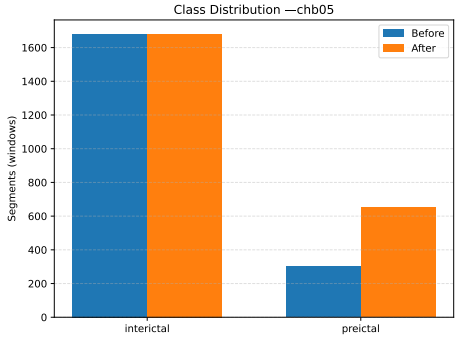
(a) Class distribution for chb01



(b) Class distribution for chb02



(c) Class distribution for chb03



(d) Class distribution for chb05

Figure 2: Class distribution of interictal and preictal segments before and after oversampling for three representative patients.

The flattened features are passed to **FC1** (256 units, sigmoid activation, dropout p=0.5) and **FC2** (two units, softmax, dropout p=0.5). Optimization used Adam with a learning rate of 5×10^{-4}, β_1=0.9, β_2=0.999, $\epsilon=10^{-8}$. This architecture was intentionally kept shallow and straightforward to reduce the risk of overfitting on the limited CHB-MIT dataset [9].

2.3.2 CNN+CBAM

To test whether attention mechanisms could improve discrimination, we added the Convolutional Block Attention Module (CBAM) [24], to our baseline CNN. In our implementation, a 3D CBAM block (channel attention followed by spatial attention) was inserted after the first 3D convolutional layer, before pooling. The channel attention used global average and max pooling with a reduction ratio of r=16, while spatial attention used a 7×7×7 convolution on concatenated channel descriptors. Attention maps modulated the feature activation via elementwise multiplication. CBAM was selected over several alternatives (SE, SAM, GAM) as it combines both channel and spatial mechanisms with negligible parameter overhead. The rest of the architecture (two additional 2D convolutional layers, pooling, and fully connected layers) was kept unchanged from the baseline CNN.

2.3.3 CNN–LSTM (one-layer BiLSTM)

We added a *single bidirectional LSTM* layer with 64 units per direction to capture temporal dependencies after CNN feature extraction. The convolutional feature maps are first flattened and then reshaped into sequences with 128 features per step, where the flattened dimension determines the sequence length. The BiLSTM output goes through dropout using the criterion p=0.5, then a dense layer with 128 sigmoid units, another dropout layer, and finally a softmax classifier. This design keeps the temporal module light in weight while extending the baseline CNN with recurrent modeling. Broader comparisons with GRU or temporal-convolution alternatives are left to future work.

All three models shared the same convolutional backbone but differed in their attention and temporal modeling strategies (FIGURE 3).

2.4 System Evaluation

Before evaluating the performance of the model for seizure prediction, it is essential to establish the key metrics and temporal parameters. Sensitivity measures the proportion of seizures that are correctly predicted, with a seizure considered as predicted if the system raises at least one alarm within the corresponding evaluation window:

$$\text{Sensitivity} = \frac{\text{Number of Predicted Seizures}}{\text{Total Number of Seizures}}.$$

The false prediction rate (FPR) quantifies the frequency of false alarms per hour of the interictal EEG:

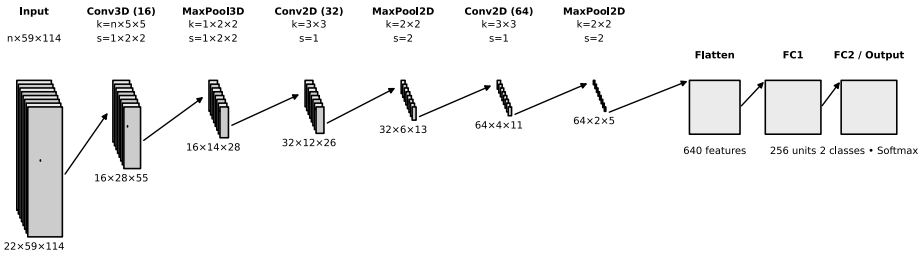
$$\text{FPR} = \frac{FP}{\text{Interictal Time (in hours)}},$$

Where FP is the total number of false alarms.

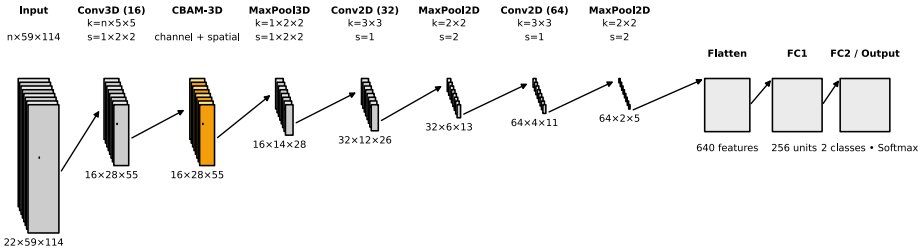
Seizure prediction further requires defining the **Seizure Prediction Horizon (SPH)** and the **Seizure Occurrence Period (SOP)** [5, 9, 25]. The SPH is defined as the minimum lead time between an alarm and the start of the SOP, where the SOP is the interval in which a seizure is expected to occur following the SPH. A prediction is considered correct if the onset of a seizure occurs *after* the SPH and *within* the SOP; a false alarm is counted when no seizure occurs within the SOP. When an alarm is triggered, it remains active until the end of the SOP.

In this study, we set the SPH to five minutes and the SOP to 30 minutes (see FIGURE 4), consistent with prior research (e.g., Truong et al. (2018) [9]). A five-minute SPH is commonly used, but longer horizons of 10–15 minutes may be more practical in clinical settings since they allow more time for intervention. Likewise, a 30-minute SOP is frequently used in research, but it may be too long for real-life applications, as it can increase patient anxiety and reduce the accuracy of seizure timing. Selecting appropriate SPH and SOP values is essential to balance clinical relevance and practical needs. However, inconsistent definitions across studies make direct comparisons challenging [9, 16, 25].

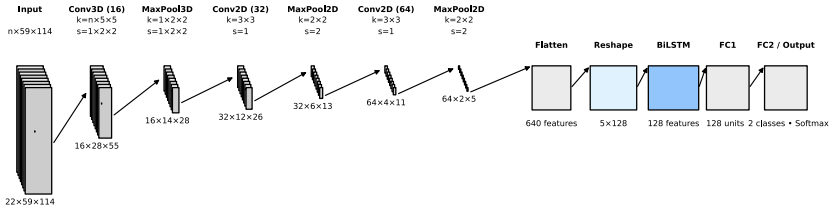
It is essential to distinguish between data segmentation and model evaluation. During data preparation, each preictal segment covered a 30-minute period ending five minutes before seizure onset.



(a) Baseline CNN (Truong et al. (2018) [9]).



(b) CNN with a 3D CBAM attention block after the first convolutional layer.



(c) CNN extended with a single Bidirectional LSTM layer.

Figure 3: Proposed patient-specific CNN architectures for seizure prediction on the CHB-MIT dataset. All models share a common convolutional backbone composed of three blocks (Conv3D → Conv2D → Conv2D), each followed by Batch Normalization, ReLU activation, and max pooling, with fully connected layers for classification. The **Baseline CNN** (a) follows the original architecture by Truong et al. (2018) [9]. The **CNN+CBAM** (b) integrates a 3D Convolutional Block Attention Module after the first convolutional layer to enhance channel and spatial feature representation with minimal parameter overhead. The **CNN–LSTM** (c) extends the baseline by adding a Bidirectional LSTM layer after feature flattening to capture temporal dependencies across extracted features. All models were trained using the Adam optimizer (learning rate 5×10^{-4}, β_1=0.9, β_2=0.999, ϵ=10^{-8}) to ensure consistency in comparison.

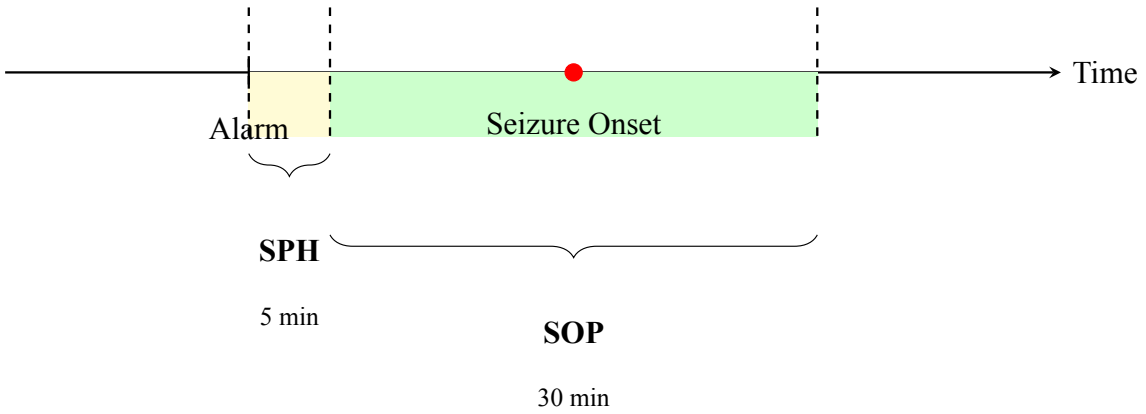


Figure 4: Illustration of Seizure Prediction Horizon (SPH) and Seizure Occurrence Period (SOP). The SPH is a 5-minute warning period, and the SOP is a 30-minute window in which a seizure is expected to occur. The red circle marks the predicted seizure onset, and a prediction is counted as correct if the seizure occurs after the SPH and within the SOP.

The five-minute gap, known as the Seizure Prediction Horizon (SPH), was excluded to ensure predictions occur at least five minutes in advance, allowing time for intervention. The preictal segments were used for both training and testing. Model performance was evaluated with respect to the SPH and the Seizure Occurrence Period (SOP) to maintain clinical relevance and consistency with previous studies.

Model evaluation followed a **leave-one-out cross-validation (LOOCV)** strategy applied separately to each patient. If a patient had N seizures, the model was trained on $N - 1$ seizures and tested on the remaining event. This process is repeated N times so that each seizure is used exactly once for testing. Interictal segments are divided into N equal parts using the same split scheme. Within each training fold, 25% of the most recent samples are reserved as a monitoring set to track validation performance and detect overfitting.

Because we adopted a patient-specific modeling approach, seizures from the same patient appear in training and testing across different folds, but never within the same fold. To prevent data leakage, all segments from a given test seizure and its corresponding interictal data were excluded from the training folds. The STFT preprocessing is a deterministic transformation without learned statistics and was therefore applied prior to cross-validation; no normalization or data-driven scaling was used.

2.5 Post Processing

The model outputs are refined before triggering an alarm to enhance the reliability of the prediction and reduce the rate of false alarms. Previous work has used *k-of-n* voting, where an alarm is issued only if at least k of the last n segments are classified as preictal [9]; other studies have employed smoothing of probabilities using temporal filters (e.g., Kalman). In this work, we propose a *Top-K soft fusion* strategy that retains confidence information by averaging the K most significant probabilities within a sliding look-back window.

2.5.1 Hard fusion (k-of-n)

Let $p_t \in [0, 1]$ denote the preictal probability at step t, with each segment of length $\Delta = 30\,\text{s}$. A look-back buffer of length n stores the most recent binary decisions: $B_t = \{b_{t-n+1}, \ldots, b_t\}$, where each b_t is obtained by thresholding the model output:

$$b_t = \mathbb{1}[p_t \geq \tau],$$

Moreover, $\mathbb{1}[\cdot]$ denotes the indicator function that returns 1 if the condition is true and 0 otherwise.

The k-of-n rule triggers an alarm at time t when the number of preictal detections within the last n segments reaches or exceeds k, i.e.,

$$\sum_{j=0}^{n-1} b_{t-j} \geq k.$$

This approach provides a simple and interpretable decision rule by aggregating recent binary outputs. However, since it relies only on thresholded values b_t, it discards the confidence magnitudes in p_t.

2.5.2 Soft fusion (Top-K)

Let $p_t \in [0, 1]$ denote the preictal probability at step t, with each segment of length $\Delta = 30\,\text{s}$. A look-back buffer of length w stores the most recent probabilities $S_t = \{p_{t-w+1}, \ldots, p_t\}$, where the total window duration $w\Delta$ corresponds to the chosen prediction time (e.g., $w = 60$ for 30 min or $w = 10$ for 5 min). The Top-K statistic is defined as

$$s_t = \frac{1}{K} \sum_{x \in \mathcal{T}_K(S_t)} x,$$

where $\mathcal{T}_K(S_t)$ denotes the K largest probabilities in S_t, each contributing equally with weight $1/K$, while the remaining $w - K$ values are ignored. An alarm is triggered when $s_t \geq \tau$. Top-K fusion focuses on the strongest signals, allowing it to trigger earlier than the k-of-n rule and to differ from averaging methods that mix all probabilities.

2.5.3 Real-time operation and SPH/SOP

The model processes one 30s segment at a time; after each update, the buffer of the most recent w segments is refreshed; the Top-K mean s_t is evaluated against τ, and an alarm is issued if $s_t \geq \tau$. Once an alarm is triggered, the system enters a *refractory* period of length equal to the SOP (no new alarms are counted), consistent with our episode-based evaluation. In deployment, an alarm is followed by the SPH (no alarm action), then the SOP during which a seizure is expected; this aligns the fusion logic with the SPH/SOP framework.

2.5.4 Implementation details

In all post-processing experiments, we set the look-back window to w=10 (corresponding to a five-minute buffer) to enable a rapid response. However, w can be extended if a longer memory is desired. The fusion parameters were set to n=10 and K=8, with $K \leq w$ ensuring a consistent comparison between hard and soft fusion methods. In the k-of-n approach, an alarm is issued when at least k of the last n binary predictions exceed the threshold. In Top-K, the average of the K highest predictive probabilities within the look-back buffer (using equal weights) is computed, and an alarm is triggered if this mean value exceeds the threshold. A refractory period equal to the SOP (30 min) was applied to prevent repeated alarms for the same episode, and an SPH of five minutes was used.

The decision thresholds τ were optimized individually for each patient using grid search $\tau \in [0.10, 0.95]$ (step 0.05) on the test outputs of Run 1, maximizing the scalar score $\overline{\text{Sens}} - \overline{\text{FPR}}$. The optimal τ was then fixed and applied unchanged to the independent Run 2 of the same folds for final evaluation.

Performance evaluation was conducted using leave-one-seizure-out cross-validation (LOOCV). The sensitivity and false prediction rate (FPR) are reported as the mean ± standard deviation across patients. This configuration facilitates a direct and controlled comparison between the k-of-n and Top-K fusion strategies.

3. RESULTS

We evaluated seizure prediction performance using two post-processing methods: **Top-K soft fusion** and the traditional **k-of-n** voting rule. In both approaches, we used $n = 10$ with a five-minute buffer and set $k = 8$.

In the k-of-n method, an alarm was triggered if at least 8 out of 10 consecutive segments were classified as preictal. In contrast, the Top-K soft fusion method averaged the highest eight probabilities among the last 10 segments, and an alarm was raised if this average exceeded a predefined threshold. To avoid multiple detections of the same event, a refractory period of 30 minutes (SOP) was applied following each alarm.

The results for the three architectures—Baseline CNN, CNN+LSTM, and CNN+CBAM—are presented in TABLE 3–TABLE 5. Across all experiments, **Top-K consistently improved model sensitivity compared with the k-of-n**. The impact on the false prediction rate (FPR) varied depending on the model architecture. In Experiment 1 (Baseline CNN), Top-K slightly increased the FPR, while in Experiments 2 (CNN+LSTM) and 3 (CNN+CBAM), the FPR remained nearly the same for both methods.

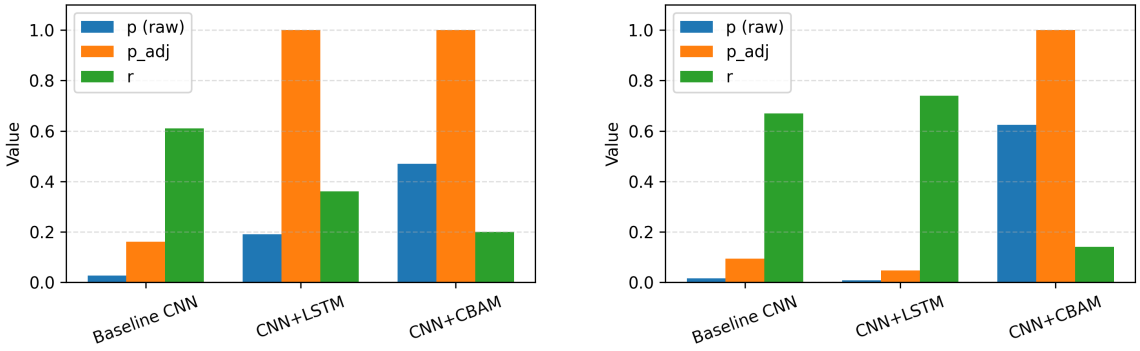
For example, in Experiment 1, Top-K achieved a macro sensitivity of 92.82% compared with 66.15% for k-of-n, with an FPR of 0.82 versus 0.53 false alarms per hour. Similar patterns were observed in Experiments 2 and 3, where Top-K improved detection performance without a significant increase in false alarms. For context, an FPR in the range of 0.7–0.9/h corresponds to approximately 17–22 false alarms per day, which exceeds the clinically acceptable limit of about 1–2 per day.

Therefore, further efforts to reduce the FPR or incorporate alarm aggregation mechanisms will be necessary for practical clinical deployment.

The patient-wise metrics are summarized as mean ± standard deviation (SD). We report both macro-averages (unweighted mean across patients) and micro-averages that weight sensitivity by the number of seizures and FPR by total interictal hours:

$$\text{Sens}_{\text{micro}} = \frac{\sum_i \text{TP}_i}{\sum_i N_i}, \qquad \text{FPR}_{\text{micro}} = \frac{\sum_i \text{FP}_i}{\sum_i T_i^{\text{inter}}}.$$

Ninety-five percent confidence intervals (95% CI) for patient-wise sensitivity used exact binomial (Clopper–Pearson) intervals based on TP_i out of N_i; FPR intervals used exact Poisson rate intervals for FP_i events over T_i^{inter} hours. For paired comparisons between Top-K and k-of-n within each experiment, we used the two-sided Wilcoxon signed-rank test across patients, with separate analyses for sensitivity and FPR. The p-values were corrected for multiple comparisons using the Bonferroni procedure; we report adjusted p-values (p_{adj}) and the effect size $r = Z/\sqrt{n}$ (see FIGURE 5).



(a) False Prediction Rate (FPR)

(b) Sensitivity

Figure 5: Statistical comparison of Top-K soft fusion vs. k-of-n post-processing across models. Bar charts show the Wilcoxon signed-rank test results for **Sensitivity** (right) and **False Prediction Rate (FPR)** (left) across the three CNN architectures. Each bar represents the raw p-value, Bonferroni-adjusted p_{adj}, and effect size r.

Across the three experiments, Top-K showed greater sensitivity; this improvement was statistically significant after correction in Exp2 (CNN+LSTM), while Exp1 (Baseline CNN) and Exp3 (CNN+CBAM) showed the same trend without reaching significance after Bonferroni adjustment.

4. DISCUSSION

The results showed that adding CBAM and LSTM layers led to only minor improvements, while the main performance gain came from the Soft Fusion method. By combining the Top-K highest probability predictions, Soft Fusion produced a smoother decision procedure that reduced the fluctuations seen with the traditional k-of-n filtering. This improvement was most evident in the reduced number of false alarms, a factor crucial for clinical application.

Table 3: Patient-wise sensitivity and FPR for Exp1 (Baseline CNN). Values are mean±SD; Overall rows report macro (mean±SD across patients) and micro (pooled counts, 95% CI).

Patient	TopK		K-of-N	
	Sens (%)	FPR	Sens (%)	FPR
chb01	100.00	0.47	100.00	0.41
chb02	100.00	0.83	0.00	0.30
chb03	100.00	0.91	66.67	0.23
chb05	80.00	1.29	20.00	0.00
chb09	100.00	0.32	100.00	0.78
chb10	83.33	1.58	50.00	1.42
chb13	80.00	0.21	80.00	0.21
chb14	80.00	2.00	80.00	1.80
chb18	83.33	0.50	50.00	0.29
chb19	100.00	1.16	33.33	0.16
chb20	100.00	0.45	80.00	0.30
chb21	100.00	0.68	100.00	0.77
chb23	100.00	0.23	100.00	0.16
Overall (macro)	92.82±9.51	0.82±0.55	66.15±33.25	0.53±0.54
Overall (micro)	92.19% [82.70, 97.41]	0.75 [0.65, 0.86]	68.75% [55.94, 79.76]	0.52 [0.43, 0.61]

This work extends the study by Truong et al. (2018) [9], who proposed the k-of-n post-processing method for seizure prediction on the CHB-MIT dataset using CNNs with STFT spectral inputs. Their model reached 81.2% sensitivity and a false prediction rate of 0.16 per hour. In comparison, the Soft Fusion method achieved higher sensitivity across all CNN-based architectures while maintaining a low false alarm rate within clinical limits. The findings indicate that averaging Top-K probabilities provides a more stable and reliable decision process than binary voting, which often overlooks early seizure patterns.

Our approach offers a good balance between sensitivity and false prediction rate (FPR) compared with other state-of-the-art methods (TABLE 1). Previous studies by Daoud and Bayoumi (2019) [18], Yang et al. (2021) [15], and Hu et al. (2023) [16], achieved high accuracy using advanced complex models such as Bi-LSTM with deep autoencoders, RDANet with dual self-attention, and hybrid Transformer models. While these models capture both spectral and temporal features effectively and often reduce the need for post-processing, they require high computational power and cause delays, which makes real-time use difficult. In contrast, the proposed Soft Fusion method offers a lightweight yet effective alternative that enhances prediction stability and performance with minimal computational overhead. Further evaluation is needed under patient-independent conditions and across multiple EEG datasets to assess the method's generalizability and clinical potential.

Recent transformer-based approaches have also reported remarkable results on the CHB-MIT dataset, as summarized in TABLE 6. For example, Yan et al. (2022) [26], used a three-tower Transformer with STFT features and achieved 96.0% sensitivity with an FPR of 0.047/h under a three-

Table 4: Patient-wise sensitivity and FPR for Exp2 (CNN+LSTM). Values are mean±SD; Overall rows report macro (mean±SD across patients) and micro (pooled counts, 95% CI).

Patient	TopK		K-of-N	
	Sens (%)	FPR	Sens (%)	FPR
chb01	100.00	0.88	100.00	0.77
chb02	33.33	0.26	33.33	0.00
chb03	100.00	0.73	66.67	0.14
chb05	60.00	0.21	40.00	0.43
chb09	100.00	0.71	75.00	0.80
chb10	100.00	2.04	66.67	1.92
chb13	100.00	0.57	80.00	0.50
chb14	100.00	1.40	80.00	1.60
chb18	83.33	0.17	66.67	0.25
chb19	66.67	0.68	33.33	0.36
chb20	60.00	0.25	60.00	0.15
chb21	75.00	0.55	75.00	0.36
chb23	100.00	0.23	100.00	0.16
Overall (macro)	82.95±22.13	0.67±0.54	67.44±21.77	0.57±0.58
Overall (micro)	85.94% [74.98, 93.36]	0.67 [0.58, 0.78]	70.31% [57.58, 81.09]	0.56 [0.48, 0.66]

minute SPH. Hussein et al. (2022) [27], applied a multi-channel Vision Transformer combined with continuous wavelet transforms in a patient-independent setting, reaching 99.8% sensitivity and 0.004/h FPR. In contrast, Hu et al. (2023) [16], proposed a hybrid Transformer with rhythm-specific embeddings that achieved 91.7% sensitivity and no false alarms. Although these attention-based models have demonstrated excellent performance, they typically require large parameter counts and multiple attention heads that increase computational costs and inference latency. Under comparable CHB-MIT conditions (30 min preictal, 5 min SPH, and 30 min SOP), our Top-K Soft Fusion achieved sensitivities of 90.82%, 82.95%, and 90.62%, with corresponding false prediction rates of 0.82, 0.67, and 0.80 false alarms per hour for the baseline CNN, CNN+LSTM, and CNN+CBAM models, respectively.

It is important to note that direct numerical comparisons across studies should be interpreted with caution, as experimental settings differ widely. Variations include the number of patients, preictal definitions, and evaluation protocols. A fair comparison requires testing all models under identical conditions, including a 30-minute preictal length, a 5-minute Seizure Prediction Horizon (SPH), and a 30-minute Seizure Occurrence Period (SOP). Even with these differences, the results show that an effective post-processing method such as Top-K Soft Fusion achieves performance comparable to recent Transformer-based models while keeping much lower complexity.

More advanced models, such as RDANet [15], and Transformer-based architectures [16], achieve high accuracy but rely on deep layers, multiple attention heads, and a large number of parameters. These factors increase memory usage, training time, and inference delay, which makes real-time de-

Table 5: Patient-wise sensitivity and FPR for Exp3 (CNN+CBAM). Values are mean±SD; Overall rows report macro (mean±SD across patients) and micro (pooled counts, 95% CI).

Patient	TopK		K-of-N	
	Sens (%)	FPR	Sens (%)	FPR
chb01	100.00	1.41	100.00	1.35
chb02	100.00	0.65	33.33	0.00
chb03	100.00	0.59	100.00	0.55
chb05	80.00	0.57	80.00	0.50
chb09	100.00	1.34	100.00	1.04
chb10	100.00	1.92	83.33	2.00
chb13	100.00	0.79	100.00	0.93
chb14	60.00	0.60	100.00	2.00
chb18	66.67	0.21	50.00	0.33
chb19	66.67	0.04	66.67	0.04
chb20	100.00	0.75	100.00	0.60
chb21	100.00	1.45	100.00	1.27
chb23	100.00	0.08	100.00	0.00
Overall (macro)	90.26±15.78	0.80±0.58	85.64±22.46	0.82±0.69
Overall (micro)	90.62% [80.70, 96.48]	0.89 [0.78, 1.00]	87.50% [76.85, 94.45]	0.79 [0.69, 0.90]

ployment difficult, especially on edge or wearable devices. In contrast, our Soft Fusion framework adds negligible computational cost since it works as a post-processing step and does not change the model's structure. This balance between simplicity and effectiveness makes it well-suited for real-time clinical monitoring and use in low-power embedded systems.

Table 6: Recent Transformer/attention-based seizure *prediction* benchmarks on CHB-MIT. "NR" denotes not reported.

Study	Year	Model	Setting	Preictal / SPH / SOP	Post-proc	Sens (%) / FPR (/h)
Yan et al. (2022) [26]	2022	Three-tower Transformer + STFT	Patient-specific	30 min / 3 min / 30 min	k-of-n (K=24,n=30)	96.01 / 0.047
Hussein et al. (2022) [27]	2022	Multi-channel Vision Transformer + CWT	Patient-independent	1 h / 5 min / NR	NR	99.8 / 0.004
Hu et al. (2023) [16]	2023	Hybrid Transformer + Wavelet Transform	Patient-specific	25 m / 5 min / 25 m	NR	91.7 / 0.00
This work (Top-K)	-	Baseline CNN + STFT	Patient-specific	30 min / 5 min / 30 min	Top-K (K=8,n=10)	92.82 / 0.82
	-	CNN + LSTM + STFT	Patient-specific	30 min / 5 min / 30 min	Top-K (K=8,n=10)	82.95 / 0.67
	-	CNN + CBAM + STFT	Patient-specific	30 min / 5 min / 30 min	Top-K (K=8,n=10)	90.62 / 0.80

Soft Fusion consistently improved the performance of the baseline CNN, CNN+LSTM, and CNN+CBAM models, showing that it is both effective and reliable. These improvements are clinically meaningful, as the Top-K fusion method reduced missed seizures without causing a major increase in false alarms, making seizure prediction systems more reliable and practical. However, the false-alarm rates at a five-minute SPH and 30-minute SOP are still higher than what is clinically acceptable. Future work should aim to lower the FPR further and enhance real-time performance by developing faster, more efficient, and better-integrated clinical systems.

5. CONCLUSION AND FUTURE WORK

This study proposed a Soft Fusion strategy using Top-K predictions as a post-processing method for seizure prediction. The framework, based on a baseline CNN and extended with CBAM and LSTM architectures, was tested on the CHB-MIT scalp EEG dataset. Results showed small gains from architectural changes, but the Soft Fusion method achieved the largest improvement. It produced smoother prediction outputs and reduced false alarms compared to the traditional k-of-n method. The findings confirm that post-processing is essential for improving seizure prediction and can be as important as model design itself.

Future work will focus on validating the Soft Fusion framework on more datasets to measure its robustness across patients and recording conditions. Further studies will explore different preictal durations and Seizure Prediction Horizons (SPHs) to assess post-processing performance under varied settings. Additional enhancements may include adaptive tuning of (K, w, and τ), stronger temporal modeling, and better handling of noise and artifacts. These steps aim to strengthen the framework and prepare it for real-time clinical use.

References

[1] Usman SM, Khalid S, Aslam MH. Epileptic Seizures Prediction Using Deep Learning Techniques. IEEE Access. 2020:8.

[2] Khan H, Marcuse L, Fields M, Swann K, Yener B. Focal Onset Seizure Prediction Using Convolutional Networks. IEEE Trans Bio Med Eng. 2018;65:2109-2118.

[3] Tsiouris KM, Pezoulas VC, Zervakis M, Konitsiotis S, Koutsouris DD, et al. A Long Short-Term Memory Deep Learning Network for the Prediction of Epileptic Seizures Using EEG Signals. Comput Biol Med. 2018;99:24-37.

[4] FU Z, DUAN L, XIE Y, JIANG K, ZHANG Y. EEG -Based Epilepsy Identification Using Multiple Feature Spaces Consistent Fusion With Label Relaxation. J Mech Med Biol. 2025;25:2540013.

[5] Maiwald T, Winterhalder M, Aschenbrenner-Scheibe R, Voss HU, Schulze-Bonhage A, et al. Comparison of Three Nonlinear Seizure Prediction Methods by Means of the Seizure Prediction Characteristic. Phys D Nonlinear Phenom. 2004;194:357-368.

[6] Dissanayake T, Fernando T, Denman S, Sridharan S, Fookes C. Deep Learning for Patient-Independent Epileptic Seizure Prediction Using Scalp EEG Signals. IEEE Sens J. 2021;21:9377-9388.

[7] Cho D, Min B, Kim J, Lee B. EEG-Based Prediction of Epileptic Seizures Using Phase Synchronization Elicited From Noise-Assisted Multivariate Empirical Mode Decomposition. IEEE Trans Neural Syst Rehabil Eng. 2017;25:1309-1318.

[8] Zhang Z, Parhi KK. Low-Complexity Seizure Prediction From IEEEG/SEEG Using Spectral Power and Ratios of Spectral Power. IEEE Trans Biomed Circuits Syst. 2016;10:693-706.

[9] Truong ND, Nguyen AD, Kuhlmann L, Bonyadi MR, Yang J, Ippolito S et al. Convolutional Neural Networks for Seizure Prediction Using Intracranial and Scalp Electroencephalogram. Neural Netw. 2018;105:104-111.

[10] Chen W, Li H, Zheng H, Yang Y, Yang L, Qin Y et al. Patient-Specific Seizure Prediction Using Convolutional Neural Networks and Scalp EEG 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). 2019:1-5.

[11] Baghdadi A, Fourati R, Aribi Y, Siarry P, Alimi AM. Robust Feature Learning Method for Epileptic Seizures Prediction Based on Long-Term EEG Signals. IEEE Trans Bio Med Eng. 2020.

[12] Ryu S, Joe I. A Hybrid Densenet-LSTM Model for Epileptic Seizure Prediction. Appl Sci. 2021;11:7661.

[13] Halawa RI, Youssef SM, Elagamy MN. An Efficient Hybrid Model for Patient Independent Seizure Prediction Using Deep Learning. Appl Sci. 2022;12:5516.

[14] Ibrahim AK, Zhuang H, Tognoli E, Muhamed Ali AM, Erdol N. Epileptic Seizure Prediction Based on Multiresolution Convolutional Neural Networks. Front Signal Process. 2023;3:1175305.

[15] Yang X, Zhao J, Sun Q, Lu J, Ma X. An Effective Dual Self-Attention Residual Network for Seizure Prediction. IEEE Trans Neural Syst Rehabil Eng. 2021;29:1604-1613.

[16] Hu S, Liu J, Yang R, Wang Y, Wang A, Li K et al. Exploring the Applicability of Transfer Learning and Feature Engineering in Epilepsy Prediction Using Hybrid Transformer Model. IEEE Trans Neural Syst Rehabil Eng. 2023;31:1321-32.

[17] Batista J, Pinto MF, Tavares M, Lopes F, Oliveira A, Teixeira C. EEG Epilepsy Seizure Prediction: The Post-Processing Stage as a Chronology. Sci Rep. 2024;14:407.

[18] Daoud H, Bayoumi MA. Efficient epileptic seizure prediction based on deep learning. IEEE Trans. Biomed Circuits Syst. 2019;13:804-813.

[19] Usman SM, Khalid S, Aslam MH. Epileptic Seizures Prediction Using Deep Learning Techniques. IEEE Access. 2020;8:39998–40007.

[20] Das K, Daschakladar D, Roy PP, Chatterjee A, Saha SP. Epileptic Seizure Prediction by the Detection of Seizure Waveform From the Pre-Ictal Phase of EEG Signal. Biomed Signal Process Control. 2020;57:101720.

[21] https://physionet.org/content/chbmit/1.0.0/

[22] Shoeb AH. Application of machine learning to epileptic seizure onset detection and treatment (Doctoral dissertation, Massachusetts Institute of Technology). 1981.

[23] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, et al. Physiobank, Physiotoolkit, and Physionet: Components of a New Research Resource for Complex Physiologic Signals. Circulation. 2000;101:E215-20.

[24] Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional Block Attention Module. Comput Vis ECCV. 2018;2018:3-19.

[25] Lu X, Wen A, Sun L, Wang H, Guo Y, Ren Y. An Epileptic Seizure Prediction Method Based on Cbam -3d CNN-LSTM model. IEEE J Transl Eng Health Med. 2023;11:417-423.

[26] Yan J, Li J, Xu H, Yu Y, Xu T. Seizure Prediction Based on Transformer Using Scalp Electroencephalogram. Appl Sci. 2022;12.

[27] Hussein R, Lee S, Ward R. Multi-Channel Vision Transformer for Epileptic Seizure Prediction. Biomedicines. 2022;10:1551.