

RESEARCH

Open Access



GEM-CRAP: a fusion architecture for focal seizure detection

Jianwei Shi^{1,2,3†}, Yuanyuan Zhang^{1,2,3†}, Ziang Song^{1,2,3†}, Hang Xu^{1,2,3}, Yanfeng Yang^{1,2,3}, Lei Jin^{1,2,3}, Hengxin Dong^{1,2,3}, Zhaoying Li^{1,2,3}, Penghu Wei^{1,2,3*}, Yongzhi Shan^{1,2,3*} and Guoguang Zhao^{1,2,3*}

Abstract

Background Identification of seizures is essential for the treatment of epilepsy. Current machine-learning and deep-learning models often perform well on public datasets when classifying generalized seizures with prominent features. However, their performance was less effective in detecting brief, localized seizures. These seizure-like patterns can be masked by fixed brain rhythms.

Methods Our study proposes a supervised multilayer hybrid model called GEM-CRAP (gradient-enhanced modulation with CNN-RES, attention-like, and pre-policy networks), with three parallel feature extraction channels: a CNN-RES module, an amplitude-aware channel with attention-like mechanisms, and an LSTM-based pre-policy layer integrated into the recurrent neural network. The model was trained on the Xuanwu Hospital and HUP iEEG dataset, including intracranial, cortical, and stereotactic EEG data from 83 patients, covering over 8500 labeled electrode channels for hybrid classification (wakefulness and sleep). A post-SVM network was used for secondary training on channels with classification accuracy below 80%. We introduced an average channel deviation rate metric to assess seizure detection accuracy.

Results For public datasets, the model achieved over 97% accuracy for intracranial and cortical EEG sequences in patients, and over 95% for mixed sequences, with deviations below 5%. In the Xuanwu Hospital dataset, it maintained over 94% accuracy for wakefulness seizures and around 90% during sleep. SVM secondary training improved average channel accuracy by over 10%. Additionally, a strong positive correlation was found between channel accuracy distribution and the temporal distribution of seizure states.

Conclusions GEM-CRAP enhances focal epilepsy detection through adaptive adjustments and attention mechanisms, achieving higher precision and robustness in complex signal environments. Beyond improving seizure interval detection, it excels in identifying and analyzing specific epileptic waveforms, such as high-frequency oscillations. This advancement may pave the way for more precise epilepsy diagnostics and provide a suitable artificial intelligence algorithm for closed-loop neurostimulation.

[†]Jianwei Shi, Yuanyuan Zhang and Ziang Song contributed equally to this work.

*Correspondence:

Penghu Wei
weipenghu@xwhosp.org
Yongzhi Shan
shanyongzhi@xwhosp.org
Guoguang Zhao
ggzhao@vip.sina.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Epilepsy is a chronic neural disease that affects approximately 70 million individuals worldwide and is the second most common neurological condition after stroke [1]. It is characterized by spontaneous seizures and poses significant diagnostic and treatment challenges. Despite the availability of over 20 antiepileptic drugs for clinical use [2], nearly one-third of patients with epilepsy cannot control their seizures through medication alone and must resort to surgical intervention [1]. For seizure foci that are widely dispersed, poorly localized, or overlap with the eloquent cortex, effective resection may be unachievable, posing potential neurological risks such as memory or language impairment, with diminished therapeutic benefits [3]. This complexity underscores the need for advanced diagnostic tools and treatments that provide more precise and actionable insights into the epileptic brain [4].

Neuroimaging provides foundational structural and functional localization for neurological disorders [5, 6]; however, in epilepsy, electroencephalography (EEG), which records electrical brain activity via scalp or intracranial electrodes, captures essential spatiotemporal neurodynamics crucial for enhancing epilepsy diagnosis and refining intervention strategies [7]. While scalp EEG offers valuable insights, intracranial EEG (iEEG), such as stereoelectroencephalography (SEEG) and electrocorticography (ECoG), provide a high-resolution view of internal brain activity with superior signal clarity and precise epileptic focus localization, which is crucial for identifying complex seizure patterns [3, 8, 5]. iEEG, integrated with artificial intelligence (AI) technologies such as deep learning, can detect subtle changes in frequency or amplitude, thereby advancing the development of sophisticated diagnostic models that improve seizure prediction and classification [9, 10].

Convolutional neural networks (CNNs) are pivotal in effectively extracting spatiotemporal features from EEG signals and capturing complex brain dynamics [11]. These networks utilize convolutional filters to analyze the data at various scales. Long short-term memory (LSTM) networks, an advanced form of recurrent neural networks (RNNs), excel at processing time-series data that are crucial for predicting epileptic seizures. LSTMs address traditional RNN challenges, such as gradient vanishing, by incorporating mechanisms such as forget gates to maintain learning stability over time [12]. The Attention Mechanism was initially introduced in Natural Language Processing (NLP) to address information loss in traditional models like RNN and LSTM when handling long sequences. It helps focus on relevant parts of the input, improving tasks like translation. However, processing epileptic EEG sequences poses challenges due to

signal non-linearity, low Signal-to-Noise Ratio (SNR), and complex seizure patterns. To capture key features, adjustments are needed, such as dynamically adjusting attention weights and optimizing distribution for seizures, enhancing detection and prediction accuracy. Support vector machines (SVMs) enhance classification by creating optimal hyperplanes in high-dimensional spaces, effectively distinguishing between neurological states [13]. The integration of these advanced models forms a robust diagnostic framework that improves seizure detection and extends the diagnosis of other neurological disorders [11, 14–21]. This synergy advances diagnostic accuracy and deepens our understanding of neural processes.

With the advancement of algorithm-based treatments, such as closed-loop neurostimulation [22], the application and real-time analysis of iEEG is gaining increasing attention [23, 24]. Existing detection algorithms that rely on line length, area, and threshold limits often overlook complex brain rhythmic activity, leading to the misclassification of normal discharges as seizures [25]. Moreover, current AI-based deep learning algorithms for seizure detection are mostly trained on public datasets with clear seizure characteristics and fail to capture subtle seizure events (Fig. 1). This study proposes an algorithmic model named GEM-CRAP (gradient-enhanced modulation with CNN-RES, attention-like, and pre-policy networks). Compared to existing deep learning models, GEM-CRAP has a significant advantage in effectively detecting more subtle focal seizures. The innovation of the GEM-CRAP model lies in its ability to accurately analyze amplitude variations in iEEG signals through the Amplitude-Aware Layer, extracting critical local features, particularly segments with high amplitude and prolonged duration. Simultaneously, the Pre-Policy Networks reduce information redundancy in interictal data, ensuring that all filtered interictal or ictal segments positively contribute to feature extraction and classification. This approach minimizes interference from irrelevant factors (such as external stimuli and normal EEG rhythm-induced amplitude-frequency changes) and redundant or similar segments during interictal periods, which is crucial for the identification of focal epilepsy. GEM-CRAP employs a custom ‘ampeak_trough’ function to automatically detect all local peaks and troughs in EEG signals, utilizing an objective function to find local optimal solutions for trigger point potentials. This process filters out segments with high amplitude and significant duration, assigning weight coefficients to them. These coefficients are consistently applied in subsequent network processing, further enhancing sensitivity and accuracy in detecting focal epileptic seizures. At the same time, it maintains considerable detection performance for generalized

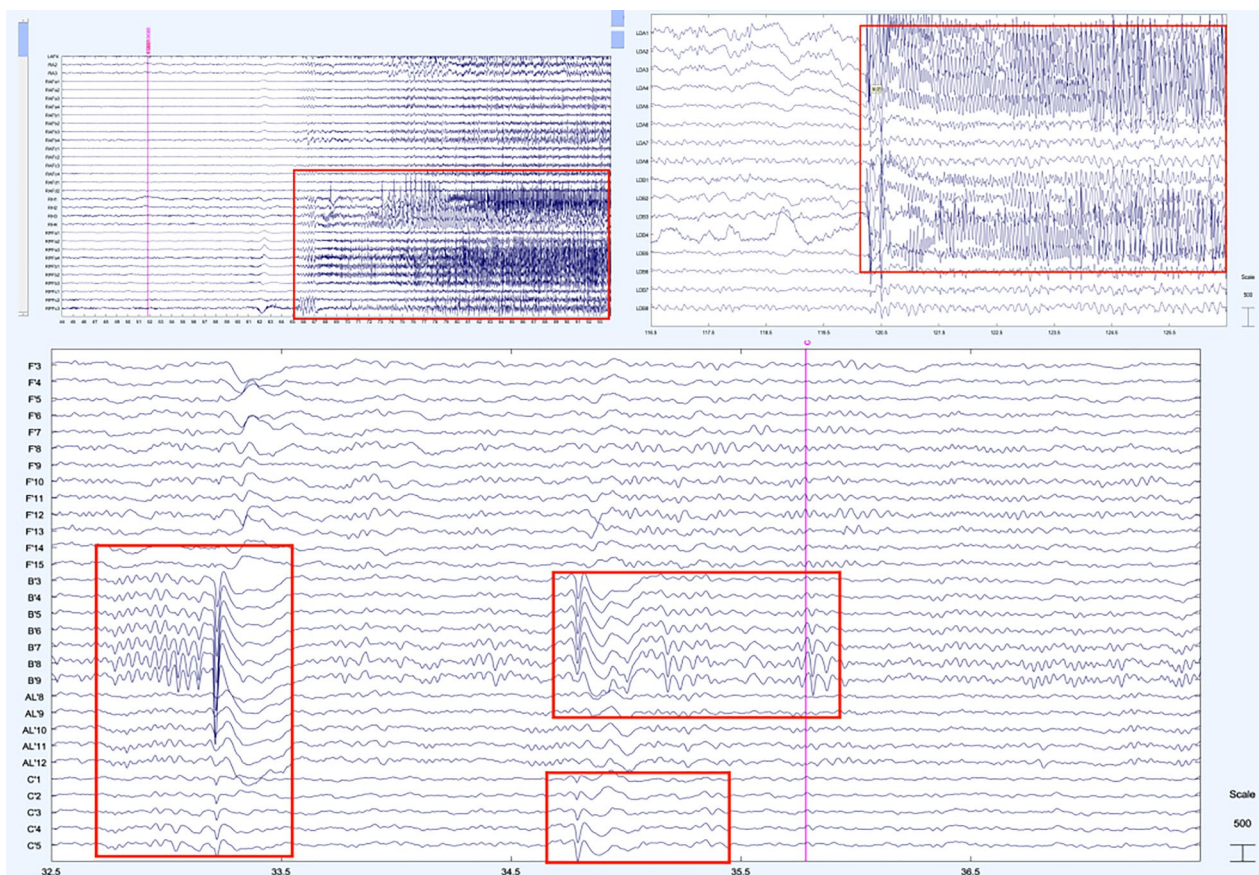


Fig. 1 Two distinct types of epileptic seizures as recorded by SEEG. The top two panels depict generalized tonic–clonic seizures, characterized by a wide spread and prolonged duration, involving multiple brain regions. In contrast, the bottom panel presents a focal seizure with limited spatial spread, shorter duration, and confined to specific brain areas. These differences demonstrate the complexity and variability of epileptic events

seizures that are more widespread and have more distinct characteristics. This capability highlights our model's sensitivity to diverse seizure patterns, making it more versatile for clinical applications involving heterogeneous seizure presentations. The overall data flow diagram of our framework can be found in Supplementary Materials—Data Stream.

Materials and methods

Datasets and pre-processing

Two datasets were used in the study. HUP iEEG dataset from the openneuro platform: this dataset contains de-identified patient data from the University of Pennsylvania Hospital, targeting surgical treatments for drug-resistant epilepsy. It included data from 54 subjects who underwent intracranial electroencephalography (iEEG) monitoring using either subdural grids, strips, and depth electrodes (ECoG) or stereotactically placed depth electrodes (SEEG). The electrophysiological data encompassed both the interictal and ictal periods and provided

electrode locations in the ICBM152 MNI space. Additionally, the dataset included clinically identified seizure onset channels and channels overlapping with resection/ablation areas determined through meticulous segmentation of the resection cavity. SEEG Dataset from Xuanwu Hospital, Capital Medical University: This dataset consists of SEEG recordings from 29 patients with mesial temporal lobe drug-resistant epilepsy captured during both awake and sleep states following SEEG implantation surgery. This use of these data was approved by the local ethics committee of the Xuanwu Hospital, Capital Medical University, Beijing, China. The ethics committee's phone, email, and address as follows: 0086-10-83919270, xwkyethics@163.com, and No. 45, Changchun Street, Xicheng District, Beijing 100053, China. All the participants provided written informed consent.

Using the open-source toolkits Brainstorm and EEGLAB, we applied a 0.5–120 Hz bandpass filter and removed 50 Hz and 60 Hz power line interference. We then downsampled the data to 500 Hz ($f_s=500$ Hz) and

performed an interpolation-based reconstruction, reducing the original electrode sampling rate from 2048 to 500 Hz. After feature point calculation, the signal was interpolated and baseline-corrected. Monopolar EEG data were converted to bipolar data, and reference electrodes were removed for global re-referencing to minimize external physical interference affecting the EEG recording accuracy.

Our study filtered HUP iEEG and Xuanwu Hospital SEEG data based on electrode localization and distribution of seizure onset channels. The HUP iEEG dataset was divided into SEEG and ECoG, based on the information provided by the platform. The SEEG data from Xuanwu Hospital were classified based on awake and sleep states and further divided into Generalized tonic-clonic seizures (GtcS) and focal seizures (FS), according to the range and intensity of the seizures. To achieve a comprehensive classification training, multiple data subsets were obtained according to the classification method presented in Table 1. FS (Subset1,2) is the status we focused on, particularly the performance of small-scale focal seizures. GtcS can be included in the SEEG category of the HUP for joint training (Subset3). Finally, all data were normalized and multiple seizure events were concatenated along the same dimension. The SEEG dataset included 64 patients from HUP(35) and Xuanwu Hospital(29), whereas the ECoG dataset comprised 19 patients from HUP. All 29 SEEG cases from Xuanwu Hospital utilized channels containing seizures. In the HUP iEEG dataset, 36 cases included seizure channels, while the remaining 18 cases provided only interictal channel data.

The concatenated EEG data were labeled interictal (0), ictal (1), or pre-ictal (2). All labeled data were saved as single-channel CSV files, with each file containing data from a single channel, including 201 s data points and corresponding feature labels for each point.

Model mechanism

Feature vectors and frequency domain analysis

The datasets were sampled at a frequency of 500 Hz with a sequence length of 250, corresponding to a sliding detection window of 0.5 s, with a 50% overlap between

each pair of adjacent sequences. The data were reshaped into a three-dimensional feature vector (batchsize, sequencelength, and features) with a potential mapping space. For the EEG data, a single data point had only one possible feature in the epileptic state; therefore, in the initial model input, the feature dimension was set to one.

The additional feature dimensions introduced into subsequent network layers can decompose different signal amplitude components or extract energy from various frequency bands. This enhances the model's ability to recognize complex dynamic characteristics of epileptic seizures more effectively.

By optimizing the moving-window technique, only half of the spectrum was obtained, and downsampling was performed to reduce the data symmetry. In the actual processing of the network, the frequency-domain distribution of each EEG time series was extracted individually, providing a reliable data source containing more hidden information for high-dimensional analysis in the subsequent convolutional layers (Fig. 2).

Attention-like mechanism in amplitude-aware module

The AmplitudeAwareLayer analyzes voltage changes in EEG data points to capture signal amplitude variations by reshaping the data into sequences of a given length. On one hand, this module extracts and calculates the difference between the maximum and minimum values of all EEG data points within each sequence, enabling feature comparison between sequences. At the same time, the module can also learn the differences in amplitude range between FS and GtcS EEG signals at the overall sequence level.

On the other hand, our study introduced a custom ampeak_trough function, designed to capture all local peaks and troughs within each sequence, enabling adaptive selection of key features within the sequence (Fig. 3). First, the irregular raw EEG signal is fitted into a regular pattern by downsampling and then reconstructing it through interpolation as a superposition of cosine functions of varying scales. This process calculates the amplitude differences (*diffs*) between each pair of adjacent peaks and troughs (based on different predefined

Table 1 Classification of the entire dataset into multiple subsets

Dataset	Class	Subset1	Subset2	Subset3	Subset4	Subset5
HUP iEEG	SEEG			√		√
	ECoG				√	√
XW SEEG	Awake GtcS			√		√
	Awake FS	√				√
	Sleep GtcS			√		√
	Sleep FS		√			√

GtcS Generalized tonic-clonic seizures, FS focal seizures, XW Xuanwu Hospital

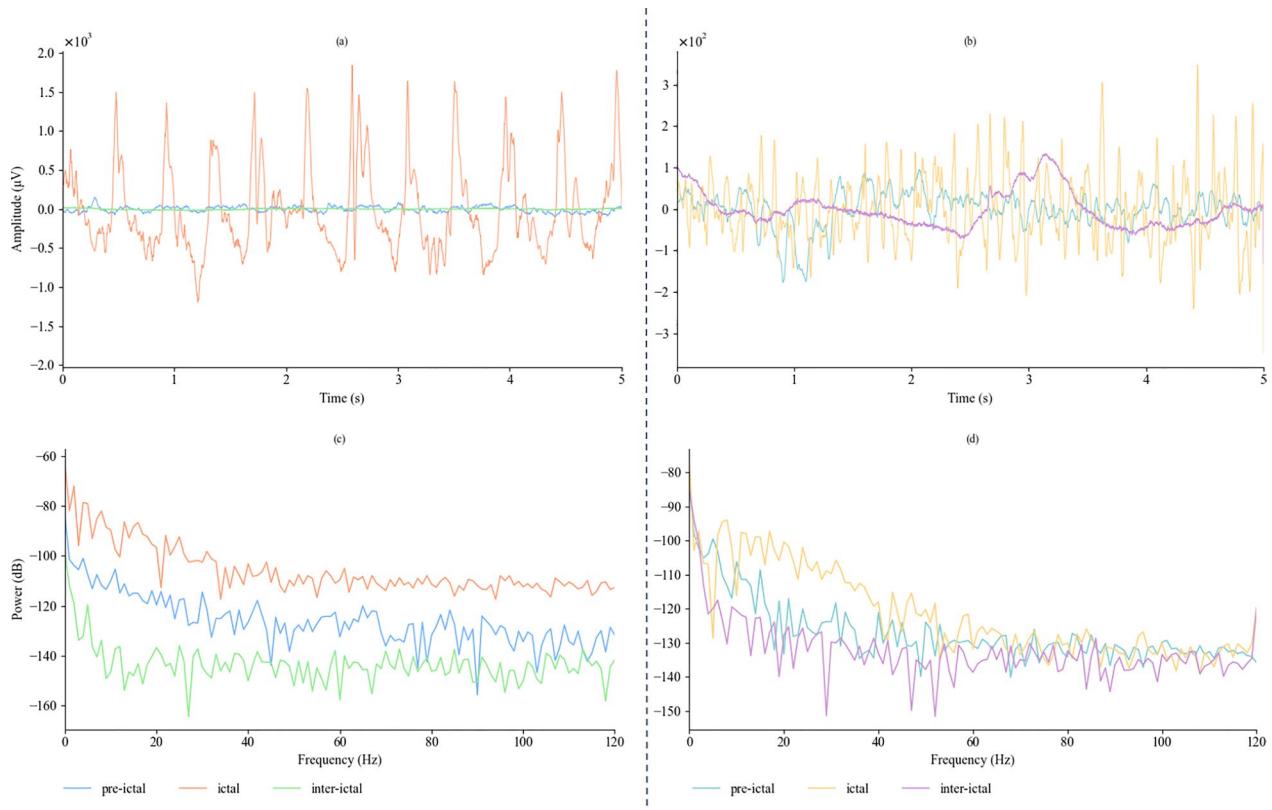


Fig. 2 Time–frequency distribution of 5-s EEG data, corresponding to three different seizure states. Comparative analysis of seizure characteristics between ECoG and SEEG based on time-domain and frequency-domain methodologies. **a** ECoG amplitude, **b** SEEG amplitude, **c** ECoG power spectrum, **d** SEEG Power Spectrum. As clearly observed from the Fig. 1, there are significant differences in seizure characteristics between SEEG and ECoG data. SEEG seizure characteristics are predominantly distributed in a higher frequency range, and in the pre-seizure state, SEEG exhibits more pronounced fluctuations, with more distinct differences from the interictal state compared to ECoG data. This observation suggests the need for classification-based training on the datasets

parameters) to filter out significant segments with high amplitude ($\text{diffs} \geq x \mu V$) and a duration of at least 50% of the sequence length. The value of x can be arbitrarily set, and when used for feature extraction in the FS, the model can filter locally optimal values for classification based on the overall amplitude information of the EEG time series. Specifically, this method is based on the amplitude of all time points within the EEG sequence. When the value at the upper quartile exceeds 150% of the mean amplitude of the entire sequence, this point is used as a trigger to begin an adaptive local optimization search for the parameter x . Starting from this trigger point, the value of x is adjusted by fluctuating upwards and downwards to determine the local optimal solution that best reflects the key features of the sequence. These high-amplitude and long-duration segments are considered the "key points" of the sequence and are fed back to the amplitude perception module as important features of the sequence. Using GtcS data as the standard, adaptively adjust the coefficients α_1 and α_2 for amplitude difference and volatility within the FS dataset.

In other words, we can amplify the network's attention to amplitude features in FS data by adjusting the coefficients accordingly, aiming to achieve the same recognition effectiveness as with GtcS. The specific implementation for finding the optimal solution x is as follows:

$$T(x) = -\alpha_1 \left(\frac{1}{|H(x)|} \sum_{i \in H(x)} \text{diffs}[i] - \frac{1}{|L(x)|} \sum_{j \in L(x)} \text{diffs}[j] \right) + \alpha_2 (\sigma_H(x) + \sigma_L(x)) \quad (1)$$

Define the objective function $T(x)$, where $H(x)$ represents the set of indices for all high-amplitude $\text{diffs}(\text{diffs}[i] > x)$, and $L(x)$ represents the set of indices for all low-amplitude $\text{diffs}(\text{diffs}[j] < x)$. $|H(x)|$ and $|L(x)|$ denote the number of high-amplitude and low-amplitude points, respectively, while $\sigma_H(x)$ and $\sigma_L(x)$ denote their respective standard deviations. The first term of the polynomial represents the amplitude difference,

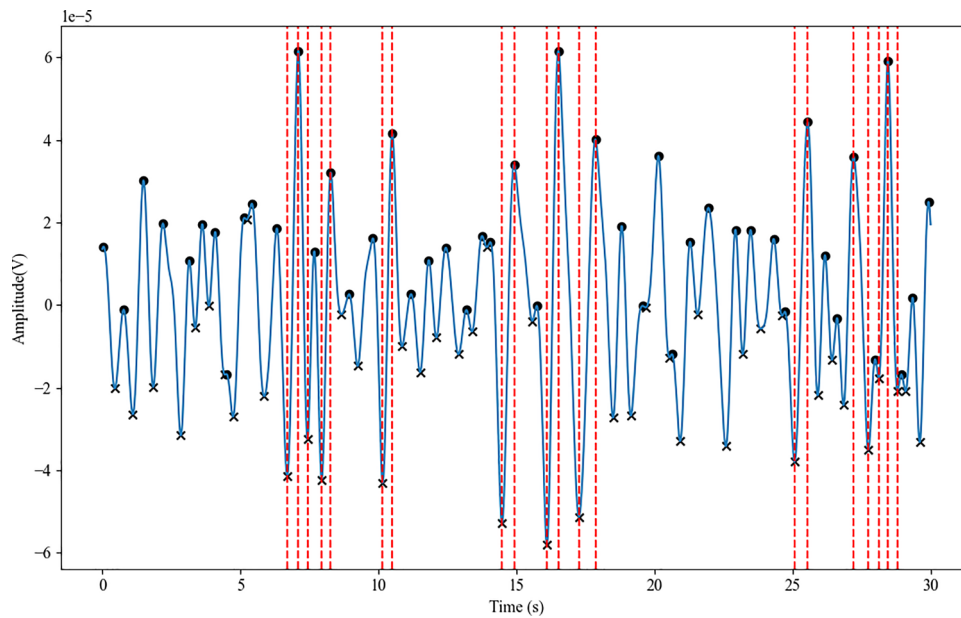


Fig. 3 The calculation and extraction of *diffs* is based on the difference between the peaks and troughs of a complete waveform. Attention features are added to waveforms that meet specific criteria

with α_1 as its weight coefficient; the second term represents volatility, with α_2 as its weight coefficient.

$$w_i(x) = \begin{cases} 1, & \text{if } \text{diffs}[i] \geq x \\ 0, & \text{if } \text{diffs}[i] < x \end{cases} \quad (2)$$

An attention mechanism is introduced based on the standard x , assigning weights to the *diffs*. The upper quartile of the *diffs* distribution serves as the initial trigger for x . $w_i(x)$ represents the weighting function, indicating whether the *diffs*[i] is classified as a high-amplitude point exceeding the threshold x .

$$\begin{cases} \frac{\partial T(x)}{\partial x} = \lim_{\epsilon \rightarrow 0} \frac{T(x+\epsilon) - T(x)}{\epsilon} \\ x_{t+1} = x_t - \alpha \frac{\partial T(x_t)}{\partial x} \end{cases} \quad (3)$$

$\frac{\partial T(x)}{\partial x}$ is the gradient of the objective function $T(x)$, and x_t is the threshold value of x at gradient update time step t . Through each update, we can find the optimal x that minimizes the objective function $T(x)$, which serves as the threshold that most effectively distinguishes between high-amplitude and low-amplitude points.

This mechanism effectively guides the “attention” of the network toward signal segments that exhibit prominent amplitudes and durations, thereby achieving a dynamic focus on key features. Filtering the significant features enhanced the overall perceptual ability of the model.

Additionally, this module uses two fully connected layers to further process the extracted features, capturing complex relationships and high-dimensional information among them. The first layer provides nonlinear

transformations to enhance feature representation, while the second layer performs dimensionality reduction and generates the final feature vector for subsequent classification or regression tasks. The fully connected layers use the Adam algorithm for gradient descent and updates of the weight matrix W and deviation vector b according to the model’s macro definition. Below are the parameter update equations:

$$\begin{aligned} g_t^W &= \nabla_W J(\theta_t) \\ g_t^b &= \nabla_b J(\theta_t) \\ m_{n+1} &= \beta_1 \cdot m_n + (1 - \beta_1) \cdot g_{n+1} \\ v_{n+1} &= \beta_2 \cdot v_n + (1 - \beta_2) \cdot g_{n+1}^2 \\ \hat{m}_n &= \frac{m_n}{1 - \beta_1^n} \\ \hat{v}_n &= \frac{v_n}{1 - \beta_2^n} \\ W_{n+1} &= W_n - \frac{\eta}{\sqrt{\hat{v}_n} + \epsilon} \cdot \hat{m}_n^W \\ b_{n+1} &= b_n - \frac{\eta}{\sqrt{\hat{v}_n} + \epsilon} \cdot \hat{m}_n^b \\ n &= 1, 2, 3, \dots, 511, 512 = \text{batch_size} \end{aligned} \quad (4)$$

where g_t^W is the gradient of weights θ_t , and g_t^b is the gradient of biases. m_{n+1} is the first moment estimate, β_1 is the decay factor, and v_{n+1} is the second moment estimate, β_2 is the decay factor. \hat{m}_n and \hat{v}_n are the bias-corrected moment estimates. η is the learning rate, and ϵ prevents division by zero.

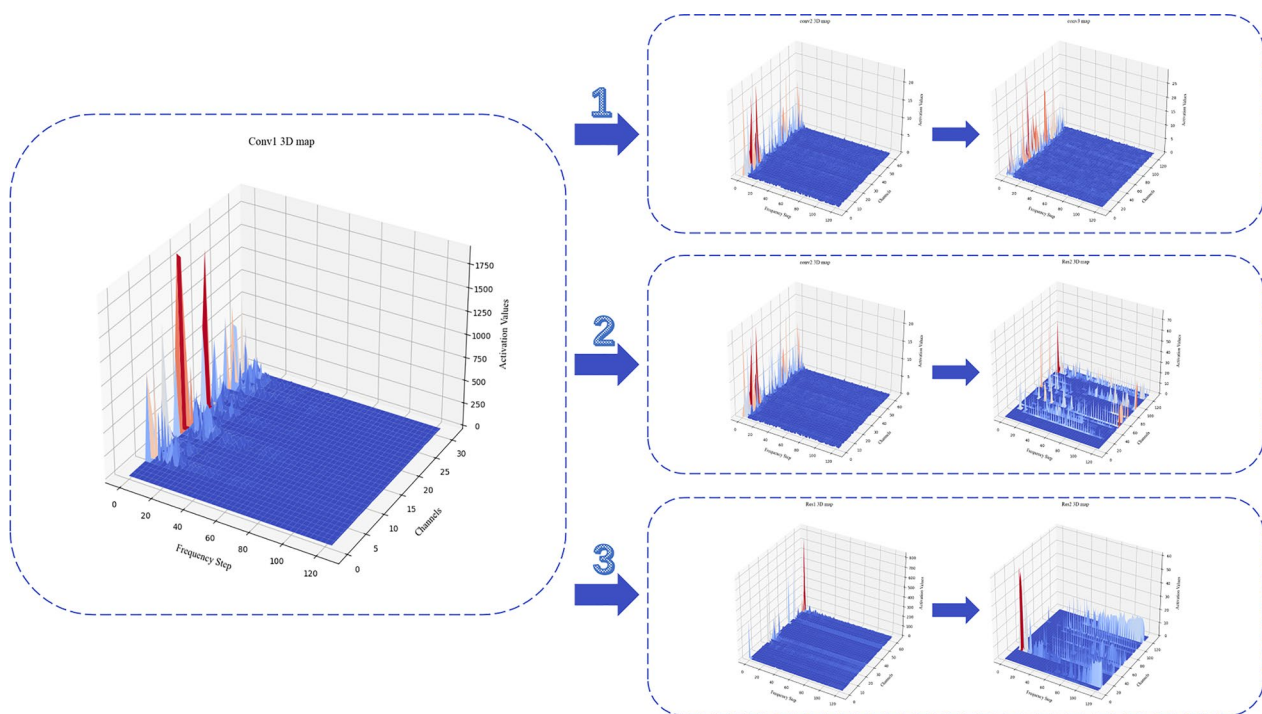


Fig. 5 The three-dimensional feature topography in the convolutional layers is composed of frequency steps, channel counts, and activation values. It encompasses a comparison of the three-layer convolution process and the CNN-RES feature extraction performance. (1) [conv1-conv2-conv3] (2) [conv1-conv2-RES2] (3) [conv1-RES1-RES2]

intermediate module, plays a transitional role in skip connections between the original convolutional layer and the residual blocks. The combined output of Conv1 and Res2 is necessary to achieve an optimal classification performance for both low- and high-frequency components. The primary performance improvement of Res2 in the third route is based on the enhancement of Res1, converting the activation values of discrete high-frequency signal components in the second route Res2 into continuous high-frequency regions, which demonstrates their capacity to extract high-frequency information from EEG signals, highlighting the impact of feature distribution across different channels on the output of the module and the final classification performance.

Pre-policy network

A policy network was integrated into the preprocessing architecture before the main model RNN input, enabling time-domain decision-making and analysis of the raw EEG data (Fig. 6). The policy network evaluates each time step in each sequence within a batch using LSTM. The hidden temporal information between time steps is mapped onto 128 related dimensions, and the output is mapped to probabilities, indicating whether

to process the input via a linear decoder. These decision probabilities, which act as gating signals, determine whether the input data are forwarded to the main model.

A reward mechanism is introduced through a softmax layer that generates the probability of processing at each time step. This mechanism trains the policy network by rewarding improvements in the performance of the main model, converting Boolean masks into integer indices, and calculating the average log probability of the decision outputs as the loss function. This allows the policy network to self-adjust and optimize the overall model performance. Specifically, if skipping certain time steps leads to more accurate predictions or reduced losses, the policy network receives positive feedback. Conversely, if it erroneously skips important time steps, the network is penalized. This feedback-based training method integrates policy loss as a “true reward,” enabling the policy network to self-adjust over time, thereby optimizing the entire model. In addition, if the policy network is disabled, all data are passed to the main RNN without affecting the overall network training or operations.

The LSTM output O_n is mapped to the probabilities of the two actions P_n for each sequence using the decoder, as described by the following equation:

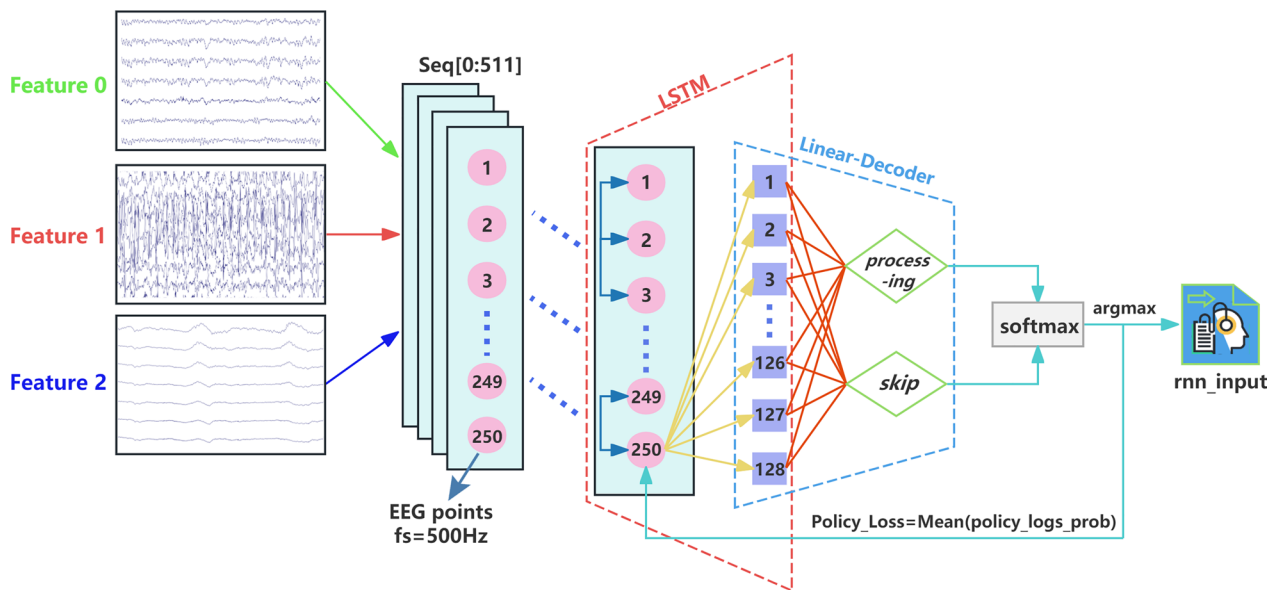


Fig. 6 The pre-policy network architecture enhances data prior to the RNN input gate, integrating loss functions to train using a reward mechanism

$$O_n = LSTM(x_n, h_{n-1}) = \sigma(W_o \cdot [h_{n-1}, x_n] + b_o)$$

$$P_n = Softmax(W_o \cdot O_n + b_o)_i = \frac{e^{(W_o \cdot O_n + b_o)_i}}{\sum_j e^{(W_o \cdot O_n + b_o)_j}} \quad (5)$$

W_o and b_o are the output weights and deviation of the decoder, h_{n-1} is the hidden state of the previous time step in the same batch, and $e^{(W_o \cdot O_n + b_o)_i}$ is the exponent of $(W_o \cdot O_n + b_o)_i$, with j iterating overall output class indices.

The loss function used for training the pre-policy network is shown as follows:

be time consuming, particularly when tuning parameters and applying kernel functions, once trained, SVM offers a considerably fast prediction speed. This is a significant advantage for applications requiring rapid real-time responses that align well with the demands of closed-loop neurostimulation systems.

Based on this, we introduce SVM as a supplemental reinforcement training channel to improve the performance of small-sample data with low classification accuracy (where the single-channel validation accuracy is below 80%). By loading a pretrained feature-extraction model into the SVM network, the features were extracted

$$L(\theta) = -E \left[\sum_{t=1}^T \alpha \cdot R_{t,FS} \cdot \log \pi(a_t | s_{t,FS}; \theta) + \beta \cdot R_{t,GtcS} \cdot \log \pi(a_t | s_{t,GtcS}; \theta) \right] \quad (6)$$

θ represents the network parameters, $R_{t,FS}$ and $R_{t,GtcS}$ represent the performance feedback rewards from the main model on the FS and GtcS data subsets, respectively; T denotes the sequence length, set to 250; α and β are the weighting coefficients for the FS and GtcS datasets, assigning different weights to different data subsets; $\log \pi(a_t | s_{t,class}; \theta)$ represents the logarithm of the average decision probability at each time step.

Channel training combined with SVM

Compared to CNN ensemble models, which require large amounts of data for training, SVM often perform better on small-sample datasets. Although SVM training can

and stored along with their corresponding labels. A pipeline that included both Standard Scaler and SVC was created, and GridSearchCV was used to determine the optimal SVM parameters. This process was completed by specifying a parameter grid and applying a five-fold cross-validation strategy with defined scoring criteria.

For the parameter grid, c serves as the regularization parameter that controls the penalty strength during model training. The gamma parameter (kernel function parameter) was applied to the radial basis function (RBF) and polynomial (poly) kernels. A smaller gamma value extends the range of influence of each data sample, resulting in a smoother decision boundary. For the SVM

network in this study, two parameter options were provided and a one-versus-one decision function was used for the three feature labels, expressed as follows:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n [x_i - \bar{x}]^2$$

$$\text{gamma}_1 = \text{'scale'} = \frac{1}{[\text{features}] \cdot S^2}$$

$$\text{gamma}_2 = \text{'auto'} = \frac{1}{[\text{features}]}$$

$$K(x, x[i]) = e^{-\gamma \|x - x[i]\|^2}$$

$$f(x) = \text{sgn} \left(\sum_{i=1}^N \alpha_i \cdot \text{Label}[i] \cdot K(x, x[i]) + b \right) \quad (7)$$

where x represents the feature vector input to the SVM (40,000 dimensions corresponding to the output from the fully connected layer of the feature extraction network). The parameters α_i and b are the decision network learned parameters, while $x[i]$ and $\text{Label}[i]$ refer to the support feature vectors from the input that influence the SVM decision and their corresponding labels. K is a kernel function that is crucial in mapping the input data to a higher-dimensional space. This mapping enables the SVM to handle nonlinearly separable datasets effectively, thereby facilitating the search for an optimal hyperplane that maximally separates different classes.

Evaluation metrics

Confusion matrix

To evaluate the prediction performance of the model for each label, we calculated a confusion matrix based on the model results for the test set. The confusion matrix derived from the three-class classification results of the model formed a 3×3 matrix, as shown below:

$$\text{Confusion matrix} = \begin{pmatrix} TP_{00} & FP_{01} & FP_{02} \\ FN_{10} & TP_{11} & FP_{12} \\ FN_{20} & FN_{21} & TP_{22} \end{pmatrix} \quad (8)$$

where the element TP_{ii} (True Positives) represents the number of instances correctly predicted as category i . FP_{ij} (False Positives) represents the number of instances where the actual category is i , but they are incorrectly predicted as j . FN_{ij} (False Negatives) represents the number of instances where the actual category is j , but they are incorrectly predicted as i .

Deviation rate

The deviation rate is a metric used to measure the overall deviation in the model recognition of seizure state intervals. In this study, when the model predicted five consecutive labels of 2, the pre-seizure state was considered to have begun, and the sequence containing the first label 2 (0–2) was identified as the predicted pre-seizure onset sequence. Similarly, when transitioning from the pre-seizure state to a full seizure (2–1), the sequence containing the last label, 1, was identified as the predicted termination of the pre-seizure state and the onset of the seizure.

$$\text{Deviation}_{0-2} = \frac{XP_{0-2} - XT_{0-2}}{XT_{2-1} - XT_{0-2}} \times 100\%$$

$$\text{Deviation}_{2-1} = \frac{XP_{2-1} - XT_{2-1}}{XT_{2-1} - XT_{0-2}} \times 100\% \quad (9)$$

deviation_{0-2} represents the deviation rate of the pre-seizure onset sequence, where XP_{0-2} indicates the model-predicted position of the pre-seizure onset sequence, and XT_{0-2} represents the actual position of the pre-seizure onset sequence. deviation_{2-1} refers to the deviation rate of the pre-seizure termination sequence, where XP_{2-1} indicates the model-predicted position of the pre-seizure termination sequence, and XT_{2-1} represents the actual position of the pre-seizure termination sequence.

Results

In this study, we trained and evaluated the classification accuracy of the model by labeling each data point within a sequence with the same label as the first point, and treating the entire sequence as a single labeled entity. We divided the data into 80%, 5%, and 15% of the training, validation, and final test sets, respectively. The model is trained, validated, and tested on various data subsets. The set of statistical features of Xuanwu Hospital is shown in Supplementary Materials-Table S1.

It is specifically stated that both the tests for deviation rates and the final correlation analysis were conducted on additional data segments of the same type as the input dataset, which were not used during the training phase. All model training and testing were conducted in the following configuration: GPU is NVIDIA GeForce RTX 3090, CPU is Intel(R) Xeon(R) Gold 6226R (base frequency 2.90 GHz, 2-core processor), and RAM was 512 GB. Python 3.12 was used for compilation in the Anaconda environment.

Model training and validation

To assess model training and convergence, we employed metrics such as cross-entropy loss, mean absolute error (MAE), and mean squared error (MSE). Figure 7

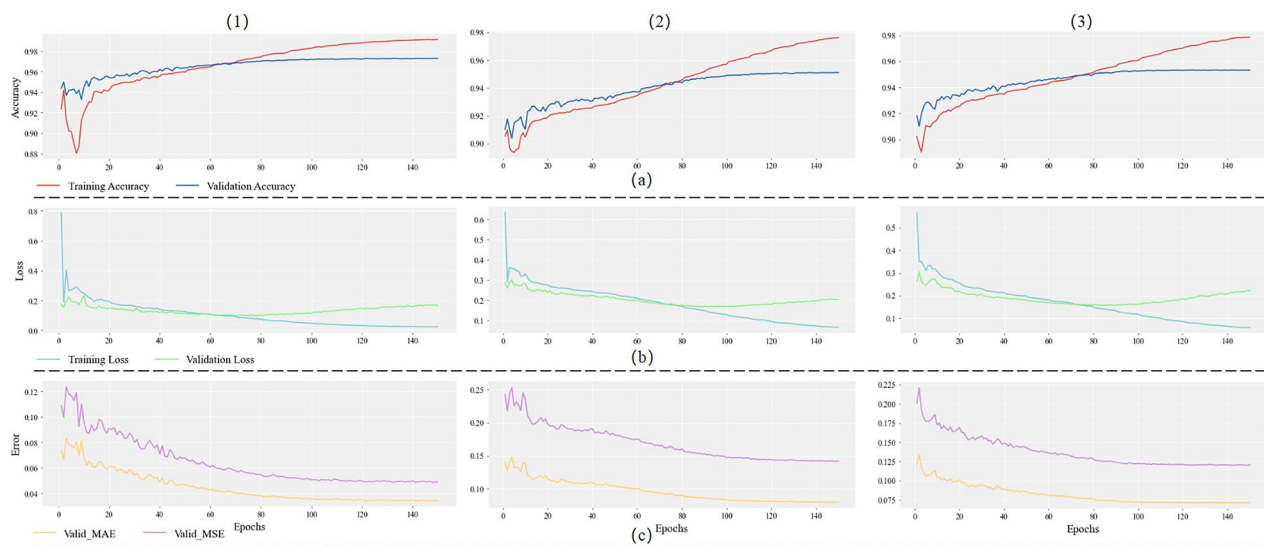


Fig. 7 Loss and Error. **a1** ECoG (Subset4) training accuracy, **a2** SEEG (Subset3) training accuracy, **a3** hybrid (Subset5) training accuracy. **b1** ECoG (Subset4) cross-entropy loss, **b2** SEEG (Subset3) cross-entropy loss, **b3** Hybrid (Subset5) cross-entropy loss. **c1** ECoG (Subset4) MAE and MSE, **c2** SEEG (Subset3) MAE and MSE, **c3** Hybrid (Subset5) MAE and MSE

illustrates the changes in the accuracy, cross-entropy loss, MSE, and MAE during the training and validation processes across the training epochs. Both the training and validation cross-entropy losses decreased progressively, indicating that the model classification performance improved over time for each dataset. To calculate the errors, this study directly used the predicted and true label values (both integers) rather than the mapping probabilities. The error exhibits minor fluctuations during training, which can be attributed to the fact that error calculations with discrete labels are less smooth than those with continuous probability-based errors. As the training progressed, the error gradually decreased and stabilized, indicating that the model's predictions became more accurate and the error converged.

Model testing

Confusion analysis

A heatmap was used to represent the numerical distribution of the confusion matrix (Fig. 8). As shown in the figure, the model performed well on both the ECoG (subset 4) and SEEG (subset 3) datasets, with most instances correctly classified. However, after removing the pre-policy network, the misclassification rate for class 0 increased significantly, with many instances of class 0 being incorrectly predicted as class 1 or class 2. This indicates that numerous interictal states were misidentified as preictal or ictal. This result strongly supports the necessity of a pre-policy network. By filtering out segments of interictal EEG containing interference, the model can more

accurately extract essential features related to resting-state brain rhythms.

Figure 9 illustrates the significant impact of the Pre-Policy Network in reducing the False Negative Rate (FNR). The False Positive Rate (FPR) exhibits relatively minor variations across different subsets, influenced by factors such as dataset complexity and the characteristics of intracranial EEG signals. Specifically, the higher temporal complexity of SEEG compared to ECoG leads to increased FPR and FNR in both SEEG and hybrid datasets. Notably, the comparison with Subset 5, which lacks the decision network, reveals a substantial increase in FNR. In the SEEG, ECoG, and Hybrid subsets, FNR remains relatively low, whereas in Subset 5 (Without Policy Network), FNR sharply rises to 12.78%, significantly higher than in other subsets. This indicates that the absence of the decision network makes the model more prone to misclassifying true positive samples as negative. Additionally, as shown in Fig. 8d, the model tends to misclassify seizure states as pre-seizure states, leading to a marked increase in FNR.

Deviation analysis

We used seven instances from the ECoG and SEEG test datasets based on the average channel deviation rate for detailed analysis. As shown in Fig. 10, the deviation rate in the ECoG data primarily leans towards the positive direction, indicating that the model tended to have a slight delay in detecting the preictal state of epilepsy. This does not necessarily mean that the model failed to detect the preictal state in a timely manner. Because we

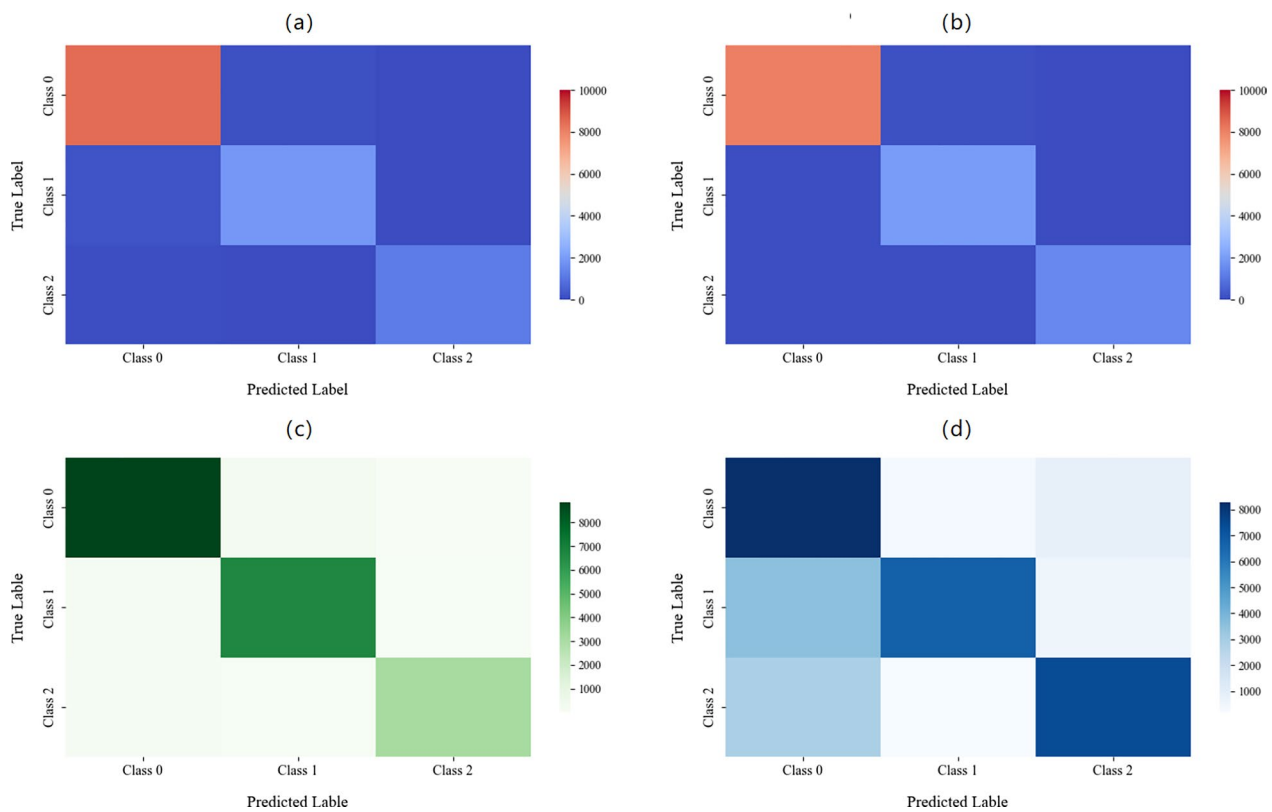


Fig. 8 Heatmap of the confusion matrix for subsets. **a** SEEG (Subset3), **b** ECoG (Subset4), **c** Hybrid (Subset5), **d** Subset5 (Without Policy Network). In this heatmap, the darker the color, the higher the proportion of correctly predicted instances in the nine categories. The cells along the diagonal represent the proportion of correctly predicted instances for each of the three categories

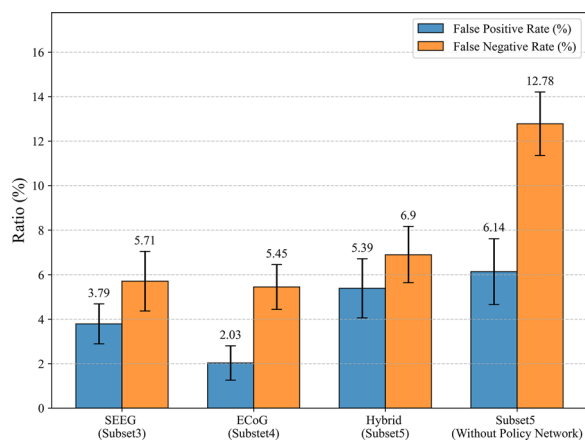


Fig. 9 The comparison of false positive rates and false negative rates across different data subsets: SEEG, ECoG, Hybrid, and Hybrid without the decision network

used uniform labels for all time points within the same sequence, if the transition point between states occurs in the latter half of the sequence, the sequence may place greater emphasis on interictal features than on preictal

ones. By contrast, the SEEG data showed a relatively higher overall deviation rate, with both positive and negative deviations distributed more evenly. Overall, the absolute deviation rate of the model remained below 6% for both datasets (see Fig. 11 for the overall statistics).

Classification accuracy of the test set

The overall prediction performance was evaluated using the accuracy, precision, F1 score, and recall. We conducted standardized testing of the model on each data subset. Table 2 presents the classification performance across different datasets. It can be observed that the model achieved optimal performance on the ECoG dataset. For all SEEG data containing GtcS, the model's ability to recognize the label '1', which indicates generalized tonic-clonic seizure states, was not significantly inferior to its performance on ECoG data. However, there was a slight decline in the performance of the overall dataset, likely due to the challenge of simultaneously handling multimodal data. This outcome was anticipated based on our analysis. Table 3 summarizes the performance differences between awake and sleep states in the FS dataset from Xuanwu Hospital. FS

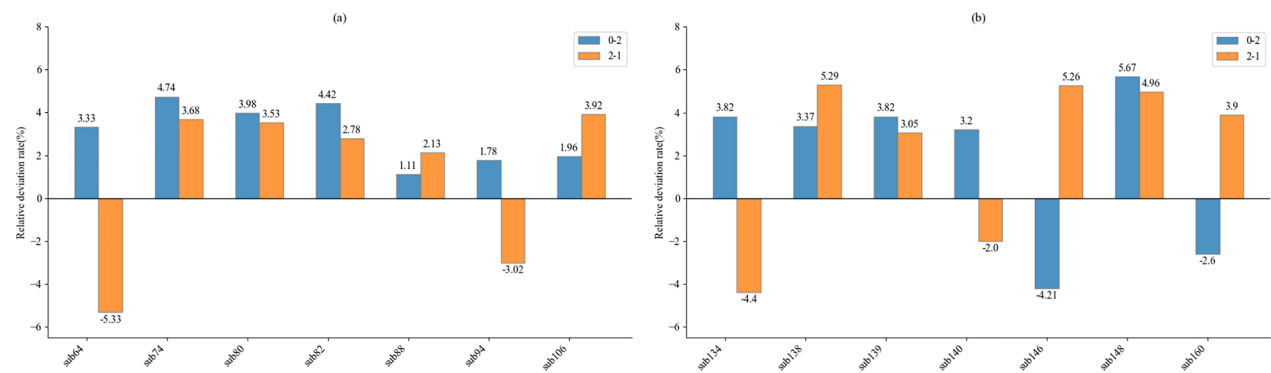


Fig. 10 Distribution of the average channel deviation rate for seven randomly selected patients. **a** Pre-seizure deviation rate in ECoG Subset4; **b** pre-seizure deviation rate in SEEG Subset3

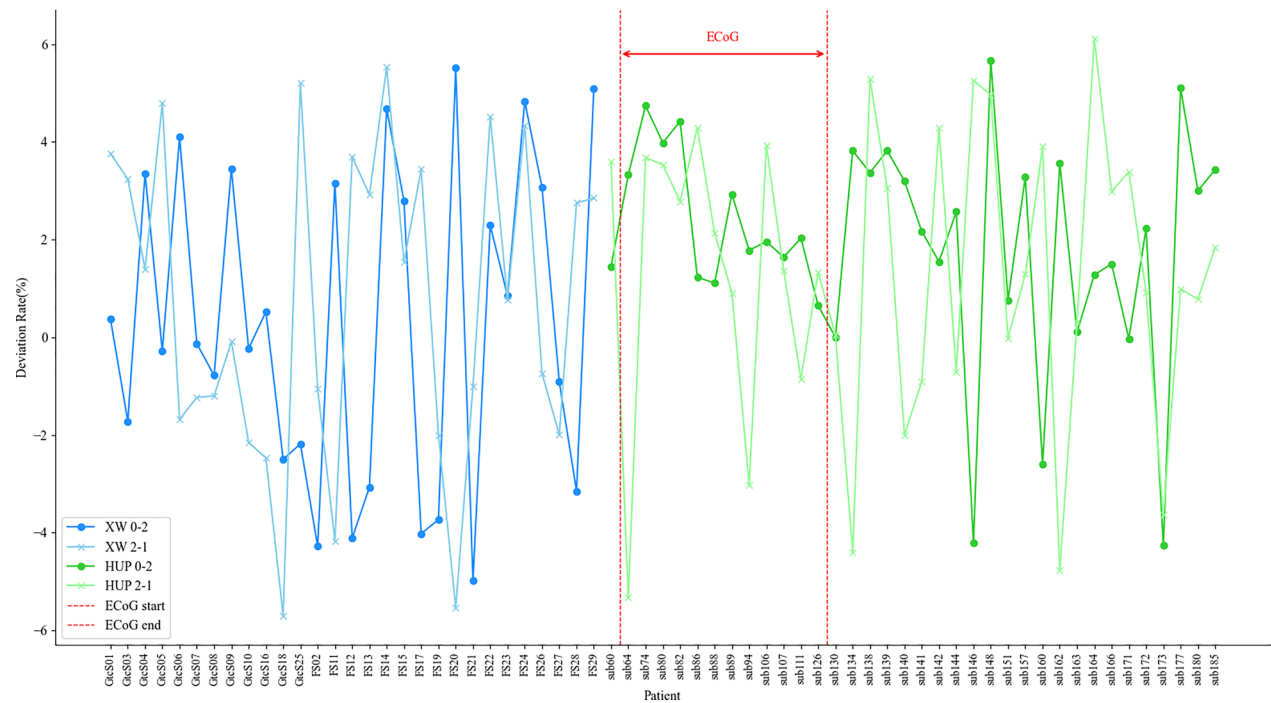


Fig. 11 Averaged deviation rate statistics for all patients with seizure channels, categorized by data source. The red dashed lines indicate the range of ECoG data, while the rest correspond to SEEG data

(subsets 1 and 2) is the focus of our analysis, as it highlights the classification performance of the model for small-scale focal seizures, an ability not present in other models. The FS during the awake state is more easily classified by the model, with a performance (94.1%) even surpassing that of the overall dataset (Subset 5). In contrast, the FS classification accuracy in the sleep EEG was the lowest among the five data subsets, although it approached 90%. Furthermore, detection of the preictal state remains at a high level. Notably, the recall rate for the seizure state did not significantly decline, indicating

that the model maintained a high sensitivity to FS during sleep.

For focal seizures, accurate labeling of specific waveforms such as spikes and high-frequency oscillations (HFOs) is essential during model training. However, most public datasets only annotate seizure intervals, primarily focusing on generalized seizures with prominent features. In contrast, our dataset from Xuanwu Hospital includes precise annotations of seizure characteristics—such as spikes, sharp waves, spike-and-slow waves, complex waves, and HFOs—down to the

Table 2 Model performance across diverse datasets

Class	Label	Accuracy	Recall	F1-Score	Precision (%)
Subset4	0	0.97±0.002	0.98±0.005	0.98±0.003	97.3±0.3
	1	0.97±0.001	0.92±0.001	0.94±0.001	
	2	0.98±0.004	0.94±0.005	0.96±0.003	
Subset3	0	0.97±0.005	0.96±0.001	0.96±0.003	95.1±0.4
	1	0.93±0.005	0.92±0.005	0.94±0.004	
	2	0.94±0.001	0.91±0.003	0.93±0.002	
Subset5	0	0.96±0.001	0.96±0.003	0.94±0.001	93.9±0.4
	1	0.85±0.001	0.94±0.005	0.86±0.004	
	2	0.92±0.002	0.87±0.002	0.89±0.003	

Table 3 Model performance on Xuanwu FS datasets

Class	Label	Accuracy	Recall	F1-Score	Precision (%)
Subset1	0	0.82±0.001	0.96±0.004	0.88±0.002	94.1±0.5
	1	0.96±0.002	0.93±0.001	0.94±0.004	
	2	0.97±0.005	0.94±0.002	0.96±0.001	
Subset2	0	0.87±0.001	0.99±0.005	0.89±0.002	89.3±0.6
	1	0.94±0.001	0.83±0.003	0.88±0.001	
	2	0.94±0.002	0.67±0.004	0.78±0.001	

Table 4 Comparison of the performance of different iEEG seizure detection methods on the same dataset

Method	Subset5			
	HUP iEEG		XuanWu SEEG	
	ACC (%)	AUC	ACC (%)	AUC
EEGNet	82.95±0.25	0.83	79.19±0.51	0.81
EEGWaveNet	87.19±0.39	0.86	83.26±0.47	0.83
CNN+BiLSTM	94.83±0.30	0.94	84.42±0.28	0.85
Deep ConvNet	95.28±0.11	0.97	86.37±0.57	0.85
GEM-CRAP	96.13±0.33	0.97	91.91±0.50	0.93

millisecond level across individual channels. The HUP iEEG dataset mainly comprises generalized seizures, and to date, no other seizure detection algorithms have utilized this dataset for training and validation. Therefore, we employed several classical seizure detection models to conduct validation tests under the same dataset and annotations. As shown in Table 4, GEM-CRAP's performance on datasets with prominent generalized seizures is nearly identical to that of Deep ConvNet. However, on the Xuanwu Hospital SEEG dataset, which includes focal seizure characteristics, GEM-CRAP significantly outperforms other

Table 5 Model performance without pre-policy network activation

Class	Label	Accuracy	Recall	F1-Score	Precision (%)
Subset4	0	0.94±0.004	0.90±0.001	0.92±0.002	87.9±0.4
	1	0.88±0.002	0.74±0.002	0.80±0.003	
	2	0.61±0.005	0.91±0.001	0.73±0.004	
Subset3	0	0.95±0.005	0.91±0.001	0.93±0.003	89.5±0.2
	1	0.86±0.001	0.82±0.001	0.84±0.001	
	2	0.59±0.001	0.90±0.004	0.71±0.005	
Subset5	0	0.88±0.003	0.95±0.005	0.91±0.004	85.9±0.1
	1	0.84±0.001	0.60±0.004	0.70±0.002	
	2	0.59±0.001	0.38±0.007	0.47±0.003	

network models, indicating its heightened sensitivity and improved recognition of short-duration focal seizure features.

Ablation experiment

We conducted comparative training with the pre-policy network for both enabled and disabled networks. Before entering the input gate of the RNN, we deactivated the activation function of the prepolicy network, allowing the entire EEG time-series input to flow directly into the next network layer. The batch size was set to 512 and the initial learning rate was set to 0.001. Additionally, the decay factor of the adaptive learning rate was modified to 0.98 to accommodate the processing of larger batches of EEG data. In Table 5, the classification performance of the model shows a significant decline across all metrics, with particularly poor results for the overall dataset (Subset 5). This suggests that as the data complexity increases, the positive impact of the prepolicy network on the model becomes more pronounced, leading to better gradient descent optimization.

A comparison of the channel accuracies before and after the reinforcement is shown in Fig. 12. Statistical analysis indicated that the average search time for each group of SVM network parameters was 12.7 s, and the average classification time for the test set was 1.12 s. The average channel accuracy after reinforcement improved by approximately 11% compared to that of the main model, reaching approximately 86%.

Correlation analysis between channel accuracy and temporal distribution of seizure states

We also analyzed the time distribution of seizure states across all seizure channels in the HUP iEEG dataset used during the model training phase, as well as the distribution of channel accuracy from the model predictions on additional test segments from the same type and the

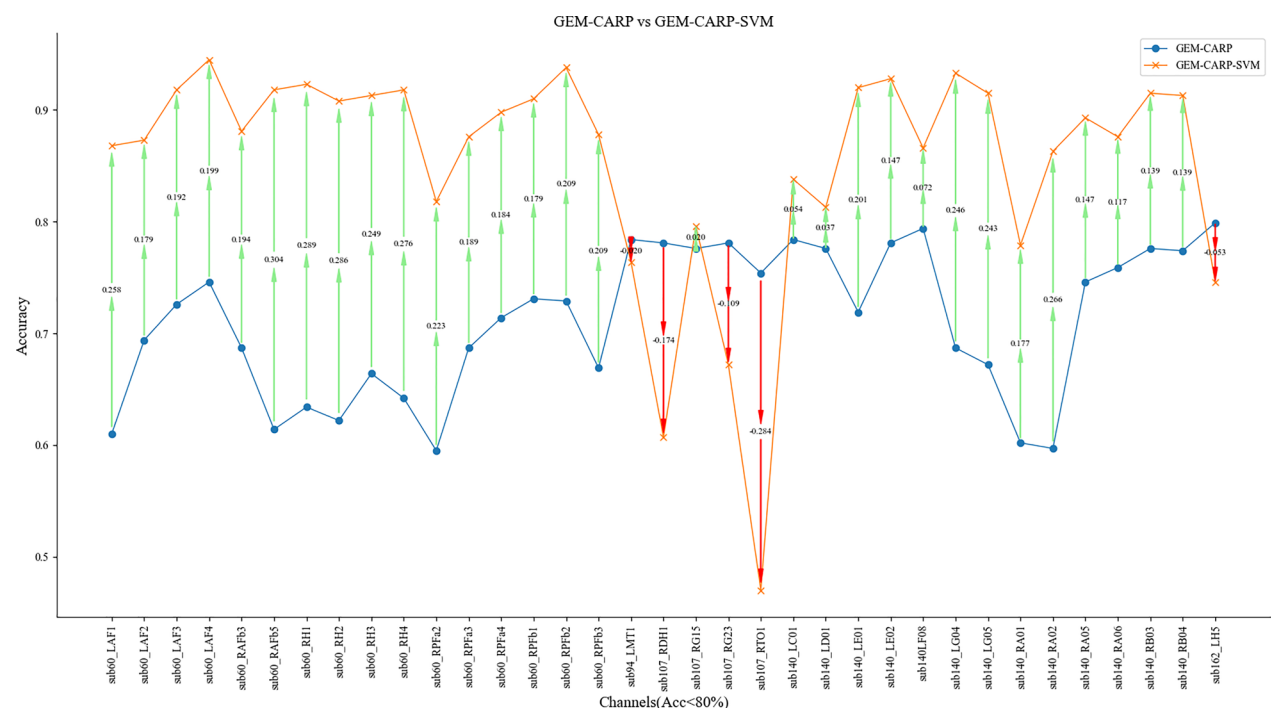


Fig. 12 The comparison of accuracy and performance improvement between the GEM-CRAP main model and the SVM-enhanced training for channels with an accuracy lower than 80%. As shown in the figure, the SVM-enhanced training achieves performance improvements over the main model for most channels

same patient. Table 6 summarizes the detailed distribution of channel accuracy for all patients in the HUP iEEG dataset, as well as the standard deviation of the seizure state time distribution for each patient in the dataset (not shown in the table, it indicates that only channels containing interictal periods were used for that patient).

The time distribution for different patients across the three states during model training showed significant variation, reflecting individual differences among patients and the diversity of epileptic seizures. Box plot illustrating the distribution of channel accuracy for predicting seizure states across different patients. As shown in Fig. 13b, the prediction accuracy for sub177 was concentrated and at a high level, whereas the accuracy for sub180 and sub185 was more dispersed and lower, indicating that the channel signals from certain patients were easier for the model to classify correctly. Correspondingly, in Fig. 13d, the time distribution of the three seizure states for sub177 was more balanced than those for sub180 and sub185. Similarly, we obtained comparable results for the ECoG data, as shown in Fig. 13a and c. To analyze this potential underlying relationship more comprehensively, we calculated the distribution information of the channel accuracy for all patients in the HUP iEEG dataset, including the maximum, minimum, standard deviation, upper and lower quartiles (Q3 and Q1),

and median (Q2) of the channel accuracy, as shown in Table 6. This table also includes the standard deviation of the seizure state time distribution for each patient in the dataset. We conducted a correlation analysis between the standard deviation of the input HUP dataset time distribution during epileptic seizures and that of the model-predicted channel accuracy distribution and found a notable positive correlation between the two. As shown in Table 7, in the SEEG data, the Pearson correlation coefficient was 0.703 ($p < 0.05$) and the Spearman correlation coefficient was 0.855 ($p < 0.05$), indicating a significant linear and monotonic relationship between the input time distribution and channel accuracy. For the ECoG data in Table 8, the Pearson correlation coefficient was 0.954 ($p < 0.05$), and the Spearman correlation coefficient was 0.896 ($p < 0.05$), suggesting an even stronger correlation in the ECoG data.

Discussion

Automated seizure detection can significantly alleviate the workload of clinicians by continuously monitoring EEG recordings and enabling the early diagnosis of epilepsy [26]. Deep learning-based EEG processing and analysis techniques facilitate more efficient and accurate medical decision making. This study proposes a supervised multi-layer hybrid learning model based on

Table 6 Seizure state time and channel accuracy distribution

Patient	Time Variance	Channel accuracy					
		Variance	Minimum	Maximum	Q1	Q2	Q3
sub60	0.3872	0.1007	0.5945	0.9453	0.6399	0.6903	0.7351
sub64	0.0950	0.0262	0.8632	0.9801	0.9502	0.9577	0.9677
sub74	0.1580	0.0331	0.7606	0.9701	0.9177	0.9377	0.9576
sub80	0.0710	0.0130	0.9328	1.0000	0.9726	0.9826	0.9857
sub82	0.1160	0.0242	0.8881	0.9950	0.9378	0.9652	0.9757
sub86	0.0870	0.0123	0.9204	1.0000	0.9801	0.9876	0.9925
sub88	0.0500	0.0058	0.9701	0.9975	0.9876	0.9925	0.9950
sub89	0.1370	0.0231	0.9080	0.9876	0.9496	0.9602	0.9701
sub94	0.2620	0.0572	0.7413	0.9801	0.8483	0.9043	0.9403
sub106	0.2080	0.0431	0.7512	0.9652	0.9179	0.9279	0.9502
sub107	0.2820	0.0775	0.7537	0.9925	0.8607	0.9552	0.9876
sub111	0.1250	0.0199	0.9254	0.9950	0.9726	0.9864	0.9900
sub126	0.1870	0.0341	0.8625	0.9700	0.9100	0.9400	0.9619
sub130	0.1330	0.0201	0.9204	0.9975	0.9751	0.9814	0.9851
sub134	0.0687	0.0134	0.9254	1.0000	0.9900	0.9950	0.9975
sub138	0.2068	0.0435	0.4801	1.0000	0.9577	0.9776	0.9900
sub139	0.0526	0.0079	0.9627	0.9975	0.9876	0.9950	0.9950
sub140	0.4032	0.0950	0.5970	1.0000	0.8321	0.9117	0.9733
sub141	0.0210	0.0039	0.9851	0.9975	0.9900	0.9925	0.9944
sub142	0.0790	0.0150	0.9428	0.9950	0.9671	0.9876	0.9900
sub144	0.1370	0.0268	0.8781	0.9975	0.9602	0.9801	0.9900
sub146	0.0491	0.0054	0.9652	1.0000	0.9900	0.9950	0.9950
sub148	0.5698	0.0087	0.9552	1.0000	0.9925	0.9975	1.0000
sub151	0.0580	0.0076	0.9701	0.9950	0.9826	0.9851	0.9925
sub157	0.0290	0.0137	0.9652	1.0000	0.9707	0.9776	0.9975
sub160	0.0674	0.0097	0.9204	0.9975	0.9876	0.9925	0.9950
sub162	0.3200	0.0566	0.7985	0.9925	0.9080	0.9627	0.9826
sub163	0.0820	0.0111	0.9055	1.0000	0.9963	1.0000	1.0000
sub164	0.3020	0.0182	0.9030	0.9801	0.9428	0.9527	0.9621
sub166	0.3090	0.0150	0.9428	0.9876	0.9577	0.9751	0.9814
sub171	0.0660	0.0079	0.9701	1.0000	0.9876	0.9925	0.9975
sub172	0.0540	0.0070	0.9552	1.0000	0.9882	0.9925	0.9969
sub173	0.0660	0.0090	0.9627	1.0000	0.9851	0.9900	0.9950
sub177	0.0960	0.0150	0.9328	0.9975	0.9776	0.9826	0.9919
sub180	0.2160	0.0392	0.8507	0.9726	0.9074	0.9266	0.9496
sub185	0.2370	0.0430	0.8184	0.9229	0.8483	0.9154	0.9179

deep learning, which was validated and analyzed using ECoG and SEEG datasets from the Hospital of the University of Pennsylvania and Xuanwu Hospital of Capital Medical University, demonstrating its excellent performance. It incorporates three parallel feature extraction channels: a CNN for frequency-domain distribution, an RNN for time-domain correlation, and amplitude variation. Additionally, a decision network layer based on LSTM was integrated as a precursor to the time-domain RNN, working with the input gate for data filtering and

enhancement. Furthermore, a hybrid cross-entropy loss reward mechanism was added to the main model training to provide feedback on the performance of the decision network, facilitating strategy updates and continuous model optimization. Finally, a post-training SVM network was included to perform secondary reinforcement training for a small subset of channels with a classification accuracy below 80%. The reliability of a model is determined by the quantity and quality of data included in the analysis. Constrained

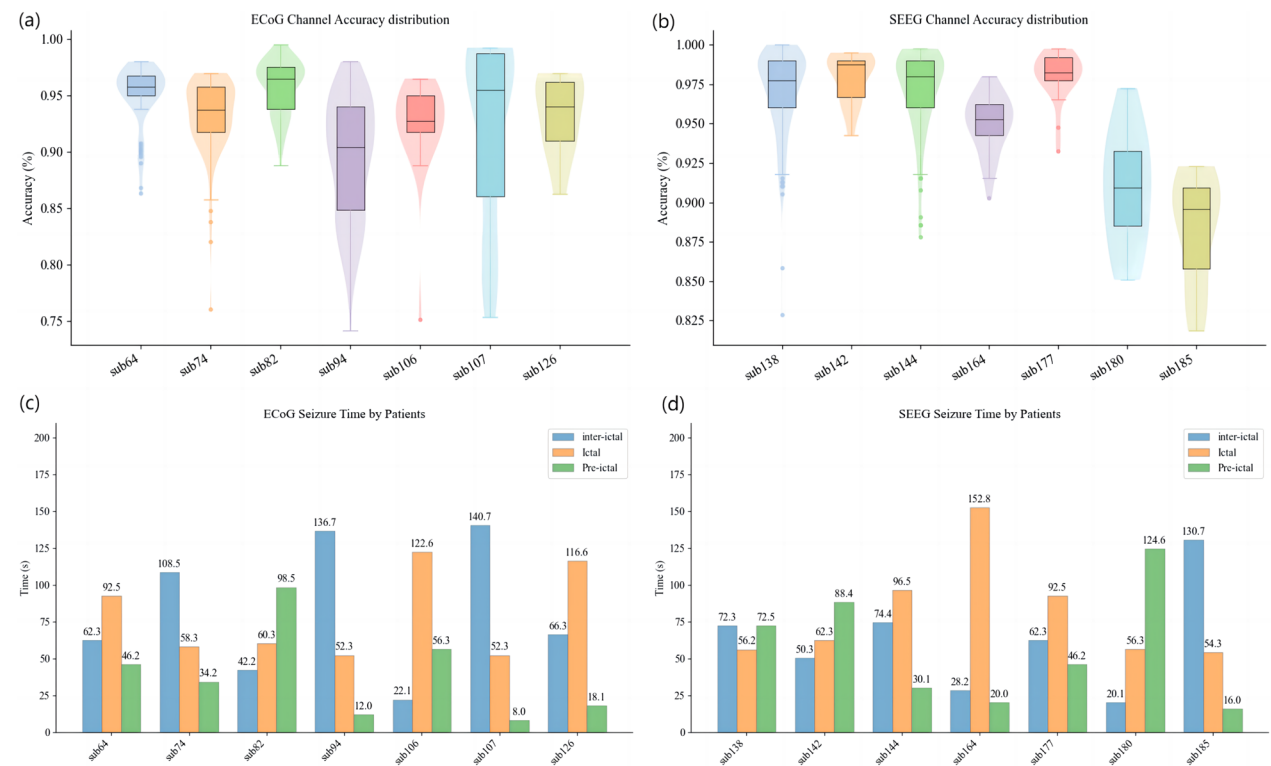


Fig. 13 Seizure state time distribution and channel accuracy distribution for seven randomly selected patients across different dataset types. **a** Channel accuracy distribution in the ECoG Subset4; **b** channel accuracy distribution in the SEEG Subset3; **c** seizure state time distribution in the ECoG Subset4; **d** seizure state time distribution in the SEEG Subset3

Table 7 Correlation between SEEG seizure time Std and channel Acc Std

Metric	Value
Pearson correlation coefficient	0.702579399
Pearson P-value	0.001660258
Spearman correlation coefficient	0.855039436
Spearman P-value	1.22E-05

Table 8 Correlation between ECoG seizure time Std and channel Acc Std

Metric	Value
Pearson correlation coefficient	0.954274759
Pearson P-value	4.16644E-07
Spearman correlation coefficient	0.895604396
Spearman P-value	3.48E-05

by the number of patients and channels, as well as noise interference in scalp EEG, many models fail to fully demonstrate their generalization capability and accuracy [27].

Our study utilized data from patients with DREs at the Hospital of the University of Pennsylvania's HUP iEEG dataset and extensive patients from Xuanwu Hospital, covering the full spectrum of pre-ictal, ictal, inter-ictal, and post-ictal EEG data. These datasets were revalidated by experienced neurosurgeons to identify seizure-onset channels and channels overlapping strictly segmented resection/ablation areas. Although the original datasets included extensive inter-ictal records, the EEG characteristics in these data primarily reflected individual EEG differences and diversity, which did not significantly enhance the generalization of the model in epilepsy recognition. Therefore, we combined the electrode location information and seizure onset channel distribution to filter out records with the most comprehensive usable electrode information and minimal interference. For each patient's non-seizure onset channels, which experienced seizures later than the onset channels and gradually spread to other channels, we classified the whole-brain records based on different seizure onset times and saved channels with similar onset times in the same EDF file. This approach saves time extracting individual channels and refines and organizes the dataset more efficiently. Notably, our model achieved a classification accuracy

of over 94% for small-scale seizures occurring during wakefulness, which were prevalent in the Xuanwu Hospital dataset, and an accuracy of approximately 90% during sleep. Consequently, this study advances intelligent detection of pre-epileptic seizures toward more generalized and widespread applications, extending beyond large-scale generalized seizures. This study provides a potential AI algorithm for future closed-loop neurostimulation therapies.

The sequence length affects the model's performance, generalization capability, and ability to capture the temporal dependencies of EEG data, making it a critical parameter for constructing neural network models. For epilepsy seizure detection, the sequence length must balance the dynamic changes in the seizure duration and the temporal characteristics of the EEG signals [28, 29]. When using traditional RNNs, processing long sequences increases the computational burden and risks gradient vanishing or exploding, thereby affecting stability and efficiency [30]. Considering our dataset's 500 Hz sampling rate and the need for real-time seizure state detection, we set the sequence length to 250, corresponding to a 0.5-s detection window, ensuring sufficient temporal resolution and manageable computational load. The data were reshaped into three-dimensional feature vectors (batch size, sequence length, and features). Initially, the feature dimensions were set to 1. This additional dimension allows the network architecture to parse different amplitude components or energies from various frequency components, thereby enhancing the model's ability to recognize complex dynamic features. Additionally, we introduced a fast Fourier transform (FFT) layer and an amplitude-aware layer to capture the amplitude and frequency characteristics of the EEG signals [31]. The FFT Layer converts the time-series data into the frequency domain and takes its absolute value, thereby enhancing the processing efficiency. The amplitude-aware layer analyzes voltage differences in the EEG data points to capture dynamic amplitude variations, which are crucial for detecting epileptic seizures. Wu et al. [9] attempted to directly input a multichannel time series of signals based on the time or frequency domain into a CNN. Building on this, we incorporated spatial distribution information and the propagation process of seizure onset channels and performed detailed channel segmentation and extraction. RNNs are well suited for processing time-series data and effectively handling sequence dependencies through internal state propagation. We integrated a Policy Network as a preprocessing RNN variant using LSTM to evaluate each time step and optimize the model performance.

Researchers have proposed various methods for detecting epileptic seizures to enhance the detection accuracy and efficiency (Supplementary Materials-Table S2). For instance, Acharya et al. [32] developed an automatic seizure detection approach based on deep CNNs. This method achieved classification accuracies of 88.7%, 90%, and 95% for the normal, intermittent, and seizure EEG signals, respectively, on the Bonn dataset. Hu et al. [33] employed a bidirectional long short-term memory network (Bi-LSTM) combined with local mean decomposition (LMD) for seizure detection, attaining mean sensitivity and specificity of 93.61% and 91.85%, respectively, for the CHB-MIT dataset. This approach leverages the bidirectional information flow to enhance the model's ability to capture temporal dependencies. Furthermore, Martis et al. [34] utilized empirical mode decomposition (EMD) and intrinsic mode functions (IMF) for feature extraction and classified them using a classification and regression tree (CART), achieving an accuracy of 93.55%. This demonstrates that traditional feature extraction methods, when combined with appropriate classifiers, can still offer a relatively efficient detection performance. Recent advances in automated seizure detection have leveraged deep learning to improve both performance and interpretability. Ma et al. [35] proposed TSD, a Transformer-based model that processes time-frequency EEG features and achieves an AUROC of approximately 92.1% on the TUH dataset, demonstrating robust seizure detection across various types. Similarly, Einizade et al. [36] developed a hybrid CNN-RNN architecture for detecting both generalized and focal seizures, achieving an accuracy of 82%, a precision of 71.7%, and a sensitivity of 85% on a heterogeneous EEG dataset. In parallel, Wong et al. [37] introduced a channel-annotated deep learning framework that combines a shallow 1D-CNN Transformer with an ensemble MLP, yielding a state-of-the-art AUC of 0.93 on public EEG data and an AUC of 0.82 on external datasets, while enhancing interpretability through channel-level explanations. Collectively, these studies significantly contribute to the field by balancing high detection performance with improved generalizability and interpretability. Models combining CNNs and long short-term memory networks (LSTMs) have demonstrated good performance for handling complex temporal information [28]. Xu et al. [38] developed a model that achieved a recognition accuracy of 82% for five-class epileptic seizure recognition tasks. Similarly, Liu's approach, which also employed this combination, showed good performance in the classification of long-term EEG signals, further validating the advantages of deep learning methods in epilepsy detection [39].

In practice, when dealing with complex high-dimensional EEG data, the difficulty of gradient descent caused by data redundancy has long been a challenge for epilepsy detection models, often affecting classification accuracy and processing time [40]. This raises several critical questions. Which time series are useful for model classification? Which sequences lack informative features and introduce noise into the classification process? For example, physical artifacts or high-frequency, high-amplitude rhythmic waves present in interictal EEG are sometimes mistakenly interpreted by network models as seizure states. Based on previous studies and practical experimentation, we found that for EEG data mapped onto a high-dimensional feature space, it is unnecessary for the network model to integrate information from all dimensions [41, 42]. For iEEG data from patients with epilepsy, the proposed CNN-RNN architecture, which incorporates a strategy network and residual encoding blocks, can effectively filter data and reduce redundant or noisy information. The proportion of time spent in each seizure state within a single channel also had a notable impact on the model classification performance. By analyzing the distribution of time and channel-specific accuracy in this study, we identified the optimal time distribution ratios for model predictions. Training the model using patient data with more evenly distributed time segments improves the precision of neural network classification of epileptic EEG patterns. Additionally, the model was trained and its classification accuracy was computed by treating each data point within a sequence as labeled using the feature label of the first point to label the entire sequence. The training and convergence of the model were evaluated using metrics such as the cross-entropy loss, mean absolute error, and mean squared error. The overall predictive performance was assessed based on the accuracy, precision, F1 score, and recall. Furthermore, we introduced an innovative evaluation metric, deviation rate, to specifically assess model performance in predicting the onset of epileptic seizures, ensuring accurate detection and high-frequency stimulation during the early stages of a seizure. The deviation rate in this study was calculated based on the temporal window misalignment rate between predicted seizure onset/offset times and expert annotations (aligned with real-time video-EEG monitoring). This metric quantifies the model's susceptibility to EEG artifacts (accounting for approximately 20–30% of interference, such as EMG artifacts and motion-induced noise). Additionally, seizure heterogeneity (e.g., focal vs. generalized onset) can lead to performance variability across seizure subtypes. For instance, the typical low-frequency discharges (4–7 Hz) observed in temporal lobe epilepsy

may be more readily captured by the model compared to high-frequency discharges (> 10 Hz) in frontal lobe seizures. Including the deviation rate also helps to identify detection blind spots for specific seizure subtypes. This new metric provides a more precise measure of the model detection capabilities during the seizure onset phase, significantly enhancing the practicality and effectiveness of the system.

In contrast to previous epilepsy detection algorithms that employed datasets in identical data formats, our study revealed that supervised learning models can discern distinct structural features even under identical labeling, and that classification is a weighted fusion of these features. The SEEG and ECoG data structures as well as the seizure characteristics were significantly different. While learning from mixed datasets, neural networks can extract various features distributed across different sequences [41]. Although the classification performance declined compared to training on homogeneous datasets, the model maintained relatively satisfactory results. Through multi-layer nonlinear transformations, our model can extract high-level discriminative features from complex raw data, even when these features manifest differently across diverse data structures.

Our model has some limitations. First, the sequence length is fixed; the model sets each sequence length to 250, which may restrict the flexibility in handling sequences of varying lengths. If the sequence length changes, then the model may require adjustments or modifications. Second, we considered only the absolute values of the magnitudes, without incorporating the phase information of the frequency components. Phase data may hold the potential for temporal feature extraction. Third, the model lacks an explicit random seed setting. The code does not fix a random seed, leading to slight variations in the results for each run. Although this issue does not substantially affect most experiments or model performances, it may need to be addressed in cases requiring fully reproducible results. The current study primarily utilizes data from Xuanwu Hospital and HUP iEEG dataset for algorithm development and validation. While they provide reliable benchmarks for seizure detection, the limited sample size may affect the model's generalizability across broader populations. In subsequent work, we plan to enhance model robustness across cross-device and cross-protocol scenarios by integrating multi-center heterogeneous data resources, including public datasets like TUSZ by Temple University, as well as proprietary data from clinical partners. Finally, our current model lacks advanced explainability algorithms, such as SHAP. This integration challenge restricts our ability to provide detailed interpretability of the model's predictions. Our future research will focus on

incorporating these explainability techniques to enhance the transparency and understanding of our model's decision-making process.

Conclusion

Compared to other algorithms, GEM-CRAP more accurately identifies key seizure characteristics of focal epilepsy. Through adaptive adjustments and attention mechanisms, it improves performance in complex signal environments, achieving higher precision and robustness in seizure detection. These advancements not only improve seizure interval detection but also enhance the identification and analysis of specific epileptic waveforms, such as HFOs, paving the way for more precise and individualized epilepsy diagnostics and treatments.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-025-06414-5>.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Author contributions

G.Z., Y.S., and Y.Z. designed the study. J.S. and Z.S. wrote the manuscript and conducted the analysis. Z.S., H.X., and Y.Z. provided statistical guidance. Z.S., H.X., J.S., L.J., Y.Y., H.D., and Z.L. conducted the experiments. G.Z., Y.S., and P.W. provided advice on the discussion. The manuscript was completed under the supervision of G.Z., Y.S., and Y.Z.

Funding

Supported by the Beijing Municipal Science & Technology Commission (Z241100009024058), Beijing Municipal Health Commission (Grant No. 2022-2-2011), and Beijing Hospitals Authority Clinical Medicine Development of special funding support (code: ZLRK202319).

Availability of data and materials

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

Declarations

Ethics approval and consent to participate

This use of these data was approved by the local ethics committee of the Xuanwu Hospital, Capital Medical University, Beijing, China. The ethics committee's phone, email, and address as follows: 0086-10-83919270, xwkyethics@163.com, and No. 45, Changchun Street, Xicheng District, Beijing 100053, China. All the participants provided written informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests in relation to this work.

Author details

¹Department of Neurosurgery, Xuanwu Hospital, Capital Medical University, Beijing, China. ²China International Neuroscience Institute (China-INI), Beijing,

China. ³Clinical Research Center for Epilepsy, Xuanwu Hospital, Capital Medical University, Beijing, China.

Received: 8 November 2024 Accepted: 24 March 2025

Published online: 05 April 2025

References

- Asadi-Pooya AA, Brigo F, Lattanzi S, Blumcke I. Adult epilepsy. *Lancet* (London, England). 2023;402:412–24. [https://doi.org/10.1016/s0140-6736\(23\)01048-6](https://doi.org/10.1016/s0140-6736(23)01048-6).
- Löscher W, Klitgaard H, Twyman RE, Schmidt D. New avenues for anti-epileptic drug discovery and development. *Nat Rev Drug Discov*. 2013;12:757–76. <https://doi.org/10.1038/nrd4126>.
- Li K, et al. Stereo-electroencephalography-guided three-dimensional radiofrequency thermocoagulation for mesial temporal lobe epilepsy with hippocampal sclerosis: a retrospective study with long-term follow-up. *Epilepsia open*. 2023. <https://doi.org/10.1002/epi4.12866>.
- Shi J, et al. The role of neuroinflammation and network anomalies in drug-resistant epilepsy. *Neurosci Bull*. 2025. <https://doi.org/10.1007/s12264-025-01348-w>.
- Shi J, et al. Applications of diffusion tensor imaging integrated with neuronavigation to prevent visual damage during tumor resection in the optic radiation area. *Front Oncol*. 2022;12: 955418. <https://doi.org/10.3389/fonc.2022.955418>.
- Shi J, Bo F, Pan R, Zhang Y, Xu Y, Chen H, Ge H, Cao X, Xia Y, Luo Z. Research on the Relationship Between Meckel's Cavity Shape, Balloon Shape, and Intracapsular Pressure During Percutaneous Balloon Compression. *World Neurosurg*. 2022 Dec;168:e369–e375. <https://doi.org/10.1016/j.wneu.2022.10.033>.
- Benbadis SR, Beniczky S, Bertram E, MacIver S, Moshé SL. The role of EEG in patients with suspected epilepsy. *Epileptic Disord Int Epilepsy J Videotape*. 2020;22:143–55. <https://doi.org/10.1684/epd.2020.1151>.
- Saez I, Gu X. Invasive computational psychiatry. *Biol Psychiat*. 2023;93:661–70. <https://doi.org/10.1016/j.biopsych.2022.09.032>.
- Zhou M, et al. Epileptic seizure detection based on EEG signals and CNN. *Front Neuroinform*. 2018;12:95. <https://doi.org/10.3389/fninf.2018.00095>.
- Zhang H, et al. The applied principles of EEG analysis methods in neuroscience and clinical neurology. *Mil Med Res*. 2023;10:67. <https://doi.org/10.1186/s40779-023-00502-7>.
- Muhammad Usman S, Khalid S, Bashir S. A deep learning based ensemble learning method for epileptic seizure prediction. *Comput Biol Med*. 2021;136: 104710. <https://doi.org/10.1016/j.combiomed.2021.104710>.
- Najafi T, Jaafar R, Remli R, Wan Zaidi WA. A classification model of EEG signals based on RNN-LSTM for diagnosing focal and generalized epilepsy. *Sensors* (Basel, Switzerland). 2022. <https://doi.org/10.3390/s22197269>.
- Wen Y, et al. A 65nm/0.448 mW EEG processor with parallel architecture SVM and lifting wavelet transform for high-performance and low-power epilepsy detection. *Comput Biol Med*. 2022;144: 105366. <https://doi.org/10.1016/j.combiomed.2022.105366>.
- Li Z, Wei C, Zhang Z, Han L. ecGBMsub: an integrative stacking ensemble model framework based on eccDNA molecular profiling for improving IDH wild-type glioblastoma molecular subtype classification. *Front Pharmacol*. 2024;15:1375112. <https://doi.org/10.3389/fphar.2024.1375112>.
- Li Z, et al. A three-stage eccDNA based molecular profiling significantly improves the identification, prognosis assessment and recurrence prediction accuracy in patients with glioma. *Cancer Lett*. 2023;574: 216369. <https://doi.org/10.1016/j.canlet.2023.216369>.
- Siddiqui MK, Morales-Menendez R, Huang X, Hussain N. A review of epileptic seizure detection using machine learning classifiers. *Brain informatics*. 2020;7:5. <https://doi.org/10.1186/s40708-020-00105-1>.
- Siddiqui MK, Huang X, Morales-Menendez R, Hussain N, Khattoon K. Machine learning based novel cost-sensitive seizure detection classifier for imbalanced EEG data sets. *Int J Interact Des Manuf (IJIDeM)*. 2020. <https://doi.org/10.1007/s12008-020-00715-3>.
- Siddiqui MK, Islam MZ, Kabir MA. Analyzing performance of classification techniques in detecting epileptic seizure. *Adv Data Min App*. 2017. https://doi.org/10.1007/978-3-319-69179-4_27.

19. Siddiqui MK, Islam MZ, Kabir MA. A novel quick seizure detection and localization through brain data mining on ECoG dataset. *Neural Comput App*. 2019. <https://doi.org/10.1007/s00521-018-3381-9>.
20. Siddiqui MK. Brain data mining for epileptic seizure-detection. Charles Sturt University; 2018. <https://researchoutput.csu.edu.au/en/publications/brain-data-mining-for-epileptic-seizure-detection>.
21. Fatma N, Singh P, Siddiqui MK. Epileptic seizure detection in EEG signal using optimized convolutional neural network with selected feature set. *Int J Artif Intell Tools*. 2023. <https://doi.org/10.1142/s0218213023500458>.
22. Yang Y, et al. Early assessment of responsive neurostimulation for drug-resistant epilepsy in China: a multicenter, self-controlled study. *Chin Med J*. 2025;138:430–40. <https://doi.org/10.1097/cm9.0000000000003292>.
23. Abibullaev B, Zollanvari A. Learning discriminative spatio-spectral features of ERPs for accurate brain-computer interfaces. *IEEE J Biomed Health Inform*. 2019;23:2009–20. <https://doi.org/10.1109/jbhi.2018.2883458>.
24. Bashashati A, Fatourehchi M, Ward RK, Birch GE. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *J Neural Eng*. 2007;4:R32–57. <https://doi.org/10.1088/1741-2560/4/2/r03>.
25. Shoeibi A, et al. Epileptic seizures detection using deep learning techniques: a review. *Int J Environ Res Public Health*. 2021. <https://doi.org/10.3390/ijerph18115780>.
26. Ayman U, et al. Epileptic patient activity recognition system using extreme learning machine method. *Biomedicines*. 2023. <https://doi.org/10.3390/biomedicines11030816>.
27. Huo Q, Luo X, Xu ZC, Yang XY. Machine learning applied to epilepsy: bibliometric and visual analysis from 2004 to 2023. *Front Neurol*. 2024;15:1374443. <https://doi.org/10.3389/fneur.2024.1374443>.
28. He C, Liu J, Zhu Y, Du W. Data AUGMENTATION for deep neural networks model in EEG classification task: a review. *Front Hum Neurosci*. 2021;15:765525. <https://doi.org/10.3389/fnhum.2021.765525>.
29. Taran S, Bajaj V, Siuly S. An optimum allocation sampling based feature extraction scheme for distinguishing seizure and seizure-free EEG signals. *Health Inf Sci Syst*. 2017;5:7. <https://doi.org/10.1007/s13755-017-0028-7>.
30. Majumdar A, Gupta M. Recurrent transform learning. *Neural Netw*. 2019;118:271–9. <https://doi.org/10.1016/j.neunet.2019.07.003>.
31. Yang J, et al. Channel selection and classification of electroencephalogram signals: an artificial neural network and genetic algorithm-based approach. *Artif Intell Med*. 2012;55:117–26. <https://doi.org/10.1016/j.artmed.2012.02.001>.
32. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Comput Biol Med*. 2018;100:270–8. <https://doi.org/10.1016/j.combiomed.2017.09.017>.
33. Hu X, et al. Scalp EEG classification using deep Bi-LSTM network for seizure detection. *Comput Biol Med*. 2020;124: 103919. <https://doi.org/10.1016/j.combiomed.2020.103919>.
34. Martis RJ, et al. Application of empirical mode decomposition (emd) for automated detection of epilepsy using EEG signals. *Int J Neural Syst*. 2012;22:1250027. <https://doi.org/10.1142/s012906571250027x>.
35. Ma Y et al. TSD: transformers for seizure detection; 2023. <https://doi.org/10.1101/2023.01.24.525308>.
36. Einizade A, Mozafari M, Sardouie SH, Nasiri S, Clifford G. In: 2020 IEEE signal processing in medicine and biology symposium (SPMB); 2020. p. 1–6.
37. Wong S, Simmons A, Rivera-Villicana J, Barnett S, Sivathamboo S, Perucca P, Ge Z, Kwan P, Kuhlmann L, O'Brien TJ. Channel-annotated deep learning for enhanced interpretability in EEG-based seizure detection. *Biomed Signal Process Control*. 2024. <https://doi.org/10.1016/j.bspc.2024.107484>.
38. Xu G, Ren T, Chen Y, Che W. A one-dimensional CNN-LSTM model for epileptic seizure recognition using EEG signal analysis. *Front Neurosci*. 2020;14: 578126. <https://doi.org/10.3389/fnins.2020.578126>.
39. Xiaohang Liu JJ, Zhang R. In: Proceedings of the 2020 4th International conference on computer science and artificial intelligence; 2020. p. 225–32.
40. Abbasi B, Goldenholz DM. Machine learning applications in epilepsy. *Epilepsia*. 2019;60:2037–47. <https://doi.org/10.1111/epi.16333>.
41. Rakhmatulin I, Dao MS, Nassibi A, Mandic D. Exploring convolutional neural network architectures for EEG feature extraction. *Sensors (Basel, Switzerland)*. 2024. <https://doi.org/10.3390/s24030877>.
42. Selvakumari RS, Mahalakshmi M, Prashalee P. Patient-specific seizure detection method using hybrid classifier with optimized electrodes. *J Med Syst*. 2019;43:121. <https://doi.org/10.1007/s10916-019-1234-4>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.