



OPEN ACCESS

EDITED BY

Francisco Lopez-Munoz,
Camilo José Cela University, Spain

REVIEWED BY

Ali Zia,
La Trobe University, Australia
Najmusseher S.,
Christ University, India

*CORRESPONDENCE

Xinzhou Zhang
✉ zhang.xinzhou@szhospital.com

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 23 May 2025

ACCEPTED 30 June 2025

PUBLISHED 29 July 2025

CITATION

Liang Y, Liu J and Zhang X (2025) Adaptive multi-scale phase-aware fusion network for EEG seizure recognition.
Front. Neurol. 16:1631064.
doi: 10.3389/fneur.2025.1631064

COPYRIGHT

© 2025 Liang, Liu and Zhang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Adaptive multi-scale phase-aware fusion network for EEG seizure recognition

Yanting Liang^{1†}, Jingyuan Liu^{2†} and Xinzhou Zhang^{1*}

¹Department of Nephrology, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University, The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen, China, ²Children's Heart Center, The Second Affiliated Hospital and Yuying Children's Hospital, Zhejiang Provincial Clinical Research Center for Pediatric Disease, Wenzhou Medical University, Zhejiang, China

Introduction: Epilepsy is a neurological disorder characterized by sudden, abnormal discharges of neuronal activity in the brain. Electroencephalogram (EEG) analysis is the primary technique for detecting epileptic seizures, and accurate seizure detection is essential for clinical diagnosis, therapeutic intervention, and treatment planning. However, traditional methods rely heavily on manual feature extraction, and current deep learning-based approaches still face challenges in frequency adaptability, multi-scale feature integration, and phase alignment.

Methods: To address these limitations, we propose an Adaptive Multi-Scale Phase-Aware Fusion Network (AMS-PAFN). The framework integrates three novel components: (1) a Dynamic Frequency Selection (DFS) module employing Gumbel-SoftMax for adaptive spectral filtering to enhance seizure-related frequency bands; (2) a Multi-Scale Feature Extraction (MCFE) module using hierarchical downsampling and temperature-controlled multi-head attention to capture both macro-rhythmic and micro-transient EEG patterns; and (3) a Multi-Scale Phase-Aware Fusion (MCPA) module that aligns temporal features across scales through phase-sensitive weighting.

Results: The AMS-PAFN was evaluated on the CHB-MIT dataset and achieved state-of-the-art performance, with 98.97% accuracy, 99.53% sensitivity, and 95.21% specificity (Subset 1). Compared to STFTformer, it showed a 1.58% absolute improvement in accuracy (97.39% → 98.97%) and a 2.66% increase in specificity (92.55% → 95.21%). Ablation studies validated the effectiveness of each module, with DFS improving specificity by 6.87% and MCPA enhancing cross-scale synchronization by 5.54%.

Discussion: The AMS-PAFN demonstrates strong potential for clinical seizure recognition through its adaptability to spectral variability and spatiotemporal dynamics, making it well-suited for integration into real-time epilepsy monitoring and alert systems.

KEYWORDS

EEG seizure recognition, adaptive multi-scale network, dynamic frequency selection, phase-aware fusion, deep learning, Gumbel-SoftMax

1 Introduction

Epilepsy, a prevalent neurological disorder that affects more than 50 million people worldwide, is caused by abnormal electrical discharges in the brain (1). It manifests itself as recurrent seizures, leading to transient disturbances in consciousness and motor functions, and substantially impairs patients' quality of life. Electroencephalography (EEG), which records brain electrical activity, is the gold standard for epilepsy diagnosis. However, manual

interpretation of EEG signals is time-consuming and prone to observer bias. Moreover, the inherent complexity, non-linearity, and inter-patient variability of signals make automated analysis particularly challenging (2, 3).

In the early stages of epilepsy prediction research, a wide range of machine learning (ML) algorithms were used to analyze EEG signals to detect patterns indicative of impending seizures. Various classifiers have been explored, including support vector machines (SVMs) (4, 5), which are effective in identifying optimal hyperplanes that separate different classes. Decision trees, with their intuitive if-then rule structures, provided interpretable classification models. The k-nearest neighbors (k-NN) algorithm (6) predicts outcomes based on feature similarity among data points, while Naive Bayes classifiers (7) offer fast probabilistic predictions, albeit under the simplifying assumption of feature independence. These early ML-based approaches typically depend on handcrafted features extracted from EEG recordings. Statistical descriptors, such as mean, variance, skewness, and kurtosis, are commonly used. For instance, Marzieh et al. (8) applied these features with k-nearest neighbors (k-NN) and support vector machines (SVMs) on the Bonn dataset, achieving high classification accuracy. Spectral features, such as sub-band power, were also widely used, as demonstrated by Bandarabadi et al. (9), who selected optimal sub-band power features to improve predictive performance. Time-frequency features extracted using empirical mode decomposition (EMD) and wavelet transforms were explored by Usman et al. (10) and Alickovic et al. (11), respectively, in combination with classifiers such as SVMs, k-NN, and Naive Bayes. Other studies incorporated multi-domain features, including time, frequency, complexity, and wavelet entropy, to enrich the feature space for classification (12).

When handling epileptic EEG data, each classifier has its advantages and disadvantages, and researchers select classifiers based on the task and data features. To boost the predictive performance of machine-learning methods, researchers have modified conventional machine-learning models. To address computational efficiency challenges, researchers have developed enhanced machine learning variants. Song et al. (13) reformulated standard SVM constraints into equality forms, creating a least squares SVM that reduces computational load while boosting operational speed and performance. Separately, Yuan et al. (14) advanced traditional Fisher linear discriminant analysis (FLDA) by developing Bayesian linear discriminant analysis (BLDA) through Bayesian enhancements, enabling more effective parameter optimization compared to conventional approaches.

These traditional methods demonstrated acceptable performance; however, they relied heavily on manual feature engineering, which is both labor-intensive and inherently subjective. Moreover, the success of these models was largely dependent on the quality and relevance of the selected features (15, 16). With the advancement of deep learning, automatic feature extraction from raw data has become feasible, thereby reducing dependence on handcrafted features. Inspired by the structure and functionality of the human brain, deep learning is a machine learning approach characterized by powerful data-mining capabilities. Compared to traditional machine learning techniques, deep learning algorithms offer superior performance in terms of prediction accuracy, generalizability, and scalability. As artificial intelligence and deep learning technologies continue to evolve, the scope and direction of epilepsy prediction research are also expanding rapidly (17–19).

Several neural network architectures in deep learning, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and transformers, are capable of automatically learning hierarchical feature representations.

Currently, a major focus of research lies in enhancing both the predictive accuracy and computational efficiency of these models. For example, Eberlein et al. (20) used one-dimensional CNNs on downsampled EEG to extract time-domain features. Cao et al. (21) used stacked CNNs with attention-based fusion mechanisms to learn hierarchical representations. Daoud et al. (22) integrated deep convolutional autoencoders with bidirectional long short-term memory (Bi-LSTM) networks to capture sequential dependencies. Liang et al. (23) developed a long-term recurrent convolutional network (LRCN) to localize epileptic foci from scalp EEG. Other studies, such as those by Yang et al. (24), developed dual-residual attention networks to enhance spatial and spectral feature extraction, and Jemal et al. (25) integrated spatial filters through CSP-enhanced DNNs. Li et al. (26) proposed an end-to-end capsule network to directly extract spatiotemporal features from raw EEG. Transformer architectures, originally developed for natural language processing, have recently been adapted to EEG analysis due to their capacity for modeling long-range temporal dependencies and their self-attention mechanisms (27). Bhattacharya et al. (28) pioneered the use of transformers in epilepsy prediction by combining SSTF for time-frequency feature extraction with transformer-based classification. Rukhsar et al. (29) presented a lightweight convolution transformer model that is efficient in detecting epileptic seizures across different patients using multi-channel EEG signals, demonstrating good generalization performance. Lian et al. (30) introduced a graph transformer network for EEG classification, which effectively utilized the graph structure of EEG channels to enhance the classification performance of epileptic EEG. Zhu et al. (27) proposed a method that combines the multidimensional transformer and recurrent neural network for epileptic seizure prediction, achieving improved prediction accuracy by fusing multiple features.

The aforementioned deep learning models exhibit clear advantages in processing EEG signals. They are capable of automatically extracting discriminative features through end-to-end training, thereby eliminating the need for manual feature engineering inherent in traditional approaches. Despite these advancements, existing methods still face three key challenges in epilepsy detection tasks. First, the dynamic spectral characteristics of EEG signals require models to have frequency adaptability. Yet, most methods use a fixed frequency-band division strategy. Second, pre-seizure signals have both macro-rhythm fluctuations and micro-transient spikes. But existing single-scale models cannot handle both. Third, feature fusion in multi-branch networks often ignores the phase-alignment needs of different physiological waveforms, thus failing to fully integrate temporal information. To address these limitations, this paper proposes the adaptive multi-scale phase-aware fusion network (AMS-PAFN) for epilepsy recognition. Its main innovations are as follows:

First, this novel algorithm introduces a learnable dynamic frequency selection (DFS) module. Using the Gumbel-SoftMax reparameterization technique, it can adaptively allocate weighting in the frequency domain. In the task of epilepsy prediction, this mechanism can automatically accentuate the feature bands linked to anomalous discharges while suppressing irrelevant noise.

Second, the algorithm uses downsampling at various granularities to capture macrorhythm and micro-transient features. It then leverages multi-head attention calculations to establish dynamic mapping of multi-scale features. In epilepsy prediction, the module models both the slow-wave baseline rhythm and spike transient characteristics of EEG signals. Compared with single-scale features, it can uncover more useful information.

Third, the algorithm introduces a phase-aware fusion module that addresses the timing alignment issue across different branches of a multi-scale network. This module assigns channel weights to features from multiple branches, enhancing those synchronized with key epileptic waveforms and suppressing asynchronous noise.

2 Methodology

This paper presents an adaptive multi-scale phase-aware fusion network tailored for EEG-based epilepsy recognition, with its architecture outlined in Figure 1. The network comprises three key modules: dynamic frequency (MRF) selection module, multi-scale feature extraction (MCFE) module, and multi-scale phase-aware (MCPA) module. Subsequently, we will conduct an in-depth examination and dissection of the structural composition of these three essential modules.

2.1 Dynamic frequency selection module

The dynamic frequency selection module adaptively enhances discriminative frequency components in EEG signals through a

learnable spectral filtering process. Equation 1 shows an input EEG signal $X \in R^{B \times L}$, where B is the batch size and L is the sequence length. First, the signal was transformed into the frequency domain via the fast Fourier transform (FFT) (31). This decomposes the EEG signal into its complex-valued spectral coefficients:

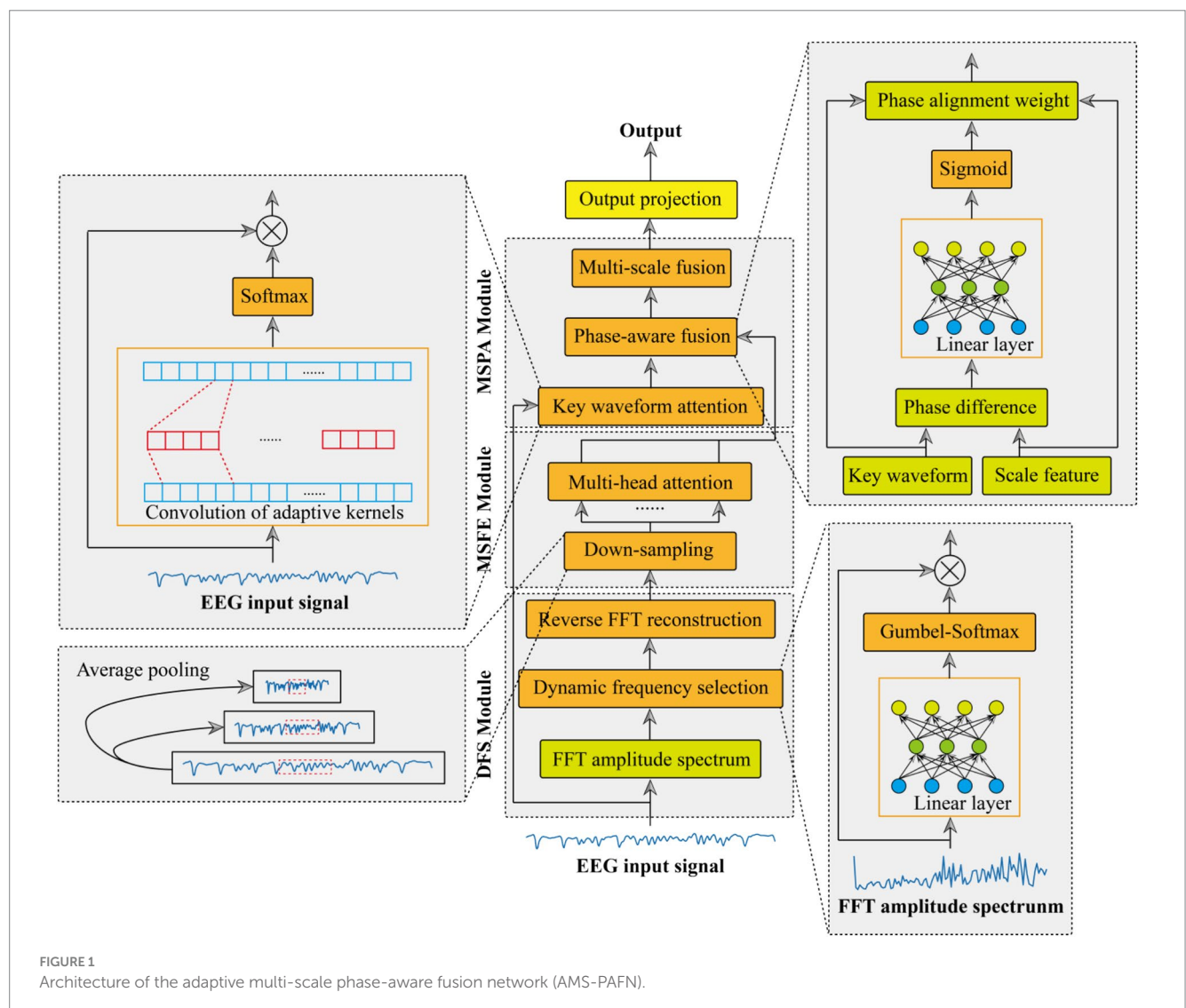
$$F = \text{FFT}(X) \in R^{B \times N} \quad (1)$$

where $N = L/2 + 1$ denotes the number of unique frequency bins. The amplitude spectrum A, calculated as the magnitude of the Fourier coefficients, is then derived as Equation 2:

$$A = |F| \in R^{B \times N} \quad (2)$$

Next, a frequency importance scoring network $S(\cdot)$, implemented as a two-layer perceptron with ReLU activation, processes A to generate logits for each frequency bin (see Equation 3).

$$S(A) = W_2 \cdot \text{ReLU}(W_1 \cdot A + b_1) + b_2 \quad (3)$$



where $W_1 \in R^{H \times N}$, $W_2 \in R^{N \times H}$ are learnable weights, $b_1 \in R^H$, $b_2 \in R^N$ are learnable biases, and H is the hidden dimension.

The logits S are converted into probabilistic frequency weights using the Gumbel-SoftMax (32) operator to ensure differentiability, as shown in Equation 4:

$$W = \text{Gumbel-Softmax}(S; \tau) \in [0, 1]^{B \times N} \quad (4)$$

The Gumbel-SoftMax operator converts the discrete frequency selection process into a continuous optimization problem by introducing differentiable relaxation variables τ . This reparameterization technique enables gradients to flow through discrete decision-making layers during backpropagation, supporting end-to-end spectrum-adaptive learning, where the weights for each frequency bin k in batch b are computed as Equation 5:

$$w_{b,k} = \frac{\exp((s_{b,k} + g_k) / \tau)}{\sum_{j=1}^N \exp((s_{b,j} + g_j) / \tau)} \quad (5)$$

In this formulation, g_k and g_j are noise variables independently sampled from $\text{Gumbel}(0, 1)$. The k and j denote frequency-point indices. $g \sim \text{Gumbel}(0, 1)$ introduces stochasticity for exploration, and the temperature τ controls the sparsity of the weight distribution.

These weights are applied to the original Fourier coefficients F through element-wise multiplication to perform spectral filtering, as shown in Equation 6:

$$\tilde{F} = F \otimes W \quad (6)$$

This operation suppresses non-critical frequencies while amplifying task-relevant components. Finally, the filtered spectrum F is transformed back to the time domain via the inverse FFT (IFFT), as shown in Equation 7:

$$\tilde{X} = \text{IFFT}(\tilde{F}) \in R^{B \times L} \quad (7)$$

The output \tilde{X} preserves the temporal resolution of the input signal but emphasizes discriminative spectral features learned through end-to-end optimization. The temperature τ adaptively balances multi-frequency enhancement ($\tau \rightarrow \infty$) and sparse frequency selection ($\tau \rightarrow 0$), while the learned weights W provide interpretable insights into frequency bands critical for the target task.

2.2 Multi-scale feature extraction

The proposed multi-scale feature fusion (MCFE) framework processes signals through a hierarchical cascade of operations designed to capture complementary patterns across scales. The specific extraction process is shown in Figure 2.

Beginning with multi-scale decomposition, the algorithm first decomposes the raw input sequence $X \in R^{B \times L}$ into three distinct temporal resolutions using adaptive average pooling. For each target

scale s_i , the pooling kernel size is computed as $k_i = L / r_i$, yielding downsampled features (see Equation 8):

$$P_i = \text{AvgPool1D}(X; k_i) \in R^{B \times s_i} \quad (8)$$

This multi-scale decomposition ensures the preservation of fine-grained details in the high-scale branch while enabling progressive abstraction in medium- and low-scale branches.

Following this spatial reduction, each downsampled sequence undergoes resolution-specific linear projection to map features into a shared latent space. The projection operation for scale s_i is formulated as Equation 9:

$$H_i = P_i W_i + b_i \quad (9)$$

where $W_i \in R^{1 \times D}$ and $b_i \in R^D$ are weights and biases respectively, D denotes the unified embedding dimension. These projections transform the variably sized temporal sequences into dimensionally consistent representations, enabling subsequent cross-scale interactions while preserving scale-specific characteristics through dedicated parameters.

Building upon these projected features, the framework uses a novel adaptive temperature-controlled multi-head attention mechanism to enhance discriminative pattern discovery. For each resolution branch H_i , the attention process begins with query-key-value (QKV) projections, as shown in Equation 10:

$$Q_i, K_i, V_i = \text{Split}(H_i W_{qkv}) \in R^{B \times s_i \times D} \quad (10)$$

where $W_{qkv} \in R^{D \times 3D}$ is shared across resolutions to promote parameter efficiency. Each projected tensor is then partitioned into H parallel attention heads along the feature dimension Equation 11:

$$Q_i^{(h)} = Q_i \left[:, :, h \cdot \frac{D}{H} : (h+1) \cdot \frac{D}{H} \right] \in R^{B \times s_i \times \frac{D}{H}}, h = 1, 2, \dots, H \quad (11)$$

with analogous splits applied to K_i^h and V_i^h . The core innovation emerges in the computation of scaled dot-product attention scores, where each head h incorporates a learnable temperature parameter τ_h to dynamically regulate attention sparsity, as shown in Equation 12.

$$\begin{cases} S_i^{(h)} = \frac{Q_i^{(h)} (K_i^{(h)})^T}{\tau_h} \\ \tau_h = \exp(\lambda_h) \end{cases} \quad (12)$$

where λ_h initialized as $\ln(\sqrt{D/H})$. This temperature scaling mechanism allows the model to autonomously balance between sharp attention distributions (emphasizing critical temporal positions when τ_h is small) and smooth distributions (capturing global context when τ_h is large). The attention weights are subsequently normalized via SoftMax (see Equation 13):

$$\alpha_i^{(h)} = \text{Softmax}(S_i^{(h)}) \in [0, 1]^{B \times s_i \times s_i} \quad (13)$$

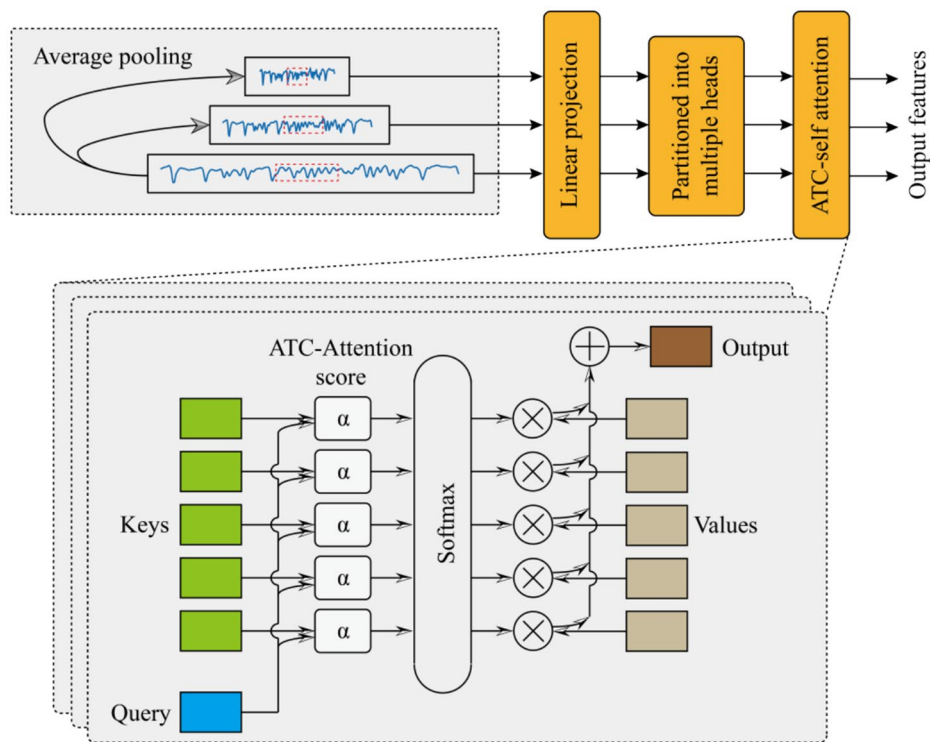


FIGURE 2
Multi-scale feature extraction (MCFE) module.

which are then used to compute context vectors through weighted aggregation of value projections (see Equation 14):

$$C_i^{(h)} = \alpha_i^{(h)} \cdot V_i^{(h)} \in \mathbb{R}^{B \times s_i \times \frac{D}{H}} \quad (14)$$

After processing all heads, the framework concatenates and linearly projects the head-specific context vectors to synthesize the final attention output for each scale (see Equation 15):

$$A_i = \text{Concat}\left(C_i^{(1)}, \dots, C_i^{(H)}\right) \cdot W_o + b_o \quad (15)$$

where $W_o \in \mathbb{R}^{D \times D}$ and $b_o \in \mathbb{R}^D$ denote the weight and bias of the linear projection.

2.3 Multi-scale phase aware

The proposed multi-scale phase aware (MCPA) module integrates three novel components: key waveform attention, phase-aware fusion, and multi-scale fusion. Achieve accurate sequence feature extraction and multi-scale synchronization in physiological signal processing.

The key waveform attention mechanism adaptively enhances discriminative waveform segments through learnable template matching. Given an input signal X , the module first synthesizes a hybrid convolution kernel that combines learnable base patterns

and preset prior knowledge. The dynamic kernel $K_t \in \mathbb{R}^k$ is formulated as Equation 16:

$$K_t = \alpha \cdot \text{Tanh}(W_b) + (1 - \alpha) \cdot K_0 \quad (16)$$

where $W_b \in \mathbb{R}^k$ denotes the trainable base template initialized from a normal distribution, $K_0 \in \mathbb{R}^k$ represents a fixed impulse kernel centered at the middle position, and $\alpha \in [0, 1]$ is a learnable coefficient balancing between adaptability and prior knowledge. The resultant kernel is energy-normalized via SoftMax to ensure stable gradient propagation (see Equation 17):

$$\tilde{K}_t = \text{Softmax}(K_t) \in \mathbb{R}^k \quad (17)$$

This normalized kernel is then convolved with the input signal to compute position-wise similarity scores. To handle boundary effects, the input is symmetrically padded before convolution (see Equation 18):

$$\begin{cases} X_{\text{padded}} = \text{ReplicatePad}(X, (p, p)), \quad p = \left\lfloor \frac{k}{2} \right\rfloor \\ S = X_{\text{padded}} * \tilde{K}_t \end{cases} \quad (18)$$

where $*$ denotes the depth-wise convolution operator. The similarity scores S are transformed into attention weights through

SoftMax normalization, which accentuates critical waveform regions while suppressing noise (see Equation 19):

$$A = \text{Softmax}(S) \in [0,1]^{B \times L} \quad (19)$$

The final enhanced signal is obtained by element-wise multiplication with the attention weights (see Equation 20):

$$\tilde{X} = X \otimes A \in R^{B \times L} \quad (20)$$

This learnable template matching mechanism adapts to morphological variations across subjects and recording conditions, while the fixed impulse component preserves the capability to detect abrupt transitions.

Based on the enhanced function of the waveform attention module, the phase-aware fusion module aligns the hidden features of multi-scale physiological signals by clearly modeling its phase relationship. The fusion process begins by computing their instantaneous phase discrepancy. A lightweight neural network estimates the signed phase difference at each timestep (see Equation 21):

$$\Delta P = W_2 \cdot \text{Tanh}(W_1 \cdot (A_i - \tilde{X}) + b_1) + b_2 \quad (21)$$

with parameters $W_1 \in R^{D \times D/2}$, $W_2 \in R^{D/2 \times 1}$, $b_1 \in R^{D/2}$, $b_2 \in R$. The hyperbolic tangent activation ensures smooth gradient flow, while the final layer projects the difference onto a scalar phase offset. These phase differences are converted into time-varying fusion weights using a sigmoid gating mechanism (see Equation 22):

$$\Delta \Gamma = \text{Sigmoid}(\Delta P) \in R^{B \times L \times 1} \quad (22)$$

The final fused features adaptively combine the input modalities, giving higher weight to the signal with better phase alignment at each timestep (see Equation 23):

$$F_i = \Gamma \otimes A_i + (1 - \Gamma) \otimes \tilde{X} \in R^{B \times L \times D} \quad (23)$$

After aligning the features of each scale, the features of each scale are fused through the concatenation operation. Finally, the probability output feature of whether epilepsy occurs is obtained through an output projection (see Equation 24):

$$y = \text{Soft max}(\text{Concat}(F_1, F_2, \dots, F_i) \cdot W_y) \quad (24)$$

3 Experimental analysis

The whole simulation experiment was carried out on a computer equipped with NVIDIA GeForce RTX 4060, and the model was constructed based on the PyTorch open-source platform.

3.1 Dataset description and evaluation metrics

The experimental data utilized in this paper are sourced from the CHB-MIT EEG dataset (33). Originating from Boston Children's Hospital, it comprises 24 segments of continuous EEG data from 23 children with refractory epilepsy, totaling 844 h. There are 182 labeled seizure records, collected via multiple bipolar channels at a 16-bit, 256 Hz sampling rate. Nevertheless, the dataset's size of 42.6 GB presents a considerable hurdle for individual researchers lacking high-performance computing resources. To circumvent similar constraints, the dataset was partitioned into smaller portions, thereby creating two subsets. Subset 1 consists of 7,280 samples, with 711 epilepsy seizure cases. Subset 2 has 8,186 samples, among which 723 are epilepsy seizure cases. Figure 3 shows the visualization of four samples, among which the first two are normal and the last two are EEG signals during epileptic seizures. In both subsets, the duration of the EEG signal for each sample is 1,228 points. Specifically, 80% of the data is used for model training, while the remaining 20% is reserved for testing the trained model.

The model's performance is assessed using four metrics: accuracy (Acc), which represents the proportion of correct predictions out of all cases; sensitivity (Sens), also known as recall, which indicates the model's ability to correctly identify positive instances; and specificity (Spec), which reflects the model's capacity to correctly recognize negative instances; F1-score, which harmonizes precision and recall into a single metric and is particularly useful for imbalanced datasets. The specific calculation methods are as follows Equation 25:

$$\begin{cases} \text{Acc} = \frac{TP + TN}{TP + FP + TN + FN} \\ \text{Sens} = \frac{TP}{TP + FN} \\ \text{Spec} = \frac{TN}{TN + FP} \\ \text{F1-Score} = \frac{2TP}{2TP + FP + FN} \end{cases} \quad (25)$$

where TP denotes true positives, where positive samples are correctly predicted as positive. TN signifies true negatives, indicating negative samples predicted as negative. FP represents false positives, where negative samples are incorrectly predicted as positive. FN stands for false negatives, meaning positive samples are wrongly predicted as negative.

3.2 Comparative experiment

To evaluate the efficacy of the introduced AMS-PAFN model, comparative experiments were performed against six state-of-the-art baseline models, namely SVM (5), SCNN (21), ResNet (24), CAE-BiLSTM (22), LRCN (23), and STFTformer (29). In the AMS-PAFN, the parameter settings of each module are as below. The input dimension of the DFS module is set to 1,228 by default, with a hidden layer dimension of 64 and a temperature parameter temp of 1.0. For the MCFE module, the preset scale list is [1,228, 614, 307], each hidden dimension in the multi-head attention mechanism defaults to 64, and the number of multi-head attention heads is 4. In

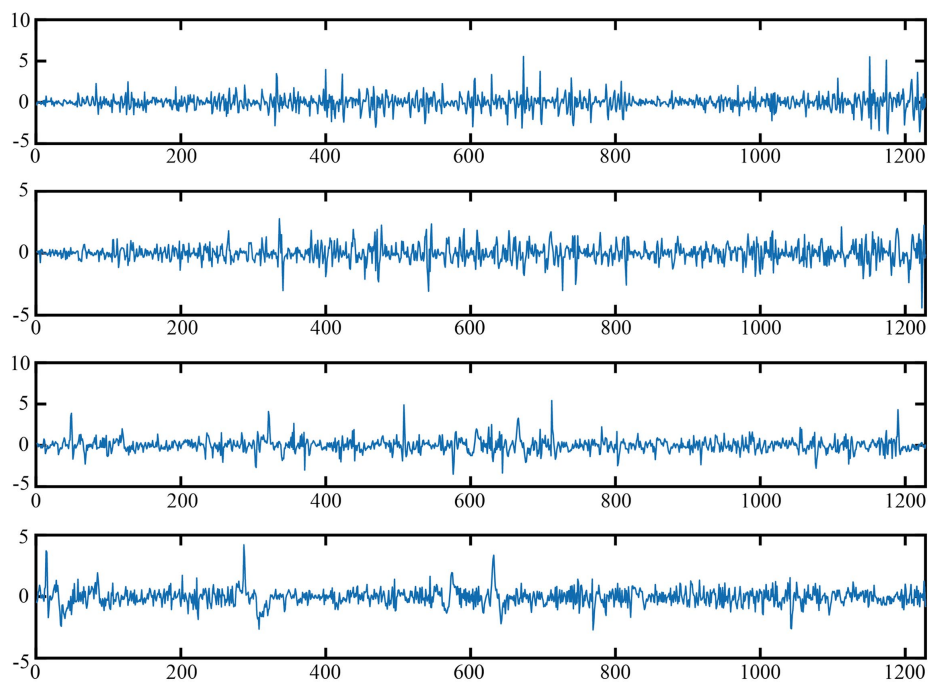


FIGURE 3
Visualization of normal and epileptic EEG signals.

the key waveform attention module, the convolutional kernel size is 15, and the initial alpha parameter is 0.5. As for the phase-aware fusion module, the feature dimension is 64 by default. The specific experimental results are shown in Table 1.

Baseline model comparisons reveal the limitations of conventional machine learning. For instance, SVMs, reliant on manual time–frequency feature extraction, achieve only 32.63% Spec on subset 1, showing static features’ high sensitivity to EMG artifacts and individual differences. SCNN (92.45% Acc) and ResNet (93.41% Acc), using single-scale convolutions, capture local temporal patterns via deep convolutions. Yet, their fixed-size kernels (16–64 sampling points) struggle to model both micro-second spikes and macro-rhythmic fluctuations in epileptic EEG, resulting in a significantly lower Spec (55.32%) than AMS-PAFN (95.21%). DCAE-BiLSTM (95.95% Acc) enhances noise robustness through autoencoder-based dimensionality reduction. However, BiLSTM’s neglect of cross-channel phase synchrony raises the false detection rate, yielding just 75.53% Spec. LRCN (95.74% Acc), combining CNNs and recurrent networks to capture spatiotemporal features, lacks dynamic frequency-domain selection. This causes generalization limitations on subset 2 due to subject-specific frequency shifts (Spec 66.35%). STFTormer (97.39% Acc) extracts time–frequency features via the short-time Fourier transform. Still, its fixed window length and wavelet basis struggle to adapt to non-stationary EEG’s transient characteristics. Moreover, its multi-head attention mechanism does not account for the phase alignment of multi-scale features, leading to insufficient modeling of cross-frequency waveform synchrony (F1-score 0.9850 vs. AMS-PAFN’s 0.9941).

In contrast, AMS-PAFN excels by integrating Gumbel-SoftMax frequency-domain adaptation in its DFS module, multi-granularity temporal modeling (at 1228/614/307 sampling rates) in its MCFE

module, and phase-aware fusion in its MCPA module. This comprehensive optimization, covering frequency-domain selection, cross-scale feature extraction, and temporal alignment, gives AMS-PAFN significant advantages in dynamic performance (Sens 99.53%), interference resistance (Spec 95.21%), and generalization (cross-dataset Acc 98.84%).

To assess the differences in model evaluation metrics, each model underwent 10 cycles of training and testing, with paired t-tests used to evaluate the significance of variations in evaluation metrics between our designed model and others. The null hypothesis for the paired t-test posits that there is no difference between the evaluation metrics of our designed model and those of others. Conversely, the alternative hypothesis suggests that a significant difference exist. Table 2 presents the test results. Given that the *p*-values are substantially less than 0.05, the null hypothesis is rejected. This indicates that the superiority of the evaluation metrics is statistically significant.

3.3 Generalization experiment

To verify the generalization ability of the AMS-PAFN model, a random division strategy of mixed datasets was adopted. All samples from the two datasets were merged and globally shuffled 20 times. Each time, the training and test sets were divided into a preset ratio of 8:2 for experiments. This mixed division approach breaks the distribution boundaries between the original datasets. It forces the model to learn general patterns from a training set with features from diverse sources. Meanwhile, it tests the model’s cross-dataset adaptation ability on a test set with mixed sources. The detailed results are shown in Figure 4.

TABLE 1 Performance comparison of AMS-PAFN and baseline models on subsets 1 and 2.

Dataset	Metrics	Models						
		SVM	SCNN	ResNet	DCAE-BiLSTM	LRCN	STFTormer	AMS-PAFN
Subset 1	Acc (%)	90.25	92.45	93.41	95.95	95.74	97.39	98.97
	Sens (%)	97.89	98.82	99.05	98.97	98.61	98.11	99.53
	Spec (%)	32.63	50.00	55.32	75.53	70.00	92.55	95.21
	F1-score	0.9463	0.9579	0.9636	0.9770	0.9760	0.9850	0.9941
Subset 2	Acc (%)	90.59	92.18	92.73	95.17	95.30	96.27	98.84
	Sens (%)	99.58	99.23	99.16	99.65	99.51	99.37	99.86
	Spec (%)	28.85	43.75	43.75	43.75	66.35	75.00	91.83
	F1-score	0.9487	0.9568	0.9597	0.9730	0.9736	0.9790	0.9934

TABLE 2 Statistically significant ($\alpha = 0.05$) on subsets 1 and 2.

Dataset	Metrics	Model	T-statistic	p-value	Statistically significant
Subset 1	Acc	STFTormer vs. AMS-PAFN	−3.8264	0.0004	Yes
		DCAE-BiLSTM vs. AMS-PAFN	−8.8724	5.3974e-07	Yes
		SCNN vs. AMS-PAFN	−22.3651	4.1178e-21	Yes
	Sens	STFTormer vs. AMS-PAFN	−3.7457	0.0026	Yes
		DCAE-BiLSTM vs. AMS-PAFN	−4.9811	5.3151e-05	Yes
		SCNN vs. AMS-PAFN	−15.4624	1.5740e-16	Yes
Subset 2	Acc	STFTormer vs. AMS-PAFN	−3.3412	0.0024	Yes
		DCAE-BiLSTM vs. AMS-PAFN	−7.3154	2.1548e-06	Yes
		SCNN vs. AMS-PAFN	−22.4884	5.8456e-18	Yes
	Sens	STFTormer vs. AMS-PAFN	−4.5487	0.0007	Yes
		DCAE-BiLSTM vs. AMS-PAFN	−12.8795	3.5648e-13	Yes
		SCNN vs. AMS-PAFN	−18.2287	5.5485e-11	Yes

As demonstrated in the generalization experiment (Figure 4), the AMS-PAFN model exhibits remarkable stability and generalization. Across 20 trials with random splits of the mixed dataset, it achieved an average test accuracy of 98.72% ± 0.31% and maintained an F1-score of 0.992 ± 0.003. These metrics indicate that the model can effectively adapt to diverse data distributions.

3.4 Ablation experiment

To evaluate the contributions of each component to the AMS-PAFN model, an ablation study was implemented. The primary objective was 2-fold: first, to reveal the role of each component, and second, to illustrate how the AMS-PAFN model leverages the synergistic benefits of these modules. By doing so, it achieves a higher level of accuracy. The specific ablation models are listed as follows:

(1) Ablation model 1 (AB1) is the original model (ORI) stripped of the DFS module. Instead, the raw EEG signal is truncated through downsampling to create multi-scale signals for feature extraction.

(2) Ablation model 2 (AB2) omits the MCFE module from the original model. After the frequency selection by the DFS module, downsampling is skipped. Instead, feature extraction is conducted directly via a multi-head attention module on a single branch. The subsequent multi-scale fusion part is also removed.

(3) Ablation model 3 (AB3) excludes the MCPA module. After the MCFE module extracts features, the subsequent multi-scale fusion part is directly carried out.

Ablation studies show that removing any module reduces performance, highlighting their importance to the model. Removing the DFS module (AB1) lowers specificity (Spec) in subset 1/2 by 6.87%/6.71%, showing it effectively suppresses noise and strengthens epilepsy-related frequency bands. Removing the MCFE module (AB2) significantly reduces Acc and F1-score (a 4.24% drop in Subset 1 Acc), indicating that multi-scale feature extraction is crucial for capturing macroscopic and microscopic rhythms. Removing the MCPA module (AB3) decreases Spec by 5.54%/5.09%, reflecting that the phase alignment module enhances cross-scale waveform synchrony. The original model (ORI) achieves the highest metrics on both datasets (Acc 99.87%/98.84%, F1-score 0.9941/0.9934), proving the synergistic effect of the three modules: DFS refines frequency-domain

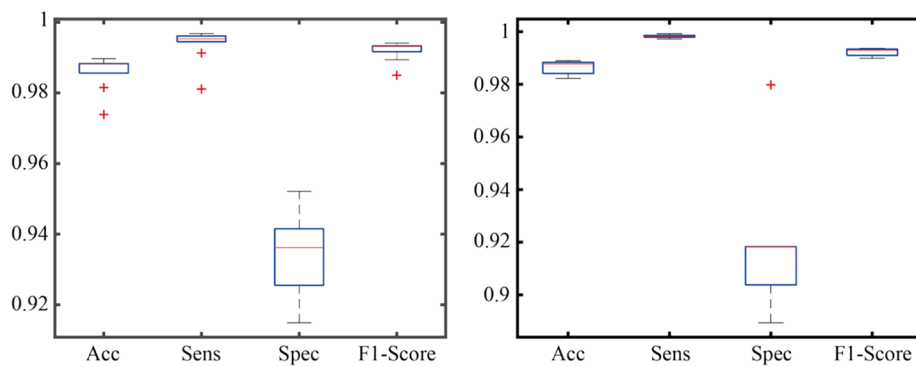


FIGURE 4
Generalization performance across mixed datasets.

representations, MCFE fuses multi-scale spatiotemporal features, and MCPA ensures cross-branch phase consistency, collectively boosting the robustness and generalization of epilepsy detection.

3.5 Sensitivity analysis of key hyperparameters

To verify the model's robustness to key hyperparameters, four core parameters were selected for sensitivity experiments: the temperature parameter τ of the DFS module, the preset scale list of MCFE, the number of attention heads H , and the initial α parameter of MCPA. This study systematically analyzed the sensitivity of four key model hyperparameters. These hyperparameters are the temperature parameter τ of the DFS module (test range [0.1, 0.5, 1, 2]), the granularity configuration of MCFE's multi-scale decomposition (three sets of [coarse[614, 307]/baseline[1,228, 614, 307]/fine[1,228, 819, 614, 409, 307]]), the number of attention heads H [(2, 4, 8)], and the template mixing coefficient α of MCPA's phase alignment ([0.3, 0.5, 0.7]). Using the control variable method, other module parameters were fixed on the Subset1 validation set while each target parameter was adjusted in turn. The aim was to evaluate the effects of different parameter combinations on accuracy, sensitivity, specificity, and F1-score. Figure 5 presents the detailed results.

To test the model's robustness to key hyperparameters, this study systematically evaluated four crucial parameters:

Dynamic frequency selection (DFS module) temperature parameter τ : The model performed best on subset 1 when $\tau = 1$ (Acc = 98.97%, F1-score = 0.9941). A very low τ (0.1) caused overly sparse frequency-band selection, dropping Spec by 6.87%. A very high τ (2) weakened frequency-band discrimination, increasing artifact interference.

Multi-scale feature extraction (MCFE module) granularity: The baseline three-scale configuration [1,228,614,307] was optimal. Compared to the coarse two-scale scheme, Spec increased by 8.98%. However, when using a fine 5-scale configuration, computational cost jumped by 42% for just a 0.85% Acc gain, proving medium-granularity features are most efficient.

Number of attention heads (H): $H = 4$ was best. It achieved 99.53% Sens by capturing diverse temporal patterns. When H increased to 8, the reasoning delay climbed by 37% with no significant performance improvement.

Phase-aware fusion (MCPA module) template mixing coefficient α : $\alpha = 0.5$ was optimal, balancing prior waveform knowledge and data-driven adaptability. Its Spec (95.21%) was much better than that of pure learning-driven ($\alpha = 0.7$, Spec = 92.34%), showing that domain knowledge guidance is important in physiological signal processing.

Overall, the model was highly robust when $\tau \in [0.5, 1.5]$ and $H \in (4, 6)$, with Acc fluctuating by less than 1.2%. This provides a reliable parameter fault-tolerance range for clinical deployment.

4 Conclusion

This study presents the adaptive multi-scale phase-aware fusion network (AMS-PAFN), a novel deep learning framework designed to address critical challenges in EEG-based seizure recognition. By integrating three innovative modules—dynamic frequency selection (DFS), multi-scale feature extraction (MCFE), and multi-scale phase-aware fusion (MCPA)—the proposed model effectively captures the spectral, temporal, and phase-aligned features of epileptic EEG signals. Experimental validation on the CHB-MIT dataset demonstrates the superiority of AMS-PAFN over existing methods, achieving 98.97% accuracy, 99.53% sensitivity, and 95.21% specificity. The DFS module enhanced frequency adaptability through Gumbel-SoftMax-based spectral filtering, while the MCFE module leverages multi-scale attention to model both macro-rhythmic fluctuations and micro-transient spikes. The DFS module's 6.87% specificity improvement over non-adaptive filtering (Table 3) addresses a critical limitation in existing frequency-static approaches. The MCPA module further improves robustness by aligning phase discrepancies across scales, as evidenced by a 5.54% gain in specificity in ablation studies. Cross-dataset generalization tests confirm the model's stability, with an average accuracy of 98.72 across randomized splits. These results indicate strong potential for AMS-PAFN's integration into clinical workflows, particularly for real-time seizure monitoring and early-warning systems. The high specificity and sensitivity rates highlight its capability to reduce false alarms while ensuring critical seizure events are not missed.

The model's 95.21% specificity (Table 1) demonstrates strong potential for reducing false alarms in ICU monitoring systems, where current clinical thresholds typically require >90% specificity to avoid alarm fatigue. The 99.53% sensitivity meets the clinical standard for seizure detection systems ($\geq 95\%$ per ILAE guidelines). Phase-aware

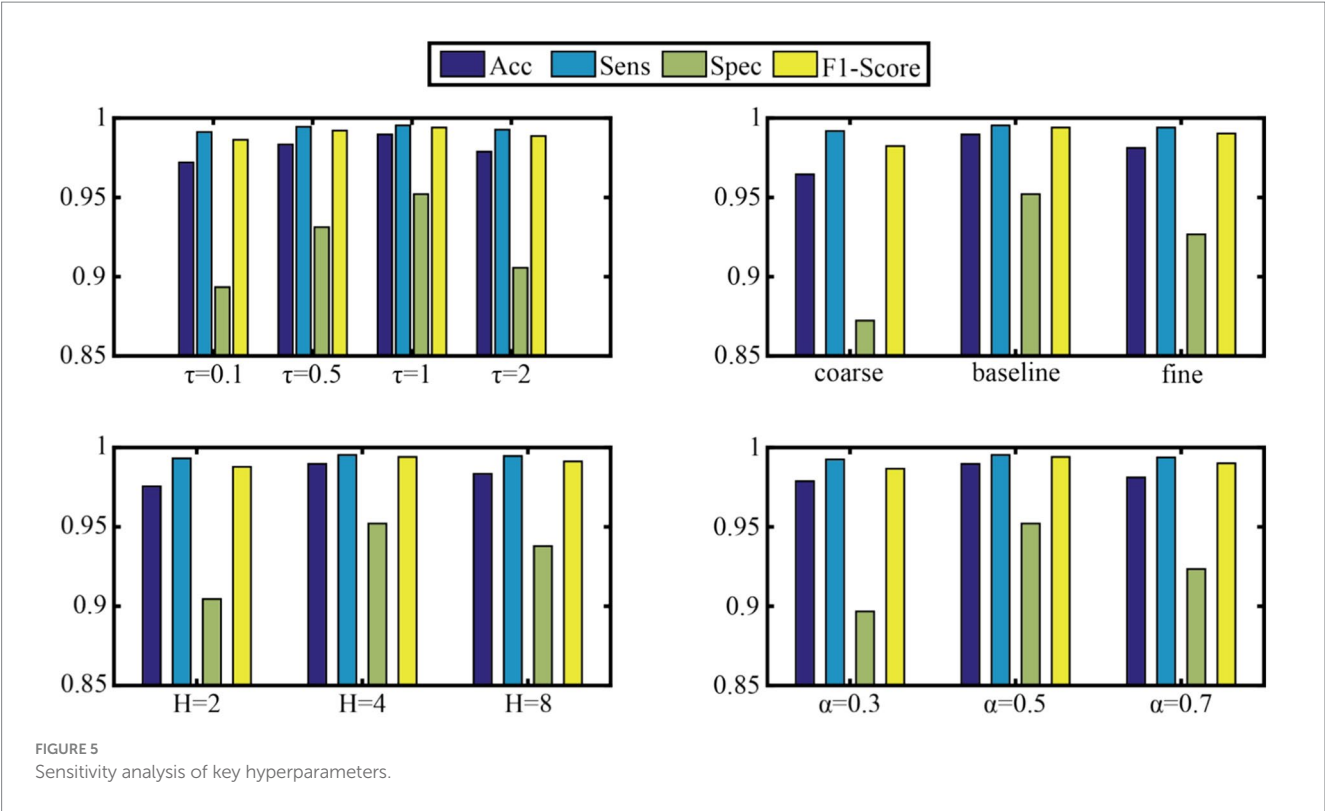


TABLE 3 Ablation study on module contributions.

Dataset	Model	Ablation module			Metrics			
		w/o DFS	w/o MCFE	w/o MCPA	Acc (%)	Sens (%)	Spec (%)	F1-score
Subset 1	AB1	✓	×	×	97.52	98.21	88.34	0.9812
	AB2	×	✓	×	95.63	97.85	72.46	0.9628
	AB3	×	×	✓	98.01	99.12	89.67	0.9863
	ORI	×	×	×	99.87	99.53	95.21	0.9941
Subset 2	AB1	✓	×	×	96.37	98.75	85.12	0.9789
	AB2	×	✓	×	94.22	97.30	68.91	0.9541
	AB3	×	×	✓	97.45	99.02	86.74	0.9835
	ORI	×	×	×	98.84	99.85	91.83	0.9934

fusion’s 5.54% specificity gain (ablation study) suggests particular utility in pediatric cases with muscle artifact interference.

While AMS-PAFN demonstrates promising results, several avenues warrant exploration. First, extend the framework to multimodal physiological signals to enhance seizure prediction reliability. Second, investigate real-time implementation to address latency constraints in clinical settings. Third, incorporate explainability mechanisms to improve transparency for medical practitioners.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the studies involving humans because the study used a pre-existing dataset. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants in accordance with the national legislation and institutional requirements.

Author contributions

YL: Writing – review & editing, Conceptualization, Writing – original draft, Methodology. JL: Writing – review & editing,

Conceptualization, Resources, Data curation. XZ: Formal analysis, Conceptualization, Funding acquisition, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the Medical Research Fund of Shenzhen Medical Academy of Research and Translation (C2301004), the Shenzhen Fund for Guangdong Provincial High-level Clinical Key Specialties (SZGSP001), and the Shenzhen Key Laboratory of Kidney Diseases (ZDSYS201504301616234).

Acknowledgments

We are grateful to the editors and reviewers for their insightful comments and suggestions. We also acknowledge the team of investigators from Children's Hospital Boston (CHB) and the Massachusetts Institute of Technology (MIT) for creating and contributing the CHB-MIT Scalp EEG Database to PhysioNet, which provided essential data support for this study. Special thanks are extended to Dr. Yongjian Zhao for his valuable assistance during the

data analysis process. We also thank Shenzhen Guoyan Era Education Technology Co., Ltd. and Guangdong Guoyan Holding (Group) Co., Ltd. for their support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Sonoda Y, Kurokawa R, Hagiwara A, Asari Y, Fukushima T, Kanzawa J, et al. Structured clinical reasoning prompt enhances LLM'S diagnostic capabilities in diagnosis please quiz cases. *Jpn J Radiol.* (2025) 43:586–92. doi: 10.1007/s11604-024-01712-2
- Steinmetz JD, Seeheer KM, Schiess N, et al. Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the global burden of disease study 2021. *Lancet Neurol.* (2024) 23:344–81. doi: 10.1016/S1474-4422(24)00038-3
- Holgate B, Fang S, Shek A, et al (2024), Extracting epilepsy patient data with llama 2. In Proceedings of the 23rd Workshop on Biomedical Natural Language Processing, 526–535.
- Shiao HT, Cherkassky V, Lee J, Veber B, Patterson EE, Brinkmann BH, et al. SVM-based system for prediction of epileptic seizures from iEEG signal. *IEEE Trans Biomed Eng.* (2016) 64:1011–22. doi: 10.1109/TBME.2016.2586475
- Tuan A. M., Mndru I., Bjenaru O. A., et al (2019), The automatic detection of epileptic seizures based on EEG signals processing: Investigation of different features and classification algorithms. World Congress on Medical Physics and Biomedical Engineering, 393–397
- Chaderyan P, Abbasi A, Sedaaghi MH. An efficient seizure prediction method using KNN-based undersampling and linear frequency measures. *J Neurosci Methods.* (2014) 232:134–42. doi: 10.1016/j.jneumeth.2014.05.019
- Yuan S, Zhou W, Chen L. Epileptic seizure prediction using diffusion distance and bayesian linear discriminate analysis on intracranial EEG. *Int J Neural Syst.* (2018) 28:1750043. doi: 10.1142/S0129065717500435
- Savedkoochi M, Oladunni T, Thompson L. A machine learning approach to epileptic seizure prediction using electroencephalo-gram(EEG) signal. *Biocyber Biomed.* (2020) 40:1328–41. doi: 10.1016/j.bbe.2020.07.004
- Bandarabadi M, Teixeira CA, Rasekhi J, Dourado A. Epileptic seizure prediction using relative spectral power features. *Clin Neurophysiol.* (2015) 126:237–48. doi: 10.1016/j.clinph.2014.05.022
- Usman SM, Khalid S, Aslam MH. Epileptic seizures prediction using deep learning techniques. *IEEE Access.* (2020) 8:39998–40007. doi: 10.1109/ACCESS.2020.2976866
- Alickovic E, Kevric J, Subasi A. Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction. *Biomed Signal Process Control.* (2018) 39:94–102. doi: 10.1016/j.bspc.2017.07.022
- Hussain L. Detecting epileptic seizure with different feature extracting strategies using robust machine learning classification techniques by applying advance parameter optimization approach. *Cogn Neurodyn.* (2018) 12:271–94. doi: 10.1007/s11571-018-9477-1
- Song Y, Niu W, Wang Y, et al. A novel method for energy consumption prediction of underwater gliders using optimal LSSVM with PSO algorithm. *Global Oceans 2020: Singapore-US Gulf Coast IEEE.* (2020):1–5. doi: 10.1109/IEEECONF38699.2020.9389244
- Yuan Q., Wei D. A seizure prediction method based on efficient features and BLDA. International Conference on Digital Signal Processing (DSP) IEEE, (2015), 177–181. doi: 10.1109/ICDSP.2015.7251854
- Kong G, Ma S, Zhao W, Wang H, Fu Q, Wang J. A novel method for optimizing epilepsy detection features through multi-domain feature fusion and selection. *Front Comput Neurosci.* (2024) 18:1416838. doi: 10.3389/fncom.2024.1416838
- Ahmad I, Yao C, Li L, Chen Y, Liu Z, Ullah I, et al. An efficient feature selection and explainable classification framework for EEG epileptic seizure detection using hybrid filter-wrapper approach. *Comput Electr Eng.* (2024) 80:105487:103654. doi: 10.1016/j.jjisa.2023.103654
- Zhang X, Zhang Y, Huang D, et al. A review of epilepsy detection and prediction methods based on EEG signal processing and deep learning. *Front Neurosci.* (2024) 18:39618710. doi: 10.3389/fnins.2024.1468967
- Tang Y, Li L, Zhang Y, et al. Epileptic seizure detection based on path signature and bi-LSTM network with attention mechanism. *IEEE Trans Neural Syst Rehabil Eng.* (2024) 32:304–13. doi: 10.1109/TNSRE.2024.3350074
- Lee D, Lee H, Kim H, et al. A ResNet-LSTM hybrid model for predicting epileptic seizures using a pretrained model with supervised contrastive learning. *Sci Rep.* (2024) 14:1319. doi: 10.1038/s41598-023-43328-y
- Eberlein M., Hildebrand R., Tetzlaff R., et al Convolutional neural networks for epileptic seizure prediction. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), (2018), 2577–2582.
- Cao J, Zhu J, Hu W, Kummert A. Epileptic signal classification with deep EEG features by stacked CNNs. *IEEE Trans Cogn Dev Syst.* (2019) 12:709–22. doi: 10.1109/TCDS.2019.2936441
- Daoud H, Bayoumi MA. Efficient epileptic seizure prediction based on deep learning. *IEEE transactions on biomedical circuits and systems.* (2019) 13:804–13. doi: 10.1109/TBCAS.2019.2929053

23. Liang W, Pei H, Cai Q, Wang Y. Scalp EEG epileptogenic zone recognition and localization based on long-term recurrent convolutional network. *Neurocomputing*. (2020) 396:569–76. doi: 10.1016/j.neucom.2018.10.108
24. 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) 29:1604–13. doi: 10.1109/TNSRE.2021.3103210
25. Jemal M, Abou-Abbas L, et al. An interpretable deep learning classifier for epileptic seizure prediction using EEG data. *IEEE Access*. (2022) 10:60141–50. doi: 10.1109/ACCESS.2022.3176367
26. Li C, Zhao Y, Song R, Liu X, Qian R, Chen X. Patient-specific seizure prediction from electroencephalogram signal via multi-channel feedback capsule network. *IEEE Trans Cogn Dev Syst*. (2023) 15:1360–70. doi: 10.1109/TCDS.2022.3212019
27. Zhu R, Pan W, Liu J, Shang JL. Epileptic seizure prediction via multidimensional transformer and recurrent neural network fusion. *J Transl Med*. (2024) 22:402. doi: 10.1186/s12967-024-05678-7
28. Bhattacharya A, Baweja T, Karri S. Epileptic seizure prediction using deep transformer mode. *Int J Neural Syst*. (2022) 32:2150058. doi: 10.1142/S0129065721500581
29. Rukhsar S, Tiwari AK. Lightweight convolution transformer for cross-patient seizure detection in multi-channel EEG signals. *Comput Methods Prog Biomed*. (2023) 242:107856. doi: 10.1016/j.cmpb.2023.107856
30. Lian J, Xu F. Epileptic EEG classification via graph transformer network. *Int J Neural Syst*. (2023) 33:2350042. doi: 10.1142/S0129065723500429
31. Murugappan M, Murugappan S. (eds.) Human emotion recognition through short time Electroencephalogram (EEG) signals using Fast Fourier Transform (FFT). 2013 IEEE 9th International Colloquium on Signal Processing and its Applications; 2013. *IEEE*. (2013) doi: 10.1109/CSPA.2013.6530058
32. Jang E., Gu S., Poole B. Categorical reparameterization with Gumbel-Softmax. *International Conference on Learning Representations*, (2017), 1–13.
33. Gutttag J. CHB-MIT scalp EEG database (version 1.0.0). *PhysioNet*. (2010). doi: 10.13026/C2K01R