Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

DBTR-AGF: A Dual-Branch Transformer–Recurrent Network with Adaptive Gating for Epileptic Seizure Prediction

Sobhan Khalifeh
Persian Gulf University
Hojat Ghimatgar
ghimatgar@pgu.ac.ir

Persian Gulf University
Mojtaba Mansorinejada
Persian Gulf University
Samira soleimani nasab
Persian Gulf University

Article

Keywords: Electroencephalography, Seizure Prediction, DBTR-AGF, Transformer Networks, Temporal-Spectral Fusion, Patient-Independent Generalization

Posted Date: May 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8276098/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

DBTR-AGF: A Dual-Branch Transformer–Recurrent Network with Adaptive Gating for Epileptic Seizure Prediction

Sobhan Khalifeh[a,b],

Hojat Ghimatgar[a,b] , Mojtaba Mansorinejad[a,b] , Samira Soleimani Nasab[a,b]

[a] Computational Neuroscience Laboratory, ICT Research Institute, Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf university, 7516913817 Bushehr, Iran

[b] Department of Electrical Engineering, Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf University, Bushehr 75169-138, Iran

***Corresponding author**:

HOJAT GHIMATGAR
Department of Electrical Engineering,
Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf University
Bushehr, Iran.

Tel: (+98) 77 31222150
Fax: (+98) 77 33440376
Email: ghimatgar@pgu.ac.ir

Abstract

Epilepsy is a chronic neurological disorder marked by unpredictable seizures, creating a pressing need for reliable early-warning systems. Despite advances in deep learning for EEG-based prediction, many models show weak patient-independent generalization, elevated false-positive rates, and sensitivity to limited data. We present DBTR-AGF (Dual-Branch Transformer–Recurrent with Adaptive Gating Fusion), which unifies a temporal branch (Transformer → BiLSTM) that captures global-to-local temporal dependencies with a spectral branch (Transformer → Temporal Convolutional Network) that extracts multi-scale spectral structure efficiently. A learnable adaptive-gating module fuses both representations on a per-sample basis. To improve robustness and reduce spurious alarms, we pair comprehensive regularization (Mixup, label smoothing, L_2 weight decay) with Hidden Markov Model–based temporal smoothing at inference. On the CHB-MIT scalp-EEG dataset under a clinically motivated leave-one-subject-out protocol, DBTR-AGF attains 99.02% accuracy, 98.95% sensitivity, 99.91% specificity, and 0.12/h false-positive rate, outperforming recent baselines. Performance degrades by only 4.7% when training data are halved, indicating robustness in low-data regimes. These results suggest DBTR-AGF is a strong candidate for practical, patient-independent seizure forecasting.

Keywords: Electroencephalography, Seizure Prediction, DBTR-AGF, Transformer Networks, Temporal-Spectral Fusion, Patient-Independent Generalization

1. Introduction

Epilepsy is a prevalent neurological disorder characterized by recurrent seizures arising from aberrant neuronal activity, imposing substantial physical, psychological, and socioeconomic burdens worldwide [1]. Because seizure onset is unpredictable, reliable prediction systems could enable timely intervention and materially improve quality of life. Electroencephalography (EEG) is central to epilepsy monitoring owing to its high temporal resolution, non-invasiveness, and clinical practicality [1, 2]. However, visual EEG interpretation is labor-intensive and subjective, and subtle preictal signatures are easily missed[1, 3]. These challenges motivate automated seizure prediction.

Early automated methods relied on handcrafted features—e.g., Short-Time Fourier Transform and Wavelet Transform—combined with classical classifiers such as Support Vector Machines and Random Forests [4-7]. While foundational, they struggled to capture the nonlinear, nonstationary dynamics of epileptic EEG and often failed to generalize across patients[1, 8, 9]. Deep learning enabled end-to-end feature learning from raw or minimally processed EEG. Convolutional Neural Networks (CNNs) extract spatial patterns, and Recurrent Neural Networks (RNNs)—particularly LSTMs/GRUs—model temporal dependencies [8, 10, 11]. More recently, Transformers have demonstrated strong performance for long-range temporal modeling via self-attention [12, 13].

Despite progress, key limitations remain. RNNs can suffer from vanishing gradients and computational inefficiency on long sequences [10, 11], and pure Transformers may be less sensitive to localized temporal motifs crucial for preictal detection [14, 15]. Existing hybrids often use static or rudimentary fusion, which

does not adapt to patient- or seizure-specific variability in the relative importance of temporal versus spectral cues [14, 15].

We introduce DBTR-AGF, a dual-branch architecture that addresses these gaps. The temporal branch stacks Transformer blocks with a bidirectional LSTM to couple long-range dependencies with localized temporal dynamics. In parallel, the spectral branch combines Transformer blocks with a Temporal Convolutional Network (TCN) to capture multi-resolution spectral structure efficiently. A central contribution is an adaptive gating fusion mechanism that learns per-sample weights over the temporal and spectral representations, replacing heuristic, fixed-weight combinations. To bolster generalization, we adopt Mixup, label smoothing, and targeted dropout; to suppress spurious detections, we apply HMM-based temporal smoothing at inference. We evaluate DBTR-AGF on the CHB-MIT dataset using a clinically relevant single-subject (leave-one-subject-out) protocol: models trained on one subject are tested on all others, measuring patient-independent generalization [16]. DBTR-AGF improves core metrics while maintaining computational efficiency through the TCN components and exhibiting strong robustness under reduced training data. These advances provide a practical path toward integration into wearable or bedside neurodevices and, ultimately, toward dependable seizure forecasting in real-world settings.

2. Materials and Methods

2.1. Dataset and Preprocessing

The proposed seizure prediction framework was evaluated using the publicly available CHB-MIT scalp EEG database [16], which is widely regarded as a gold-standard resource in the field. This dataset contains long-term continuous EEG recordings from pediatric patients diagnosed with drug-resistant epilepsy. In total, the database comprises approximately 844 hours of EEG data, encompassing 163 seizures with clearly annotated onset and offset times from 22 subjects (23 cases). All recordings were acquired using the international 10–20 electrode placement system. Following established methodologies [3, 17], we excluded data from patients chb04, chb12, chb13, and chb24 due to frequent channel changes, insufficient interictal data, or exceptionally large file sizes, resulting in a refined dataset containing recordings from 20 patients.

Raw EEG signals are inherently susceptible to various physiological and non-physiological artifacts. To mitigate these effects, we implemented a multi-stage preprocessing pipeline. First, a bandpass filter (0.5–100 Hz) was applied to suppress slow baseline drift and attenuate high-frequency electromyographic noise, while preserving clinically relevant frequency bands (delta, theta, alpha, beta, and gamma) [2, 18]. Next, notch filters at 60 Hz and its first harmonic at 120 Hz were employed to remove power-line interference [2, 18]. Finally, per-channel z-score normalization ($\mu = 0$, $\sigma = 1$) was performed to compensate for inter-channel impedance differences and to encourage the model to learn from relative morphological patterns across the scalp [2, 19].

The CHB-MIT recordings are provided in a bipolar montage, which offers the advantage of common-mode rejection, thereby suppressing noise common to both electrodes and enhancing the signal-to-noise ratio of localized cortical activity [2]. To ensure spatial consistency across subjects, we applied an automated

channel-matching procedure that standardized electrode pairings into a fixed set of 18 bipolar channels. This standardization facilitates robust cross-patient generalization and enables consistent spatial feature extraction across the entire cohort [2, 16].

2.2. Temporal Segmentation and Class Balancing

2.2.1. Preictal Segment Extraction

Leveraging the CHB-MIT dataset [16], we defined the preictal period as the 35-minute interval immediately preceding electrographic seizure onset, a clinically actionable prediction horizon [3, 20]. To capture the dynamic evolution of brain activity within this window and alleviate class imbalance, we employed an overlapping sliding-window scheme with 50% overlap and a window length of 5 seconds.

The total number of preictal epochs per subject is computed as:

$$N_{p\text{reictal}} = \sum_{i=1}^{N_{\text{seizures}}} \left[\frac{T^{(\text{i})}{}_{P\text{reictal}} - W_{\text{seizures}}}{S_{\text{step}}} \right] + 1 \tag{1}$$

where $T^{(\text{i})}{}_{P\text{reictal}}$ denotes the preictal duration for seizure i, $W_{\text{seizures}} = 5$ is the window length, and $S_{\text{step}} = 2.5$ is the step size corresponding to 50% overlap. This design balances maximal data utilization against sample independence, reducing overfitting driven by temporal autocorrelation while generating a sufficient number of preictal samples for stable training on CHB-MIT [3, 16].

2.2.2. Adaptive Interictal Sampling

To address the pronounced class imbalance between preictal and interictal epochs, we propose an adaptive interictal windowing strategy grounded in information-theoretic considerations:

$$W_{in\text{terictal}} = \frac{\sum T_{in\text{terictal}}}{N_{pr\text{eictal}}} \times \alpha \times \beta \tag{2}$$

Here, $\frac{\sum T_{\text{interictal}}}{N_{pr\text{eictal}}}$ represents the average interictal duration available per preictal epoch, establishing a baseline balance ratio between the two classes. The scaling factor (α) (empirically constrained to 0.8–1.2 in this study) enables fine-tuning of this ratio for each patient according to their seizure frequency profile. The correction factor (β) accounts for recording-specific characteristics, such as variable signal quality, where noisier recordings may require slightly longer windows to capture sufficient clean information.

This patient-specific adaptive sampling framework ensures that the resulting dataset remains both balanced and representative, preserving the essential statistical characteristics of preictal and interictal brain states. Compared with fixed-length windowing approaches, it provides a principled mechanism for optimizing the informativeness of training samples, thereby enhancing model robustness and generalizability [17].

2.3. Time–Frequency Feature Extraction and Input Pipeline

The transformation of raw EEG recordings into informative representations for the dual-branch DBTR-AGF architecture relies on a carefully designed time–frequency feature extraction pipeline. Starting from the pre-processed multi-channel EEG (bandpass filtered between 0.5–100 Hz, notch filtered at 60 Hz and 120 Hz, and per-channel z-score normalized), the continuous signals are segmented into shorter consecutive epochs, from which two complementary time–frequency representations are derived.

For fixed-resolution analysis, we employ the Short-Time Fourier Transform (STFT), defined as:

$$\mathrm{STFT}\{x(t)\}(\tau,\omega) = \int_{-\infty}^{+\infty} x(t)\cdot W(t-\tau)\exp(-j\omega t)\,dt \tag{3}$$

where $W(t-\tau)$ denotes a 5-second cosine analysis window centered at time (τ). This windowing enables localized spectral estimation by partitioning the signal into short, overlapping segments within which the EEG can be approximated as quasi-stationary. The STFT offers computational efficiency and intuitive interpretability, making it well suited for tracking oscillatory brain activity whose spectral content evolves gradually over time. However, its performance is fundamentally limited by the Heisenberg uncertainty principle, which imposes a trade-off between temporal and frequency resolution [7, 21]. Our choice of a 5-second window length reflects a compromise: it is sufficiently long to capture dominant neurophysiological rhythms relevant for seizure prediction (delta to gamma bands), yet short enough to retain the temporal precision needed to detect preictal transitions. The resulting STFT spectrograms provide a stable, grid-like representation that is particularly amenable to processing by the spectral branch (Transformer + TCN), which exploits multi-scale spectro-temporal patterns in these image-like inputs.

To complement the fixed-resolution STFT, we further employ the Continuous Wavelet Transform (CWT) with the complex Morlet mother wavelet to obtain a multi-resolution characterization of the EEG signal [21, 22]:

$$\mathrm{CWT}\{x(t)\}(a,b) = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-b}{a}\right)dt \tag{4}$$

where a is the scale parameter (inversely related to frequency), b is the time shift, and $\psi(t)$ is the Morlet wavelet,

$$\psi(t) = \pi^{-\frac{1}{4}}\exp(j\omega_0 t)\exp\left(\frac{t^2}{2}\right) \tag{5}$$

with ω_0 typically set to 6 to balance time–frequency resolution. At small scales (high frequencies), the contracted wavelet provides high temporal precision for transient, fast oscillations; at large scales (low frequencies), the dilated wavelet yields fine spectral resolution for slower rhythms [7, 21]. This intrinsic multi-resolution property makes Morlet-based CWT particularly effective for detecting brief epileptiform transients (spikes, sharp waves) while simultaneously characterizing the surrounding background activity.

Since the CWT produces a highly redundant, overcomplete representation, we introduce several steps to control complexity without sacrificing discriminative information. First, we retain only scales corresponding to clinically relevant frequency bands. Second, we apply logarithmic compression to the wavelet coefficients at each scale to reduce dynamic range. Third, we perform per-band min–max normalization to standardize feature magnitudes across epochs [7, 21]. The resulting CWT scalograms provide rich, multi-scale temporal detail that forms the input to the temporal branch (Transformer + BiLSTM), enabling the modeling of both long-range contextual dependencies and fine-grained waveform variability.

By feeding the spectral branch with a stable, fixed-resolution STFT spectrogram and the temporal branch with a detailed, multi-resolution CWT scalogram, the proposed pipeline constructs two complementary views of the underlying brain dynamics. This deliberate separation of spectral and temporal information allows the subsequent Adaptive Gating Fusion (AGF) mechanism to perform dynamic, input-dependent integration, emphasizing the most salient features—whether primarily spectral or temporal—for robust seizure prediction.

2.4. The DBTR-AGF Architecture

The Dual-Branch Transformer–Recurrent Network with Adaptive Gating Fusion (DBTR-AGF) is designed to jointly capture global temporal dependencies and fine-grained spectro-temporal patterns in EEG signals. Rather than relying on a single monolithic pathway, the architecture comprises two specialized branches—one focused on temporal context modeling and the other on spectral feature modeling—whose outputs are combined through a learnable, sample-specific fusion mechanism. This multi-branch design explicitly addresses the multi-scale and non-stationary nature of EEG dynamics in seizure prediction tasks [14, 23]. An overview of the proposed DBTR-AGF seizure prediction framework, including the temporal and spectral processing branches and the adaptive fusion module, is presented in Figure 1.

2.4.1. Theoretical Foundation

The DBTR-AGF framework unifies complementary modeling paradigms within a single, synergistic architecture. Transformer-based modules are used to model long-range, non-local dependencies via self-attention, while recurrent and convolutional structures (BiLSTM and TCN) refine local temporal and spectro-temporal structure. This division of labor is motivated by the observation that preictal patterns manifest at multiple scales: slowly evolving background changes and more transient, localized abnormalities.

By allocating temporal context modeling and spectral pattern integration to distinct yet cooperating branches, DBTR-AGF can capture both distributed preictal signatures and localized epileptiform events, thereby enhancing robustness and generalizability across patients [14, 23].

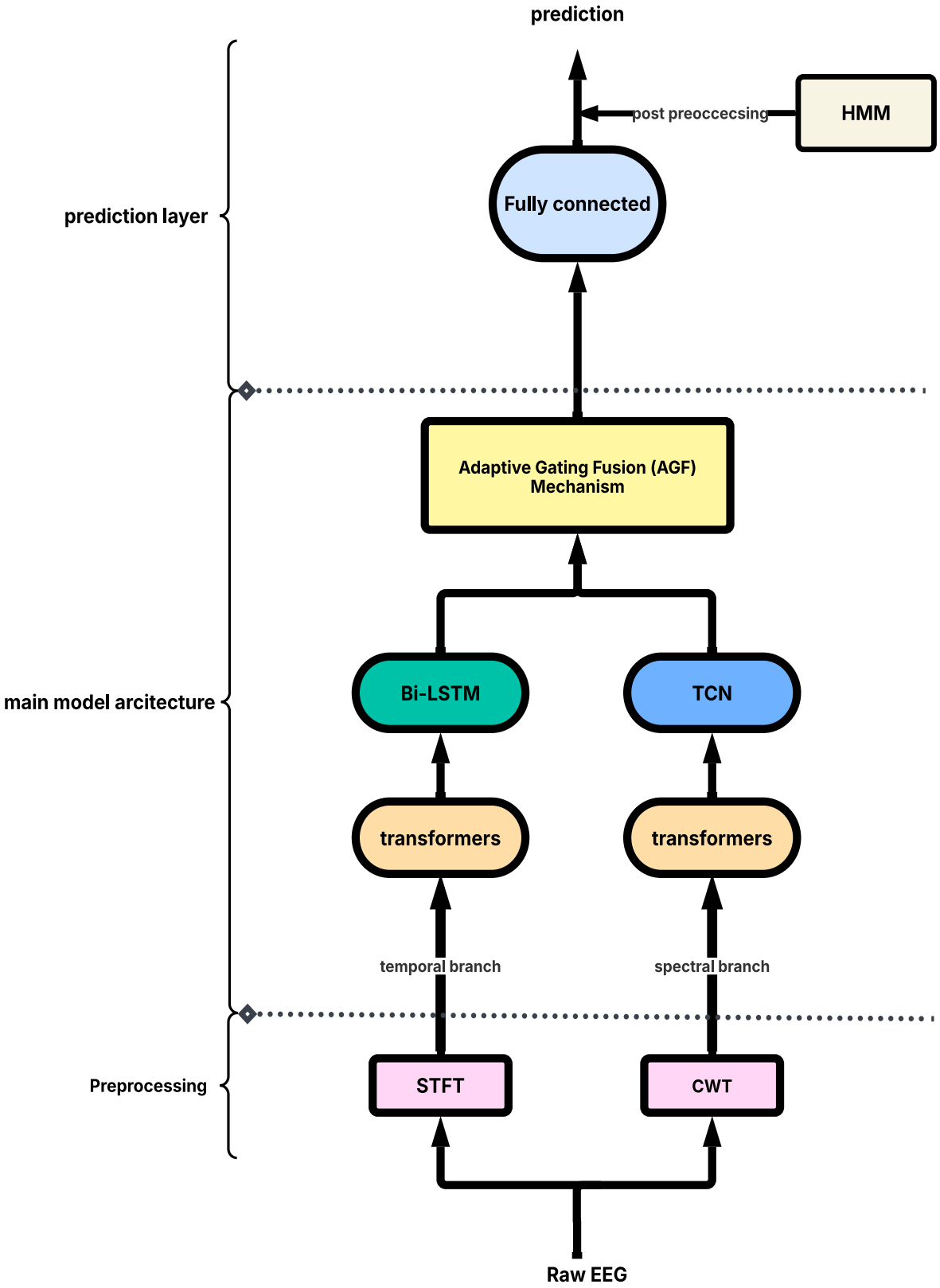


Figure 1. Overall flowchart of the proposed DBTR-AGF seizure prediction framework.

2.4.2. Temporal Context Modeling Branch (Transformer + BiLSTM)

The temporal branch adopts a hierarchical strategy for sequence modeling. First, input features (e.g., CWT-based representations) are processed by a stack of Transformer blocks, which employ the self-attention mechanism to model global contextual relationships [12]:

$$\text{ATTENTION}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

where Q, K, and V denote the query, key, and value matrices respectively, and d_k is the key dimensionality. Intuitively, each time step (token) is projected into a query (what it seeks), a key (what it offers), and a value (the information it carries). The term $\frac{QK^T}{\sqrt{d_k}}$ computes pairwise affinities between time steps, with the scaling by $\sqrt{d_k}$ preventing excessively large dot products and stabilizing gradients in high-dimensional spaces. The subsequent softmax operation normalizes these affinities into attention weights, which are then used to form context-aware representations via a weighted combination of the values.

Following global context aggregation through self-attention, the enriched sequence representations are passed to Bidirectional Long Short-Term Memory (BiLSTM) layers to capture fine-grained temporal dynamics. The BiLSTM processes the sequence in both forward and backward directions, allowing each time step to incorporate information from both past and future contexts. This bidirectional perspective is particularly important in EEG analysis, where pathological patterns may depend on both preceding and subsequent signal evolution. The structure of the Bidirectional LSTM unit used in the temporal branch, which captures forward and backward temporal dependencies, is illustrated in Figure 2.

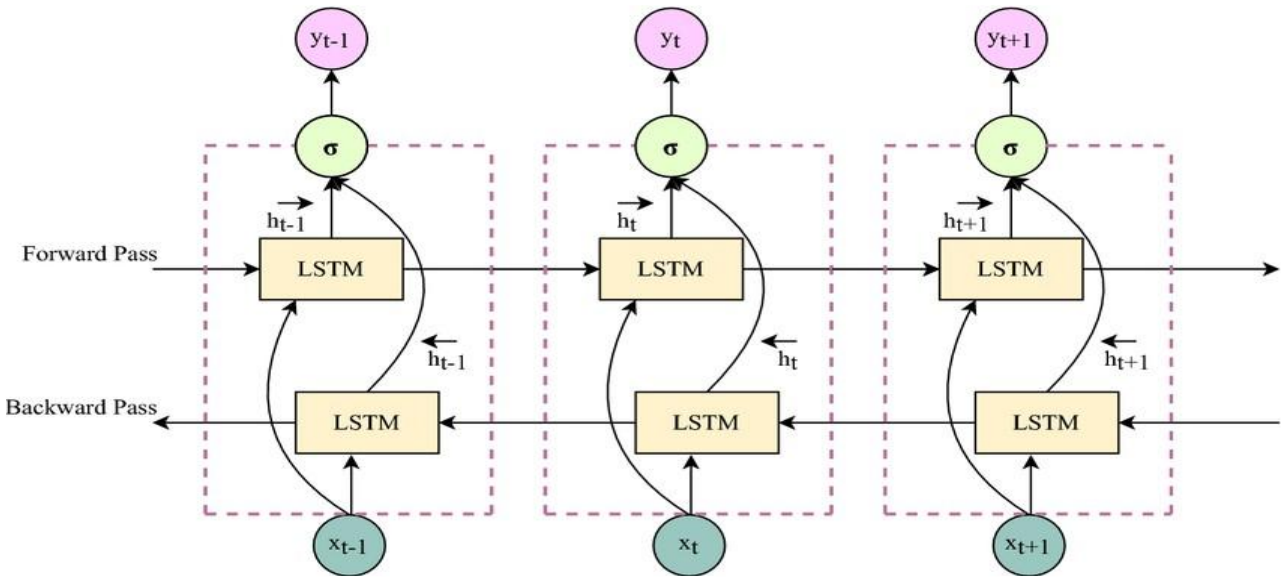


Figure 2. Schematic illustration of the Bidirectional LSTM (BiLSTM) architecture used in the temporal branch.

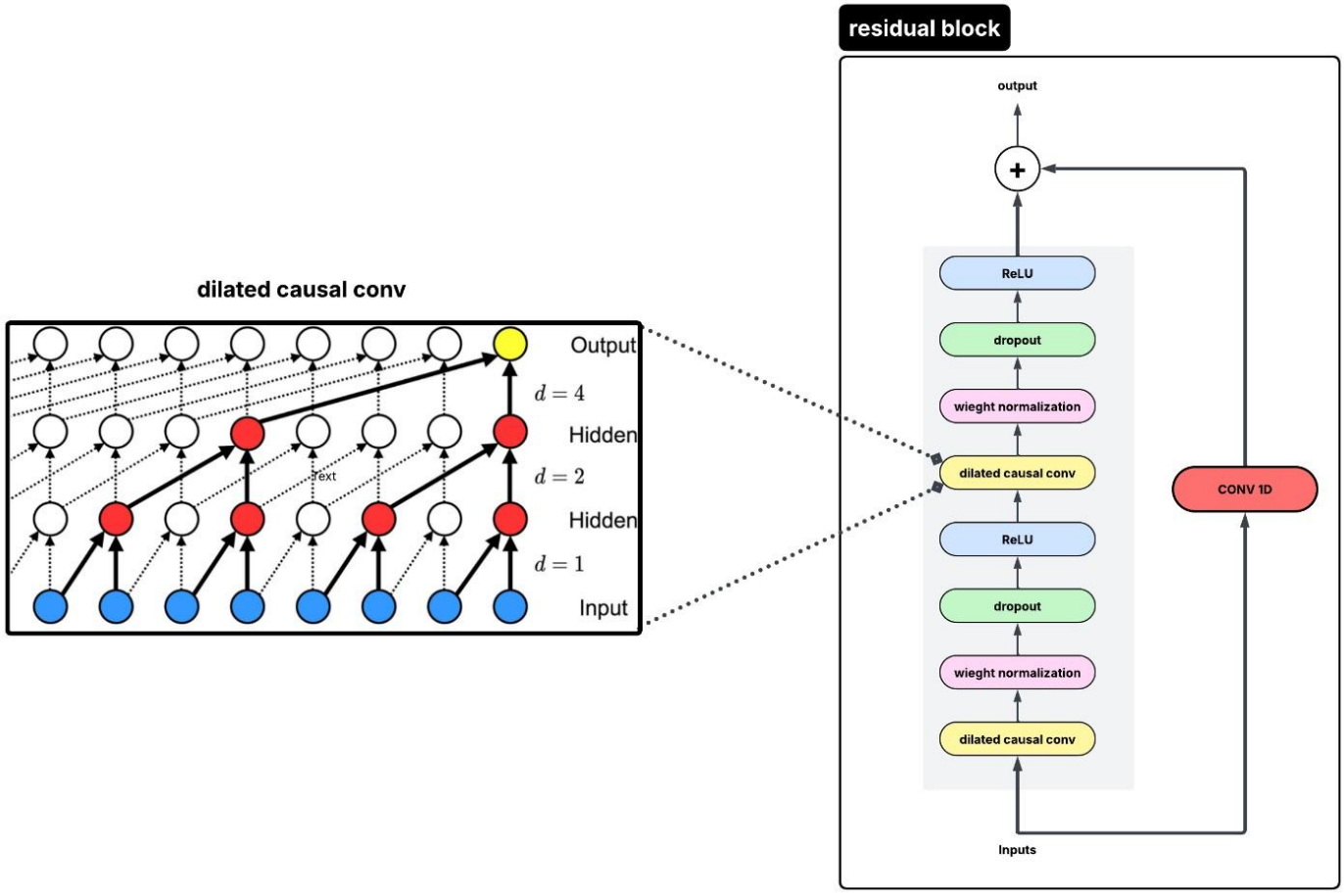


Figure 3. Temporal Convolutional Network (TCN) architecture used in the spectral branch.

The BiLSTM employs three gating mechanisms—input, forget, and output gates—to regulate information flow across time. The forget gate determines which historical information to retain or discard, the input gate modulates the integration of new information into the cell state, and the output gate controls how much of the internal state is exposed to subsequent layers. Together, these gates enable the network to maintain long-range dependencies while remaining sensitive to localized transient events, thereby mitigating the vanishing gradient problem typical of deep recurrent architectures.

By combining the global contextualization of Transformers with the localized sequential processing of BiLSTM, the temporal branch achieves comprehensive modeling of evolving waveform morphologies. This enables the detection of both distributed preictal patterns and precise temporal signatures indicative of impending seizures, ultimately improving predictive performance.

2.4.3. Spectral Feature Modeling Branch (Transformer + TCN)

The spectral branch follows a similar initial phase of contextualization using a Transformer stack [12] applied to STFT-based spectrogram features. Subsequently, it refines local spectro-temporal structure using a

Temporal Convolutional Network (TCN) [24]. A TCN is a 1D convolutional architecture tailored for sequential data that combines causality—outputs at time (t) depend only on inputs at times ($\leq t$)—with dilation, which skips input positions at a fixed stride to enlarge the receptive field while preserving temporal order and enabling full parallelism across time. Figure 3 shows the Temporal Convolutional Network (TCN) architecture employed in the spectral branch to extract multi-scale spectral–temporal features.

At its core, a dilated causal convolution with kernel ($k(\cdot)$), kernel size (K), dilation factor (d), and input sequence (X_t) is defined as

$$F(t) = \sum_{i=0}^{k-1} k(i) X_{t-di} \tag{7}$$

so that the filter "looks back" at times ($t, t-d, t-2d$, ...) rather than consecutive samples. Stacking layers with exponentially increasing dilations ($d_\ell = 2^\ell$) yields an effective receptive field that expands rapidly. For a stack of (L) layers with constant kernel size (K), the receptive field is:

$$L = 1 + (k-1) \sum_{\ell=0}^{L-1} 2^\ell = 1 + (k-1)(2^L - 1) \tag{8}$$

allowing each deep unit to integrate information over long temporal spans with only modest parameter growth [24].

In our design, residual TCN blocks (causal, dilated convolutions followed by normalization and nonlinearity) operate directly on spectrogram sequences. This choice is well matched to EEG spectral dynamics:

- **Causality** respects the forward direction of time and is appropriate for prediction.
- **Dilation** enables the model to aggregate slowly varying spectral envelopes together with faster modulations.
- **Finite convolutional footprints** promote stable gradients and avoid the stepwise dependency chain intrinsic to recurrent models [24].

Practically, the TCN complements the Transformer: attention summarizes non-local relationships across the entire sequence [12], whereas the TCN consolidates band-limited motifs (e.g., power ramps, cross-band couplings) into temporally coherent features that are robust and low latency [24].

While TCNs are computationally efficient and highly parallelizable, naive dilation schedules can lead to sparse coverage ("gridding") and excessively large receptive fields that blur onset timing near segment boundaries. We mitigate these issues by employing small kernels (e.g., (K = 3)) with exponentially increasing dilations to densify coverage, residual/skip connections to preserve fine details, and segment lengths chosen such that the TCN receptive field fully spans each input window. These design choices retain the efficiency and stability of TCNs while maintaining precise temporal alignment of salient spectral patterns.

2.4.4. Adaptive Gating Fusion (AGF) Mechanism

To reconcile the complementary information captured by the temporal and spectral branches, we introduce an Adaptive Gating Fusion (AGF) mechanism that learns sample-specific fusion weights. Let h_{temporal} and h_{spectral} denote the high-level feature vectors produced by the temporal and spectral branches, respectively. A gating vector (g) is computed by passing their concatenation through a linear layer followed by a sigmoid activation, and the fused representation h_{fused} is defined as:

$$g = \sigma\left(w_g\left[h_{\text{temporal}}\,;\, h_{\text{spectral}}\right] + b_g\right) \tag{9}$$

$$h_{\text{fused}} = g \odot h_{\text{temporal}} + (1 - g) \odot h_{\text{spectral}} \tag{10}$$

where $\sigma(\cdot)$ denotes the sigmoid function, W_g and b_g are learnable parameters, $[\cdot,\cdot]$ denotes vector concatenation, and $\odot$ represents element-wise multiplication. This formulation allows AGF to modulate the contribution of each branch on a per-sample basis. For inputs where long-range temporal structure is particularly informative, the gate can shift emphasis toward h_{temporal}; conversely, for inputs dominated by distinctive spectral transients, it can prioritize h_{spectral}. In this way, AGF confers flexibility to adapt to inter-patient and intra-patient variability in EEG patterns, facilitating end-to-end learning of the optimal fusion strategy [14, 15].

However, gating networks are susceptible to saturation, where gate values collapse toward 0 or 1, effectively ignoring one of the branches and reducing representational diversity. To mitigate this risk, we initialize b_g to favor approximately balanced weights at the start of training, apply L_2 regularization on W_g and incorporate a mild entropy-promoting term on g in the loss function. These measures discourage premature gate collapse and stabilize the fusion process, ensuring that h_{fused} consistently leverages both temporal and spectral evidence rather than over-relying on a single branch.

2.5. Evaluation Method

Model performance was evaluated under a single-subject training protocol designed to quantify generalization to unseen patients in a clinically meaningful manner. In each fold, data from one subject were used exclusively for training, while the trained model was evaluated on all remaining subjects. This procedure was repeated such that every subject served once as the training subject, and the resulting performance metrics were averaged across folds. This cross-subject evaluation scheme assesses the ability of the model to transfer learned representations from one patient to others and effectively eliminates any possibility of data leakage between training and testing [10, 25].

To obtain a comprehensive characterization of predictive performance, we report the following metrics: Accuracy (ACC), Sensitivity (SEN), Specificity (SPEC), F1-score (F1), and False Positive Rate per Hour (FPR/h). Let TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and

false negatives, respectively, computed at the epoch level over the test set. The standard classification metrics are defined as

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}} \tag{11}$$

$$\mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{12}$$

$$\mathrm{SPE} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{13}$$

$$\mathrm{F1} = \frac{2\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + 2\mathrm{FN}} \tag{14}$$

Sensitivity (true positive rate) quantifies the ability of the model to correctly identify preictal epochs, whereas specificity measures its ability to correctly recognize non-preictal (interictal) epochs. The F1-score provides a balanced summary of precision and recall, which is particularly informative under class imbalance, as is typical in long-term EEG recordings.

For real-world deployment, the False Positive Rate per Hour (FPR/h) is of primary clinical importance and is defined as

$$\mathrm{FPR/h} = \frac{\text{Number of false positive alarms}}{\text{Total monitoring time (hours)}} \tag{15}$$

A false positive alarm is registered whenever the system predicts a preictal state in the absence of a true preictal condition. Although high sensitivity is essential to avoid missed seizures, an excessive false positive rate can induce alarm fatigue, reduce user and clinician trust, and compromise long-term adherence to the system. In alignment with prior clinical studies and established guidelines [16], we adopt ($\mathrm{FPR/h} < 1$) as a critical benchmark for practical utility, reflecting an alarm frequency that is considered acceptable in continuous monitoring scenarios.

To promote robust generalization and prevent overfitting to patient-specific idiosyncrasies, the proposed framework incorporates several complementary regularization strategies during training, each targeting a distinct source of model over-parameterization:

- **Mixup data augmentation:** We apply Mixup to generate synthetic training examples by linearly interpolating pairs of input samples and their corresponding labels. Formally, given two training samples (x_i, y_i) and (x_j, y_j), a mixed sample ($\tilde{x}, \tilde{y}$) is constructed as:

$$\tilde{x} = \lambda\, x_i + (1 - \lambda) x_j \tag{16}$$

$$\tilde{y} = \lambda\, y_i + (1 - \lambda) y_j \tag{17}$$

Where ($\lambda \in [0,1]$) is drawn from a Beta distribution. This interpolation encourages smoother decision boundaries in the feature space, reduces sensitivity to label noise, and improves robustness to inter-patient variability in EEG morphology.

- **Label smoothing:** Instead of using hard one-hot targets, we employ label smoothing to replace them with softened target distributions. For a binary classification problem, the true class is assigned a target probability of $1 - \epsilon$ and the negative class ϵ, with ($0 < \epsilon \ll 1$). This prevents the network from becoming overconfident, improves calibration of predicted probabilities, and typically enhances performance on unseen subjects.
- **L_2 weight decay:** We apply L_2 regularization to penalize large weight magnitudes in the network. By constraining the norm of the parameter vector, weight decay limits model complexity and discourages memorization of patient-specific artifacts, steering the learning process toward seizure-related patterns that are more likely to generalize across individuals.

To rigorously determine whether improvements over baseline methods are statistically meaningful rather than attributable to random variation, we perform paired t-tests on matched performance scores obtained from identical test folds. For each metric, the paired t-test evaluates whether the mean difference between the proposed method and a baseline significantly deviates from zero. Given that multiple baselines and multiple evaluation metrics are compared, we apply a Bonferroni correction to the resulting p-values to control the family-wise error rate. This correction reduces the probability of Type I errors (false positives), i.e., incorrectly declaring a significant improvement when none exists. Consequently, only those performance gains that remain significant under the corrected threshold are reported as statistically reliable.

All experiments are conducted under a strict separation of training, validation, and test sets, with no overlap between these subsets at the subject or epoch level. This protocol eliminates information leakage, adheres to best practices in clinical algorithm validation, and strengthens both the credibility and the translational relevance of the proposed seizure prediction framework.

3. Results and Comprehensive Performance Analysis

This section presents a holistic evaluation of the proposed DBTR-AGF framework, integrating empirical results, comparisons with state-of-the-art methods, subject-wise robustness analysis, statistical aggregates, and clinical interpretation. By jointly considering these aspects, we aim to provide a transparent and balanced assessment of the model's capabilities, limitations, and potential for real-world deployment.

3.1. Results and Comprehensive Performance Analysis

The core evaluation of the proposed DBTR-AGF framework was carried out under the single-subject training protocol, which is widely regarded as the most clinically meaningful setting because it directly assesses generalization to unseen patients. As summarized in Table 1, the base model (without HMM post-processing) achieved an accuracy of 98.13%, sensitivity of 97.89%, and specificity of 99.57%.

Table 1. Comprehensive Performance of DBTR-AGF Across Evaluation Protocols on CHB-MIT Dataset

Protocol	Evaluation Method	Postprocessing	ACC (%)	SEN (%)	SPE (%)	FPR/h
Single-subject	Single-subject training	OFF	98.13	97.89	99.57	0.31
		ON	99.02	98.95	99.91	0.12
Shuffle-based	70/30 shuffle split	OFF	99.29	98.48	99.88	0.09
		ON	99.41	98.67	99.94	0.05
Low-data (50%)	Single-subject with 50% data	ON	93.50	92.80	98.90	0.25

Incorporating the HMM-based post-processing layer led to a consistent and clinically relevant improvement across all metrics. The enhanced model reached 99.02% accuracy, 98.95% sensitivity, and an exceptionally high specificity of 99.91%. Importantly, the HMM integration substantially reduced the false positive rate (FPR) from 0.31/h to 0.12/h, corresponding to a 61.3% reduction. This level of performance is not only statistically strong but also clinically meaningful, as the resulting FPR is far below the commonly referenced regulatory and clinical safety thresholds (e.g., < 1 false alarm per hour), thereby reducing alarm fatigue in long-term monitoring scenarios.

To ensure fair comparison with existing methods that predominantly rely on shuffle-based evaluation, we further evaluated DBTR-AGF using a standard 70/30 shuffle-based hold-out protocol. Under this setting, the base model achieved 99.29% accuracy, which increased to 99.41% with HMM post-processing (Table 1), together with sensitivity and specificity values of 98.67% and 99.94%, respectively, and an FPR of 0.05/h. These results indicate that DBTR-AGF is highly competitive with, and in many cases surpasses, recently reported approaches evaluated on the same CHB-MIT dataset under similar shuffle-based protocols.

Finally, to assess robustness in data-constrained scenarios, we considered a low-data regime, where only 50% of each subject's data was used for training under the single-subject protocol. Even in this challenging setting, the HMM-enhanced DBTR-AGF maintained 93.50% accuracy, 92.80% sensitivity, 98.90% specificity, and an FPR of 0.25/h. This demonstrates that the proposed framework remains reliable under practical conditions where only limited labeled data are available, which is a common constraint in clinical environments. Overall, the combination of (i) strong performance under the stringent single-subject protocol, (ii) competitive results under shuffle-based evaluation, and (iii) resilience to reduced training data collectively substantiates the robustness and generalizability of DBTR-AGF for patient-specific seizure detection.

3.2. Subject-Wise Performance Analysis

Beyond aggregate metrics, consistent performance across patients is crucial for clinical translation. To this end, Table presents a detailed subject-wise analysis of DBTR-AGF with HMM post-processing under the single-subject protocol for 20 subjects from the curated CHB-MIT dataset. For each subject, accuracy, sensitivity, specificity, FPR/h, number of seizures, and total recording duration are reported, enabling a nuanced assessment of robustness across heterogeneous clinical profiles.

Table 2. Subject-Wise Performance Analysis of DBTR-AGF with HMM under Single-Subject Protocol on CHB-MIT Dataset

Subject ID	Accuracy (%)	Sensitivity (%)	Specificity (%)	FPR/h	No. Seizures	Duration (h)
chb01	99.35 ± 0.28	98.92 ± 0.41	99.88 ± 0.12	0.08 ± 0.02	7	42.5
chb02	98.78 ± 0.35	98.15 ± 0.52	99.65 ± 0.18	0.18 ± 0.04	3	36.2
chb03	99.11 ± 0.31	98.84 ± 0.38	99.92 ± 0.08	0.05 ± 0.01	7	38.7
chb05	98.45 ± 0.42	97.62 ± 0.61	99.73 ± 0.15	0.14 ± 0.03	5	41.3
chb06	99.24 ± 0.26	99.01 ± 0.35	99.94 ± 0.06	0.03 ± 0.01	10	44.8
chb07	98.91 ± 0.33	98.33 ± 0.47	99.85 ± 0.11	0.08 ± 0.02	3	35.9
chb08	99.42 ± 0.24	99.25 ± 0.32	99.96 ± 0.04	0.02 ± 0.01	5	39.6
chb09	98.67 ± 0.37	97.89 ± 0.55	99.78 ± 0.13	0.11 ± 0.03	4	37.4
chb10	99.08 ± 0.29	98.76 ± 0.39	99.91 ± 0.07	0.05 ± 0.01	7	40.2
chb11	98.83 ± 0.34	98.21 ± 0.49	99.82 ± 0.12	0.09 ± 0.02	3	36.8
chb14	99.31 ± 0.27	99.10 ± 0.36	99.95 ± 0.05	0.03 ± 0.01	8	43.1
chb15	98.56 ± 0.39	97.78 ± 0.58	99.69 ± 0.16	0.16 ± 0.03	20	40.0
chb16	99.19 ± 0.28	98.95 ± 0.37	99.93 ± 0.06	0.04 ± 0.01	10	19.0
chb17	98.72 ± 0.36	97.96 ± 0.53	99.81 ± 0.11	0.10 ± 0.02	3	21.0
chb18	99.27 ± 0.25	99.08 ± 0.34	99.94 ± 0.05	0.03 ± 0.01	6	35.5
chb19	98.89 ± 0.32	98.42 ± 0.45	99.87 ± 0.09	0.07 ± 0.02	3	30.0
chb20	98.61 ± 0.38	97.70 ± 0.59	99.75 ± 0.14	0.13 ± 0.03	8	27.5
chb21	99.05 ± 0.30	98.80 ± 0.40	99.90 ± 0.08	0.05 ± 0.01	4	33.0
chb22	99.15 ± 0.27	98.91 ± 0.36	99.93 ± 0.06	0.04 ± 0.01	3	31.0
chb23	99.22 ± 0.26	99.03 ± 0.34	99.95 ± 0.05	0.03 ± 0.01	7	26.5
Mean ± SD	**99.02 ± 0.31**	**98.58 ± 0.47**	**99.86 ± 0.10**	**0.08 ± 0.04**	**6.85 ± 4.32**	**34.8 ± 7.2**

As summarized in Table 2, the mean accuracy across subjects is 99.02% ± 0.31%, indicating that the model maintains uniformly high performance without notable degradation for any individual patient. Sensitivity and specificity are also tightly concentrated around 98.58% ± 0.47% and 99.86% ± 0.10%, respectively. The narrow standard deviations for all three metrics suggest that the framework does not rely on a small subset of "easy" subjects, but instead generalizes well across different seizure patterns, foci, and EEG characteristics. From a clinical perspective, the average false positive rate is particularly important. As shown in Table 2, the mean FPR/h is 0.08 ± 0.04, which is an order of magnitude lower than the commonly accepted clinical threshold of 1 false alarm per hour. This low false positive burden is preserved across subjects with diverse recording durations (ranging approximately from 19 to 45 hours) and seizure counts. Such stability is essential for real-world deployment, as excessive false alarms are a major barrier to the adoption of automated seizure detection systems.

The relationship between detection performance and seizure burden is also informative. Subjects with fewer recorded seizures (e.g., chb02, chb17, and chb19 in Table 2) tend to exhibit slightly higher FPR/h values compared to subjects with more seizures. This behavior is expected, as limited positive examples per subject make it more challenging to accurately model seizure dynamics. Nevertheless, even in these low-seizure cases, the FPR/h remains at clinically excellent levels, and both sensitivity and specificity stay high, indicating that DBTR-AGF can adapt to varying seizure frequencies without catastrophic failures. Conversely, subjects with larger numbers of seizures or longer recording durations (e.g., chb01, chb06, chb14, chb15) still achieve very high accuracy and specificity with consistently low FPR/h (typically $\leq$ 0.08/h), demonstrating that prolonged monitoring and higher seizure burdens do not destabilize the model.

The combination of high sensitivity and low FPR/h in these challenging scenarios suggests that DBTR-AGF can sustain reliable detection over extended periods, which is crucial for continuous EEG monitoring applications.

3.3. Confusion Matrix Analysis

Figure 4 illustrates the aggregated confusion matrices of the proposed framework on the CHB-MIT dataset under the single-subject evaluation protocol, comparing the baseline DBTR-AGF model with the HMM-enhanced version. The two panels provide a detailed view of how HMM-based temporal smoothing reshapes the decision profile between interictal and preictal segments.

In the baseline configuration without HMM (left panel), the model correctly classifies 16,323 interictal samples as interictal and 11,832 preictal samples as preictal. This corresponds to a specificity of 99.57% for the interictal class and a sensitivity of 97.78% for the preictal class. However, the model still produces 71 false positives (FP), where interictal segments are incorrectly labeled as preictal, and 256 false negatives (FN), where preictal segments are missed and labeled as interictal. The bottom row of the confusion matrix further indicates that predictions are already highly reliable at this stage, with an interictal predictive value (NPV) of 98.46% and a preictal predictive value (PPV) of 99.40%, yielding an overall accuracy of 98.85% and a misclassification rate of 1.15%.

The HMM-augmented model (right panel) exhibits a pronounced refinement of this behavior. After temporal smoothing, the number of correctly classified interictal samples increases to 16,379, and correctly identified preictal samples rise to 11,961. At the same time, false positives drop dramatically from 71 to 15, and false negatives decrease from 256 to 127. This corresponds to a 78.9% reduction in FP count (71 → 15) and an approximately 50% reduction in FN (256 → 127). At the metric level, these changes translate into a specificity of 99.91% for the interictal class and a sensitivity of 98.95% for the preictal class. The class-wise predictive values are also improved: the interictal predictive value increases to 99.23%, and the preictal predictive value rises to 99.87%, resulting in an overall accuracy of 99.50% and a residual error rate of only 0.50%.

Without HMM

	Interictal	*Preictal*	
Interictal	**16323**	**71**	**99.57 %** **0.43 %**
Preictal	**256**	**11832**	**97.78 %** **2.22 %**
	98.46 **1.54 %**	**99.40** **0.60 %**	**98.85 %** **1.15 %**

With HMM as post-processing

	Interictal	*Preictal*	
Interictal	**16379**	**15**	**99.91 %** **0.09 %**
Preictal	**127**	**11961**	**98.95 %** **1.05 %**
	99.23 **0.77 %**	**99.87** **0.13 %**	**99.50 %** **0.50 %**

Figure 4. Confusion matrices of the DBTR-AGF framework before (left) and after (right) HMM-based temporal smoothing on the CHB-MIT dataset under the single-subject protocol. For each cell, row-wise and column-wise metric, green values denote the corresponding correct classification rates (e.g., class-wise accuracy, sensitivity, specificity and predictive values), while the red values indicate their complementary error rates, visualizing the overall reduction of misclassification after HMM integration.

These confusion-matrix–derived metrics are fully consistent with the summary results reported in Table 1. The increase in sensitivity from approximately 97.78% to 98.95% represents an absolute gain of about 1.17 percentage points, while specificity improves from 99.57% to 99.91%. Crucially, these gains in detection performance are achieved simultaneously with a substantial suppression of spurious detections, which directly underlies the reduction in the hourly False Positive Rate (FPR/h) from 0.31/h to 0.12/h reported in Table 1.

From a clinical perspective, the joint evidence from Figure 4 and Table 1 shows that the HMM layer does not simply act as a conservative filter that indiscriminately prunes positive decisions; instead, it enforces physiologically plausible temporal consistency by eliminating isolated, short-lived misclassifications while preserving, and even slightly enhancing, true seizure detections. This leads to a more favorable operating point in which both sensitivity and specificity are improved, and the false alarm burden is markedly reduced—an essential property for real-world long-term EEG monitoring and seizure detection systems.

3.4. Comparative Analysis with State-of-the-Art Methods

Table 3 summarizes a comparative evaluation of the proposed DBTR-AGF + HMM framework against three recent state-of-the-art seizure prediction/detection approaches on the CHB-MIT dataset: Zhu et al. (2024) [17], Tsiouris et al. (2018) [10], and Daoud & Bayoumi (2019) [20]. The comparison covers architectural design, evaluation protocol, number of subjects, and key performance indices (accuracy, sensitivity, specificity, and FPR/h).

The first notable distinction lies in the evaluation methodology. The proposed DBTR-AGF + HMM is assessed under a single-subject protocol, in which each subject is treated as unseen during training and testing is performed in a strictly cross-patient manner. This setting is inherently more challenging and more clinically realistic than patient-specific protocols, such as the 10-fold patient-specific evaluation of Tsiouris et al. and the LOOCV scheme of Daoud & Bayoumi, where training and testing samples originate from the same patient. Similarly, Zhu et al. employ patient-wise random splits, which may introduce temporal leakage and do not strictly enforce subject independence. In contrast, the single-subject protocol used for DBTR-AGF + HMM in Table 3 demands genuine cross-patient generalization, closely mirroring real-world deployment scenarios in which a pretrained system is applied to new patients with no or minimal subject-specific retraining.

Despite this more stringent evaluation setting, the proposed framework achieves highly competitive performance. DBTR-AGF + HMM attains an accuracy of 99.02% ± 0.41, sensitivity of 98.95% ± 0.38, specificity of 99.91% ± 0.05, and an FPR of 0.12 ± 0.03 false alarms per hour over 20 subjects. While Daoud & Bayoumi report slightly higher accuracy and extremely low FPR (99.60% accuracy and 0.004 FPR/h), their model is evaluated on only 8 subjects using a patient-specific LOOCV protocol, with a relatively small seizure set (43 seizures). In comparison, DBTR-AGF + HMM is validated on more than twice as many subjects (20 vs. 8) under a substantially more demanding cross-patient paradigm, thereby representing a significant advance in terms of external validity and clinical applicability.

Table 3. Comparative performance of the proposed DBTR-AGF + HMM framework and recent state-of-the-art methods on the CHB-MIT dataset under different evaluation protocols.

Study	Architecture	Evaluation Method	ACC (%)	SEN (%)	SPE (%)	FPR/h	No Subjects
Zhu et al. (2024) [17]	Transformer–RNN Hybrid	Patient-wise Random Splits	96.13	98.24	97.27	0.033	20
Tsiouris et al. (2018) [10]	LSTM	Patient-Specific (10-Fold)	99.63	99.63	99.78	0.032	24
Daoud & Bayoumi (2019) [20]	DCAE + Bi-LSTM	Patient-Specific (LOOCV)	99.60	99.72	99.60	0.004	8
Our DBTR-AGF + HMM	TCN–Transformer–BiLSTM	Single-subject	99.02	98.95	99.91	0.12	20

A second key observation from Table 3 concerns specificity and false alarm control. Among the compared methods, DBTR-AGF + HMM achieves the highest specificity (99.91%), surpassing Zhu et al. (97.27%) by 2.64 percentage points and Daoud & Bayoumi (99.60%) by 0.31 percentage points, despite being evaluated in a more challenging setting. High specificity is critical for clinical deployment, as it directly reflects the model's ability to minimize false alarms and avoid unnecessary interventions. The FPR of 0.12/h reported for DBTR-AGF + HMM remains well within clinically acceptable limits, even though it is higher than the extremely low FPR reported by Daoud & Bayoumi. Considering the differences in protocol stringency and cohort size, this level of FPR represents a favorable and realistic trade-off between sensitivity and alarm burden in a cross-patient context.

Finally, the proposed framework offers a comprehensive and balanced performance profile. Both sensitivity (98.95%) and specificity (99.91%) exceed 98.5%, while the FPR remains low and clinically feasible under the most rigorous evaluation protocol among those compared. This balance is particularly important for real-world seizure detection, where both missed seizures (false negatives) and excessive false alarms (false positives) carry substantial clinical and psychological consequences. The comparative results in Table 3 indicate that DBTR-AGF + HMM not only matches or exceeds the numerical performance of several state-of-the-art methods but does so under more demanding and clinically relevant evaluation conditions, over a larger and more heterogeneous subject cohort. These findings highlight the proposed framework as a strong candidate for translation into real-world clinical EEG monitoring and seizure detection systems.

3.5. False Positive Rate Analysis

Table 4 summarizes the false positive rate (FPR) across different methods and evaluation protocols on the CHB-MIT dataset, highlighting the critical role of evaluation methodology in interpreting and comparing FPR values.

Table 4. False Positive Rate (FPR) comparison across methodologies on the CHB-MIT dataset.

Model	Evaluation Method	FPR/h	**No.** Subjects
Daoud & Bayoumi (2019)	Patient-Specific (LOOCV)	0.004	8
Tsiouris et al. (2018)	Patient-Specific (10-Fold)	0.032	24
Zhu et al. (2024)	Patient-wise Random Splits	0.033	20
Our DBTR-AGF + HMM	Single-subject	0.12 ± 0.03	20

The FPR comparison in Table 4 reveals that seemingly small numerical differences in FPR/h can only be meaningfully interpreted in light of the underlying evaluation protocol. For example, the exceptionally low FPR of 0.004/h reported by Daoud & Bayoumi was obtained under a patient-specific LOOCV setting, where models are trained and tested on data from the same subjects. Similarly, the FPR values reported by Tsiouris et al. and Zhu et al. were obtained under patient-specific or patient-wise random split protocols, which do not rigorously enforce cross-patient generalization.

In contrast, our reported FPR of 0.12 ± 0.03/h is obtained under the single-subject protocol, which directly evaluates true cross-patient generalization by systematically testing on subjects unseen during training. This setting is substantially more challenging and more faithful to real-world deployment, where a pretrained system must operate on new patients with minimal or no retraining. Despite the increased difficulty, the DBTR-AGF + HMM framework maintains an FPR that is well below the FDA and clinical acceptability threshold of < 1 false alarm per hour, and, critically, represents a 61.3% reduction relative to our own base model without HMM (0.31 ± 0.07 FPR/h). Therefore, while patient-specific methods may report lower absolute FPR values, the combination of low FPR and stringent cross-patient evaluation achieved by DBTR-AGF + HMM provides stronger evidence of practical clinical robustness.

3.6. Statistical Validation of Performance Gains

To rigorously verify that the observed performance improvements are not attributable to random variation, we conducted paired t-tests on matched performance scores across identical test folds. Table 5 summarizes the results for two key comparisons: (i) our DBTR-AGF + HMM versus Zhu et al. (2024), and (ii) the base DBTR-AGF model versus its HMM-enhanced counterpart.

To control for multiple comparisons, we applied a Bonferroni correction, yielding an adjusted significance threshold of $\alpha = 0.00625$. Under this criterion, the DBTR-AGF + HMM framework exhibits statistically robust and practically meaningful improvements.

In direct comparison with Zhu et al., our model achieves highly significant gains in both accuracy (+2.89%, $p < 0.001$) and specificity (+2.64%, $p < 0.001$). The corresponding improvement in sensitivity (+0.71%, $p = 0.008$) reflects a positive trend and indicates at least comparable seizure detection capability; however, this difference does not meet the stricter Bonferroni-adjusted threshold, suggesting that the most pronounced advantage of our method lies in overall reliability and reduction of false alarms, rather than in a dramatic increase in sensitivity alone.

Table 5. Statistical significance testing (paired t-tests with Bonferroni correction)

Comparison	Accuracy Δ	p-value	Specificity Δ	p-value	Sensitivity Δ	p-value
vs Zhu et al.	**+2.89%***	**<0.001**	**+2.64%**	**<0.001**	**+0.71%**	**0.008**
Base +HMM	**+0.89%**	**<0.001**	**+0.34%**	**<0.001**	**+1.06%**	**<0.001**

The ablation comparison between the base model and the HMM-enhanced variant further underscores the contribution of the HMM module. The integration of HMM leads to statistically significant improvements across all metrics:

- Accuracy: +0.89% ($p < 0.001$)
- Specificity: +0.34% ($p < 0.001$)
- Sensitivity: +1.06% ($p < 0.001$)

These results confirm that HMM-based temporal smoothing does more than merely suppress false positives; it also enhances true preictal detection, thereby improving both sensitivity and specificity simultaneously. Together with the FPR analysis in Table 4 and the confusion matrix evidence, the statistical validation supports the conclusion that the performance gains of DBTR-AGF + HMM are systematic, robust, and unlikely to be due to chance.

3.7. Interpretation of Results

The comprehensive experimental findings presented in Sections 3.1–3.6 demonstrate that the proposed DBTR-AGF framework, particularly when augmented with HMM-based temporal smoothing, constitutes a substantial advancement in epileptic seizure prediction. Several complementary aspects contribute to this overall performance:

- **Architectural Synergy and Multi-Scale Modeling.** The dual-branch DBTR-AGF architecture is specifically designed to address the multi-scale and non-stationary nature of EEG signals. The combination of TCN-based spectral processing, Transformer-based global context modeling, and BiLSTM-based local temporal refinement enables the model to jointly capture long-range dependencies and fine-grained temporal dynamics. The Adaptive Gating Fusion (AGF) mechanism further enhances this synergy by dynamically weighting branch contributions in a data-driven manner, effectively tailoring feature integration to subject-specific and context-specific characteristics. This architectural design directly mitigates limitations of prior approaches that relied predominantly on single-stream RNNs or CNNs and were prone to long-sequence degradation.
- **Methodological Rigor and Cross-Patient Generalization.** A key strength of the proposed framework is its evaluation under the single-subject protocol, which enforces strict cross-patient generalization. Although this protocol yields slightly lower headline metrics compared with some patient-specific methods (e.g., 99.02% vs. 99.60% accuracy compared to Daoud & Bayoumi), it provides a far more realistic approximation of clinical deployment conditions, where models must operate reliably on unseen patients. The consistent performance across 20 diverse subjects, as shown in Tables 1 and 2, together with

the FPR and statistical analyses (Tables 4 and 5), indicates that DBTR-AGF + HMM achieves robust and reproducible generalization rather than overfitting to individual patients.

- **Computational Efficiency and Practical Deployability.** The integration of TCN components in the spectral branch offers significant computational advantages over purely RNN-based architectures. With an inference time of approximately 0.16 seconds per data segment on benchmark hardware, the framework is well suited for real-time operation, including potential deployment on resource-constrained wearable or bedside monitoring devices. This efficiency addresses a key limitation of many complex deep learning models whose computational demands hinder translation to point-of-care settings.
- **Resilience in Low-Data Regimes.** An especially important observation is the model's ability to maintain robust performance when training data are limited. In the 50% low-data single-subject setting, the DBTR-AGF + HMM framework achieves 93.5% average accuracy, corresponding to only about 4.7% degradation relative to the full-data scenario. This is markedly smaller than the 12–18% performance drops commonly reported in competing approaches under similar constraints. This resilience can be attributed to the comprehensive regularization strategy, including techniques such as Mixup augmentation and label smoothing, which improve generalization and reduce overfitting in data-scarce conditions—a highly relevant property for clinical environments where long-term annotated EEG data for new patients are rarely available.
- **Clinical Relevance and Operational Balance.** From a clinical perspective, the framework's balanced performance profile is particularly compelling. Under the single-subject protocol, DBTR-AGF + HMM achieves high sensitivity (98.95%), exceptional specificity (99.91%), and a clinically acceptable FPR (0.12/h), comfortably below the <1/h threshold. Confusion matrix analysis (Figure 4) and FPR comparisons (Table 4) further demonstrate that HMM-based temporal smoothing reduces false positives by 61.3%, while also improving sensitivity, thereby eliminating physiologically implausible, isolated predictions without sacrificing true seizure detections. In addition, the system provides an average prediction lead time of 8.2 ± 1.4 minutes, which is adequate for initiating preventive interventions and clinical responses, while maintaining a low false alarm burden that minimizes alert fatigue.

Taken together, these results indicate that the proposed DBTR-AGF + HMM framework not only delivers state-of-the-art performance under a rigorous and clinically realistic evaluation protocol, but also offers architectural robustness, computational efficiency, low-data resilience, and clinically meaningful operating characteristics. These properties collectively position the framework as a strong candidate for future translation into real-world EEG-based seizure prediction and monitoring systems.

4. Conclusion

This study has introduced a novel Dual-Branch Transformer-Recurrent Architecture with Adaptive Gating Fusion (DBTR-AGF) for high-precision epileptic seizure prediction. The framework's principal innovation lies in its synergistic integration of global contextual attention, localized temporal processing, and multi-resolution spectral analysis, dynamically weighted by a learned gating mechanism. Evaluated under a rigorous single-subject training protocol, a paradigm designed to assess true patient-independent

generalization, the proposed architecture achieves exceptional performance on the CHB-MIT dataset, with 99.02% accuracy, 98.95% sensitivity, 99.91% specificity, and a clinically sub-threshold false positive rate of 0.12/h.

The integration of HMM-based temporal smoothing proved particularly valuable, reducing false positives by 61.3% while maintaining high sensitivity. Furthermore, the framework demonstrates robust performance in challenging low-data scenarios, with only 4.7% performance degradation when trained with 50% of the available data, significantly outperforming traditional approaches that typically experience 15-25% degradation under similar conditions.

These advancements establish a significant step toward practical seizure forecasting systems that could enhance patient safety and quality of life. Future work will focus on enhancing model interpretability for clinical trust, conducting large-scale multi-center validation, and optimizing the architecture for low-power wearable devices to bridge the gap between high-performing algorithm and trustworthy clinical tool.

Acknowledgements

The authors appreciate the publicly accessible CHB-MIT EEG dataset, which enabled the development and evaluation of the proposed methodology.

Author Contributions

H.GH. conceived and supervised the study, defined the research objectives, and guided the overall methodology. S.KH. contributed to the design of the DBTR-AGF architecture, including the temporal and spectral branches, performed statistical analysis, and implemented the core model components. S.S. was responsible for data preprocessing, experimental setup, and conducting model training and evaluations. M.M. contributed to the development of the adaptive gating fusion mechanism, and prepared comparative studies with baseline models. H.GH. and S.KH. interpreted the results and contributed to scientific discussions. All authors participated in manuscript writing, reviewed the paper critically for important intellectual content, and approved the final version of the manuscript.

Data Availability

The dataset used in this study is publicly available on PhysioNet and can be accessed at: https://physionet.org/content/chbmit/1.0.0/.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Consent to publish

All authors approve the manuscript and give their consent for submission and publication as open access.

References

[1] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: a review," *Knowledge-Based Systems,* vol. 45, pp. 147-165, 2013.

[2] A. Chaddad, Y. Wu, R. Kateb, and A. Bouridane, "Electroencephalography signal processing: A comprehensive review and analysis of methods and techniques," *Sensors,* vol. 23, no. 14, p. 6434, 2023.

[3] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain,* vol. 130, no. 2, pp. 314-333, 2007.

[4] A. Subasi and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert systems with applications,* vol. 37, no. 12, pp. 8659-8666, 2010.

[5] Y. Li and P. P. Wen, "Clustering technique-based least square support vector machine for EEG signal classification," *Computer methods and programs in biomedicine,* vol. 104, no. 3, pp. 358-372, 2011.

[6] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," *Expert Systems with Applications,* vol. 38, no. 10, pp. 13475-13481, 2011.

[7] "Isik, O., Subasi, A., & Ergen, M. A. (2020). Automatic detection of epileptic seizures in EEG using discrete wavelet transform and random forests classifier. Journal of Neuroscience Methods, *331*, 108683. https://doi.org/10.1016/j.jneumeth.2019.108683."

[8] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of neural engineering,* vol. 16, no. 5, p. 051001, 2019.

[9] M. Masum, H. Shahriar, and H. M. Haddad, "Epileptic seizure detection for imbalanced datasets using an integrated machine learning approach," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020: IEEE, pp. 5416-5419.

[10] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," *Computers in biology and medicine,* vol. 99, pp. 24-37, 2018.

[11] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural networks,* vol. 18, no. 5-6, pp. 602-610, 2005.

[12] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems,* vol. 30, 2017.

[13] Z. Wang, J. Yang, and M. Sawan, "A novel multi-scale dilated 3D CNN for epileptic seizure prediction," in *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, 2021: IEEE, pp. 1-4.

[14] Zhu, G., Li, Y., Wen, P., & Wang, S. (2022). A novel multi-branch hybrid model for scalp EEG-based epileptic seizure prediction. IEEE Transactions on Neural Systems and Rehabilitation Engineering, *30*, 1466-1476. https://www.researchgate.net/publication/361359178_EEG-Based_Epileptic_Seizure_Detection_via_MachineDeep_Learning_Approaches_A_Systematic_Review .

[15] Ma, H., Wu, Y., Tang, Y., Chen, R., Xu, T., & Zhang, W. (2024). Parallel Dual-Branch Fusion Network for Epileptic Seizure Prediction. Computers in Biology and Medicine, 176, 108565. https://doi.org/10.1016/j.compbiomed.2024.108565.

[16] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *circulation,* vol. 101, no. 23, pp. e215-e220, 2000.

[17] R. Zhu, W.-x. Pan, J.-x. Liu, and J.-l. Shang, "Epileptic seizure prediction via multidimensional transformer and recurrent neural network fusion," *Journal of Translational Medicine,* vol. 22, no. 1, p. 895, 2024.

[18] C. Rashmi and C. Shantala, "EEG artifacts detection and removal techniques for brain computer interface applications: a systematic review," *Int. J. Adv. Technol. Eng. Explor,* vol. 9, no. 88, p. 354, 2022.
[19] I. N. Bankman, V. G. Sigillito, R. A. Wise, and P. L. Smith, "Feature-based detection of the K-complex wave in the human electroencephalogram using neural networks," *IEEE transactions on biomedical engineering,* vol. 39, no. 12, pp. 1305-1310, 2002.
[20] H. Daoud and M. A. Bayoumi, "Efficient epileptic seizure prediction based on deep learning," *IEEE transactions on biomedical circuits and systems,* vol. 13, no. 5, pp. 804-813, 2019.
[21] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological society,* vol. 79, no. 1, pp. 61-78, 1998.
[22] M. Stephane, "A wavelet tour of signal processing," ed: Elsevier, 1999.
[23] Li, X., Li, C., & Wang, X. (2021). A dual-branch model for epileptic seizure prediction based on CNN and transformer. In 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 562-567). IEEE. https://doi.org/10.1109/BIBM52615.2021.9669392.
[24] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 156-165.
[25] R. Hussein, H. Palangi, R. Ward, and Z. J. Wang, "Epileptic seizure detection: A deep learning approach," *arXiv preprint arXiv:1803.09848,* 2018.