Journal of Network Intelligence
Taiwan Ubiquitous Information
©2025 ISSN 2414-8105
Volume **10**, Number **1**, February **2025**

FF-DGAT: Feature Fusion Dual Graph Attention Network for Epileptic Seizure Detection Based on EEG Channel Correlation

Jian Wang

Kunming University of Science and Technology, Kunming 650504, China
jianwang@kust.edu.cn

Jia-Le Zhao

Kunming University of Science and Technology, Kunming 650504, China
zhao@stu.kust.edu.cn

Ping Zhu

Kunming Children's Hospital, Kunming 650103, China
zhuxiaopingzxp@foxmail.com

Lei Li

Kunming Children's Hospital, Kunming 650103, China
crlee801003@foxmail.com

Li Wang*

Kunming Children's Hospital, Kunming 650103, China
wangli197804@foxmail.com

*Corresponding author: Li Wang
Received February 18, 2024, revised May 8, 2024, accepted July 20, 2024.

Abstract. *Using traditional manual inspection methods to identify epileptic seizure segments in continuous electroencephalograms (EEG) is a complex and time-consuming task. Automatic epilepsy detection techniques play a pivotal role in expediting the diagnosis of epilepsy. However, current research on the extraction of information from EEG signals is insufficient, with most studies focusing solely on the temporal information from individual channels, neglecting or inadequately leveraging the spatial relationships among channels to extract more expressive spatial features. Some models adopt an end-to-end architecture that simplifies the signal processing procedure, yet they overlook the importance of the frequency domain features of EEG signals, which are pivotal for enhancing the precision in identifying epilepsy. This study introduces a novel approach for epilepsy detection utilizing the multi-scale Feature Fusion Dual Graph Attention Network (FF-DGAT), This approach innovatively addresses the challenge of underutilized spatial relationships and frequency domain features among EEG signal channels in an end-to-end network architecture by leveraging the feature fusion capabilities of Graph Attention Networks (DGAT). Our model employs two parallel GATs as front-end processing units to extract both spatial and frequency information, one for processing the raw EEG signals and the other for incorporating the extracted frequency domain information. Moreover, the model incorporates an Attentional Bi-directional Gated Recurrent Unit (Att-Bi-GRU) as a back-end network, which further investigates into the temporal relationships of EEG data, effectively enhancing the accuracy of epilepsy detection. Experiments conducted on the CHB-MIT dataset demonstrate that FF-DGAT outperforms currently known deep methods in epilepsy detection, with comprehensive improvements in accuracy, sensitivity, and specificity metrics. This research introduces a novel method for utilizing the inherent characteristics of EEG signals, thus offering a fresh perspective on the automatic detection and diagnosis of epilepsy.*

Keywords: Biomedical Signal Processing; Epilepsy Detection, Electroencephalogram, Graph Attention Network, EEG Channel Correlation, Artificial Intelligence.

---

1. **Introduction.** Epilepsy, also known as seizure disorder, ranks among the most common neurological disorders [1,2] noted for its episodic nature and lack of predictability. While seizures associated with epilepsy often last for just a brief period, their impact on the brain's normal operations can be significant, and in some cases, they may lead to fatal outcomes. Thus, it is crucial to promptly detect epileptic seizures and rapidly implement treatment measures.

Electroencephalography (EEG) is a technique capturing the brain's electrical activity through scalp-placed sensors, without being invasive [3]. To date, EEG has been crucial for identifying and managing epileptic conditions. Nevertheless, brain specialists often need to employ visual analysis methods to distinguish between normal brain electrical signals and epileptic seizure signals. This not only consumes substantial time and effort but also introduces potential subjectivity and inconsistencies in interpretation among experts. Therefore, exploring automated EEG-based epilepsy recognition offers significant real-world benefits. Beginning in the early 1970s, automatic detection of epileptic seizures has garnered the attention of numerous researchers. In the early stage, statistics-based and nonlinear machine learning techniques within the realm of time domain, frequency domain, and time-frequency domain were extensively applied for seizure detection [4]. In recent years, the development of deep learning models has gained the upper hand with their advantage of automatically extracting features from large datasets, and has been favored by those researching on automatic epilepsy detection [5,6].

Common approaches include analyzing EEG data using Convolutional Neural Networks (CNNs) [7,8] and Recurrent Neural Networks (RNNs) [9,10]. However, these methods still fall short of adequately utilizing the spatial relationships between EEG channels. Some studies have employed methods like Graph Convolutional Networks (GCN) to explore

spatial information [11-15], However, there has been little innovation in the selection of channel correlation algorithms. Therefore, there is a need for the improvement in the extraction of spatial relationships, particularly in refining algorithms for channel adjacency relations.

Furthermore, existing methods often neglect crucial frequency domain features, as many GCN implementations opt for an end-to-end approach that processes raw EEG data without distinct frequency feature extraction. This simplifies preprocessing but severely restricts the model's ability to utilize comprehensive frequency domain insights, which are essential for accurate seizure detection.

Our study introduces an innovative multi-scale Feature Fusion Dual Graph Attention Network (FF-DGAT), designed to overcome these limitations by enhancing both spatial and frequency domain analysis. FF-DGAT integrates a novel approach for channel correlation using a multi-scale feature fusion strategy. This method not only refines the extraction of spatial relationships by employing advanced algorithms for defining channel adjacency but also emphasizes the incorporation of frequency domain features into the model's architecture. The FF-DGAT method has demonstrated superior performance over existing deep learning models by effectively capturing and integrating these multidimensional features, thereby improving the accuracy and reliability of automatic epilepsy detection systems.

1.1. **Related Work.** Research into epilepsy detection using deep learning models has surged since 2016. CNNs [16,17] and RNNs, including Long Short-Term Memory networks (LSTMs) [18,19] and Gated Recurrent Units (GRUs) [20], have attracted considerable interest. For instance, Chen et al. achieved an accuracy of up to 96.67% in detecting epileptic seizures on a private dataset using a 3-layer GRU network. Some researchers have found that combining CNN with RNN can process epilepsy data more efficiently, leading to the proposition of a CNN-LSTM architecture [21,22].

However, merely analyzing EEG data from time series alone can be inadequate. Frequently, investigators neglect the inter-channel spatial relationships in EEG data, potentially leading to the omission of vital spatial correlations among electroencephalographic signals [23,24]. Some endeavors have attempted to integrate spatial relationships between EEG channels [25-30]. Acharya et al. [25] and Vidyaratne et al. [26] incorporated spatial information into EEG using Deep Convolutional Neural Networks (DCNN) and Deep Recursive Neural Networks (DRNN), respectively. Chen et al. pioneered the use of Graph Convolutional Networks (GCN) for epilepsy signal analysis to mine spatial information between channels [29]. Moreover, Covert et al. introduced the Time Graph Convolutional Network (T-GCN), an innovative approach that transforms input time series into a graph structure, efficiently combining EEG's time-based and spatial details [27]. In addition, Wei et al. and Zeng et al. employed 3D-CNN models and hierarchical GCN models, respectively, to further enhance epilepsy detection performance [28,30]. Notably, Zhao et al. were the first to employ the Graph Attention Network (GAT) [31] to allocate attention coefficients between channels.

However, these approaches primarily focus on spatial information and frequently use an end-to-end network architecture that processes raw EEG signals without extracting frequency domain features, limiting the model's capacity to capture information from the frequency spectrum. In addition, the calculations of spatial information by these approaches rely mainly on spatial distance measures, such as Euclidean Distance, and amplitude-based methods, such as Pearson correlation, both of which overlook the complexity of EEG signals. Channels not only have distance but also frequency, phase, and

other attributes. Moreover, temporal context around a specific point in the EEG signal is often not considered.

Our proposed method directly addresses these shortcomings by incorporating both spatial and temporal information through the use of a Dual Graph Attention Network (DGAT). By leveraging attention mechanisms and multi-scale features, we capture spatial relationships while incorporating frequency domain characteristics, thus ensuring a comprehensive extraction of EEG signal information.

1.2. **Motivation and contribution.** 1) The extraction of EEG spatial information by validating and adopting a Wavelet Correlation algorithm for EEG channel correlation extraction and constructing GAT to improve the utilization of EEG spatial information.

2) The methodology enhances the extraction of frequency domain details by integrating PSD, band power, and peak frequency measurements, supplemented by the incorporation of DGAT modules.

3) The proposal of FF-DGAT, a novel deep neural network framework for epileptic seizure detection, by concatenating Dual GAT and Att-Bi-GRU.

This paper initiates with an introduction of the study and a literature review, proceeds to elucidate the FF-DGAT approach in depth, presents the outcomes of experiments along with their interpretation, and wraps up by achievements, limitations and future works.

2. Materials and Methods.

2.1. **FF-DGAT Framework.** Figure 1 shows the process of the FF-DGAT framework.

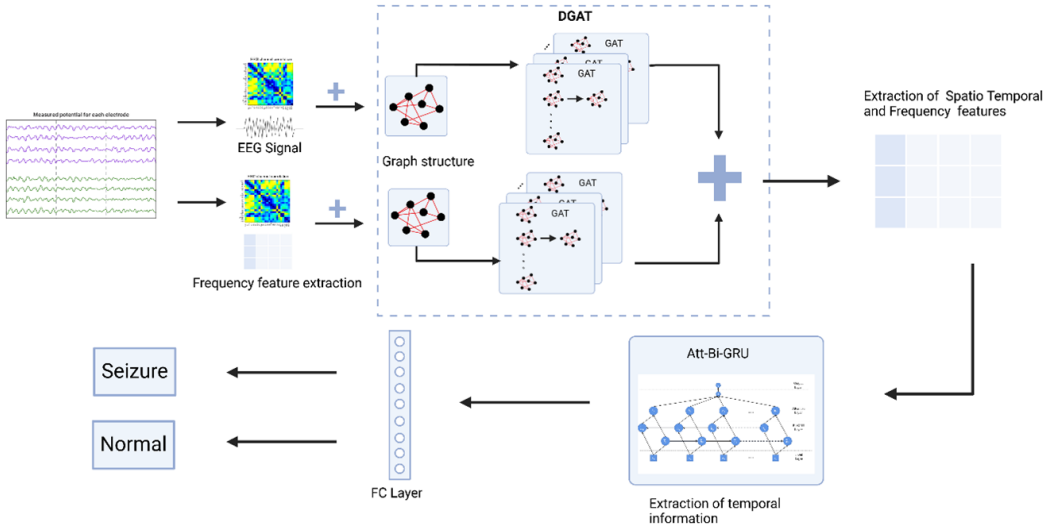
Raw EEG signals first undergo a preprocessing pipeline involving filtering, electrode re-referencing, Independent Component Analysis (ICA), and normalization, resulting in denoised input data. This preprocessed data forms the basis of a distinct graph structure, in which each EEG channel data, raw or processed, represents a graph node, while correlations between channels form the graph's edges. This graph structure enables the exploration of temporal, spatial and frequency representations of EEG signals and facilitate the effective extraction of features via the DGAT module.

The DGAT module, an enhanced version of the GAT, handles two graphs concurrently. One GAT branch processes a raw EEG-based graph, while the other focuses on a graph that incorporates frequency-based features like Power Spectral Density (PSD), Discrete Fourier Transform (DFT), and frequency band power. The dual branches of DGAT comprehensively capture spatial temporal and frequency information through an attention mechanism that effectively weighs the edges between nodes. The features extracted from the two branches are then fused through an averaging operation to form a comprehensive multi-scale feature vector.

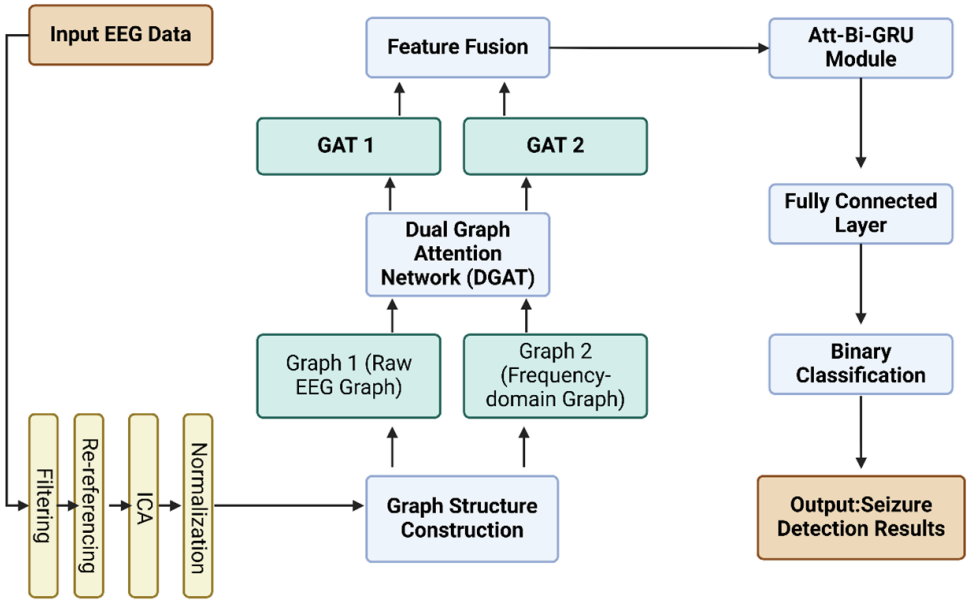
Finally, this fused feature vector is fed into an Attentional Bi-directional Gated Recurrent Unit (Att-Bi-GRU) module, which focuses on the temporal dependencies of EEG signals. By assigning appropriate weights to different time points, the Att-Bi-GRU module enhances the precision of temporal information extraction. The new features generated are then classified via a densely connected layer to detect epilepsy seizures effectively.

The function of the graph structure module is firstly to capture the spatial-temporal and frequency features of EEG data, and secondly, to lay the groundwork for the subsequent application of GAT modules, which enables the training of GAT to be possible.

Figure 2 illustrates the construction of the graph structure. As mentioned at the start of this section, each node of the graph corresponds to an EEG signal channel, and the correlation (or similarity) between channels is depicted by the graph's edges. Note that the edge formulations in both graphs are identical, that is, the inter-channel similarity is initially determined by employing a Wavelet Correlation algorithm, and subsequently,



(a) FF-DGAT framework



(b) FF-DGAT algorithm flow chart

Figure 1. FF-DGAT epilepsy detection framework

connections are formed between nodes when inter-channel similarity measures exceed a certain threshold. However, the node construction methods in the two graphs are essentially different. In the upper branch of Figure 2, raw EEG channel data is directly stored as nodes in the graph structure without any modifications, while in the lower branch of the graph structure, nodes are formed through frequency-related feature extraction on raw EEG channel data, including features extracted through Discrete Fourier Transform (DFT), Power Spectral Density (PSD) features on five key brainwave bands(delta, theta, alpha, beta, gamma), frequency band power features, and peak frequency features.

2.2. **Data Preprocessing and Feature Extraction.** In this research, EEG data collection was performed through channels following the international 10-20 electrode placement system. These multi-channel recordings are noise-contaminated due to external

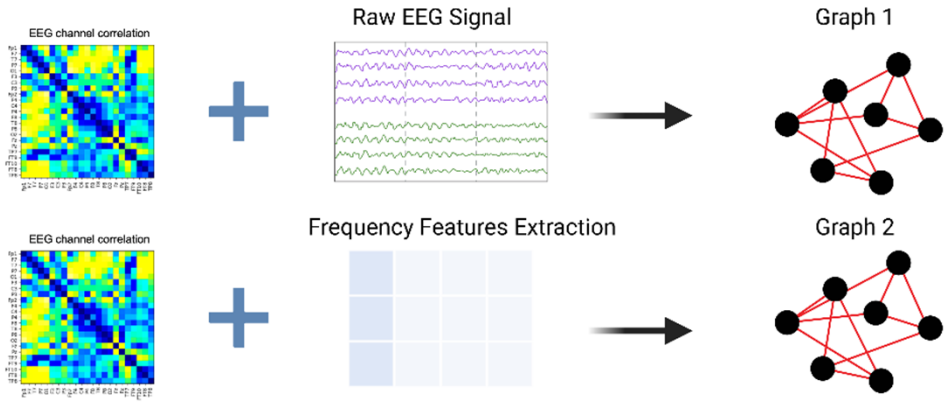


Figure 2. Two types of graphs

interferences like eye movements, heartbeats, and blinks. To improve signal clarity and interpretation, we undertook specific pre-processing steps.

1. The EEG signals are passed through a band-pass filter, constraining the frequency range from 1Hz to 40Hz. This process effectively filters out or minimizes interference and irrelevant elements.

2. A re-referencing technique is employed to adjust the reference electrode of the EEG signals. The primary objective of this technique is to diminish certain offsets or common reference effects present in the signals, thereby augmenting their interpretative power.

3. Independent Component Analysis (ICA) is leveraged to remove noise resulting from eye movements and limb activities.

4. Z-score normalization is utilized to standardize the signal magnitude. The formula is as follows:

$$X_{\text{normalization}} = \frac{x - \mu}{\delta} \tag{1}$$

Once the EEG data is preprocessed, we start to identify critical frequency domain characteristics contained in EEG data, such as Power Spectral Density (PSD), band power, and peak frequency. We carefully choose these attributes because they disclose vital details of the frequency domain pertinent to epilepsy, highlighting shifts in energy distribution across distinct frequencies and modifications in the main frequency elements.

1) Power Spectral Density (PSD) PSD illustrates how a signal's power is distributed among different frequency bands, showcasing how energy is allocated within the frequency spectrum. Through PSD analysis, we can detect unusual patterns of energy distribution in EEG data during seizures, which supports the diagnosis and tracking of epilepsy. Calculating PSD for each channel requires first applying the Discrete Fourier Transform (DFT) to raw EEG data. The DFT's equation is presented as follows:

$$F(f) = \sum_{n=0}^{N-1} f(n) \cdot e^{-j\frac{2\pi}{N}kn} \tag{2}$$

where $F(f)$ is the frequency domain signal after DFT, $f(n)$ represents the time-domain discrete signal. N signifies the overall count of signal samples, and k is the frequency index.

Subsequently, the PSD at each frequency is calculated, with the following formula for PSD:

$$P(f) = \frac{1}{N}|F(f)|^2 \tag{3}$$

where $F(f)$ is the Fourier transform of the EEG signal, N is the length of the signal, and $|F(f)|^2$ is the squared magnitude of the Fourier transform result.

2) Band Power Dividing the EEG signal into specific frequency bands, such as delta, theta, alpha, and beta waves, enables the analysis of various brain activity patterns. Since epileptic seizures often accompany significant changes in power within certain frequency bands, analyzing the power of specific bands can help identify EEG patterns related to epilepsy, thereby enhancing the accuracy and specificity of epilepsy detection.

$$B = \sum_{f_1}^{f_2} P(f) \tag{4}$$

where f_1 and f_2 represent the lower and upper frequency limits of the band considered, respectively.

3) Peak Frequency Peak frequency refers to the frequency where the power spectrum attains its maximum power, signifying the most dominant frequency component of the EEG signal. During epileptic seizures, the peak frequency may change. By tracking these fluctuations, the initiation and end of epileptic seizures can be more accurately pinpointed, particularly for seizures that manifest subtly or at their inception, which may elude detection by conventional time-domain analytical approaches.

$$F_{\text{peak}} = argmax_f P(f) \tag{5}$$

where $argmax_f$ represents finding the frequency that maximizes $P(f)$ across all frequencies.

In summary, focusing on the frequency domain features of EEG signals, particularly power spectral density, band power, and peak frequency, is crucial for enhancing the accuracy and sensitivity of epilepsy detection. Therefore, these three features are carefully chosen and then extracted and fused into the graph's nodes.

2.3. **EEG Channel Correlation Analysis Methods.** During the EEG signal acquisition process, different channels will produce distinct voltage fluctuations at the same time point. The correlation between channels corresponding to these different channels is pivotal to interpreting these voltage fluctuations.

Through examining works in related fields, we have discovered that researchers typically utilize Euclidean Distance or Pearson Correlation Coefficient to compute channel correlation. Yet, there are many beneficial techniques still to be utilized, including Phase Locking Value (PLV) and wavelet coherence. PLV is employed to assess the phase stability and synchrony in the interactions between two signals, while wavelet coherence provides a unique approach by evaluating the inter-relation of two signals across various frequencies, pinpointing areas with high shared power and consistent phase over time. Thanks to wavelet analysis's ability to capture signals at different frequency scales, it can intricately unveil interactions between channels across multiple frequencies, rendering it advantageous for EEG signal analysis.

This study selected the aforementioned four algorithms, namely Euclidean Distance, Pearson Correlation Coefficient, PLV, and Wavelet Correlation, for channel correlation computation. Through comparative experiments, we aim to identify an algorithm that offers unique advantages in channel correlation analysis and apply it as an improved algorithm in the experiments presented in this paper. Next, we look at four correlation algorithms in detail:

1) Euclidean Distance Euclidean Distance measures the geometric difference or absolute distance between two channels. It doesn't involve signal correlation, frequency characteristics, or phase information and purely judges based on spatial distance. For two n-dimensional vectors x and y, one can calculate the Euclidean Distance by the formula:

$$D = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{6}$$

2) Pearson Correlation Coefficient Contrasting with Euclidean Distance, the Pearson Correlation Coefficient stands as a foundational approach for determining the linear correlation between two signals. This approach determines if the temporal evolution patterns of two signals are alike, offering a direct insight into their linear correlation. The calculation formula is as follows:

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{7}$$

3) PLV - Phase Locking Value

PLV emphasizes evaluating the phase consistency between two signals. Specifically, it investigates whether the phase changes of the two signals are synchronized, providing vital information for studying functional connections in EEG. The mathematical formula is as follows:

$$PLV = \left| \frac{1}{N} \sum_{n=1}^{N} e^{j(\phi_{1,n} - \phi_{2,n})} \right| \tag{8}$$

Where N is the number of sample points within the time segment, $e^{j(\phi_{1,n}-\phi_{2,n})}$ represents the complex representation of the phase difference between the two EEG signals at the nth time point. $\phi_{1,n}$ and $\phi_{2,n}$ respectively denote the phases of the first and second EEG signal at the nth time point. j is the imaginary unit.

4) Wavelet Correlation

Wavelet Coherence is utilized as a technique to assess the consistency among signals by combining information from both frequency and temporal domains to analyze signal interactions. Compared to other methods, it not only considers the linear association of the signals but also delves into specific frequency components, offering a multi-dimensional perspective.

$$W_x(t, f) = \int_{-\infty}^{+\infty} x(u) \cdot \theta^*_{t,f}(u) du \tag{9}$$

$$CW_{xy}(t, f) = \int_{t-\frac{\theta}{2}}^{t+\frac{\theta}{2}} W_x(\tau, f) \cdot W^*_y(\tau, f) \tag{10}$$

Where $W_x(t, f)$ represents the wavelet transform value of signal $x(u)$ at time t and frequency f. $x(u)$ denotes the original signal, where u is the time variable, and $\theta_{t,f}(u)$ is the wavelet function, dependent on time t and frequency f.

$CW_{xy}(t, f)$ represents the wavelet coherence between two signals x and y at time t and frequency f. $W_x(\tau, f)$ and $W_y(\tau, f)$ respectively represent the wavelet transform values of signals x and y at time τ and frequency f. θ is the width of the window.

2.4. **Feature Extraction Network.**

1) Dual Graph Attention Network [19, 20]

We have opted for two parallel GAT models to explore the spatial relationships among EEG signal channels. Both models share the same structural configuration, with the sole difference residing in the graph data they receive. One module inputs a graph with raw EEG signals as node features, while the other module inputs a graph with node features derived from extracted data. Through the incorporation of information from both the time and frequency domains, the dual GAT network can examine EEG data through two distinct perspectives. This multi-faceted approach enhances the model's ability to understand brain electrical activity, thereby improving the efficiency and accuracy of epilepsy detection.

The graph attention coefficient, α_{ij}, serves as the edge weight in the GAT network, quantifying the correlation between channels. Given the often lengthy nature of EEG time series, we employ a sliding window approach to segment the original data, ensuring each window captures local features of the time series effectively. These segmented data chunks are then fed into the GAT module, allowing for a more precise analysis of the dynamic relationships between channels. The GAT module's detailed explanation is as follows:

Assuming there are n EEG images, denoted as $S = [S_1, S_2, ..., S_n]$, each subject can be represented by $S_i \in R^{N\times F}$. Where N is the number of EEG channels for each subject, and F denotes the vector dimension in each channel.

The input to the GAT model is a set of node features, $\vec{h} = \{\vec{h}_1, \vec{h}_2, ..., \vec{h}_N\}, \vec{h}_i \in R^F$, where N represents the number of nodes (EEG channels) and F represents the feature dimension of each node. To accurately depict each node's attributes, an initial linear transformation is conducted by applying a weight matrix W to every node in the graph. Following this, we compute the shared attention coefficients between each node pair:

$$e_{ij} = a(W\vec{h}_i, W\vec{h}_j) \tag{11}$$

where $W \in R^{F\times F'}, \alpha : R^{F'} \times R^{F'} \to R$, i and j denote any two nodes, F represents the input feature dimension, F' signifies the hidden unit weight, and e_{ij} quantifies the importance of association between node i and node j. In our experiments, to make the weight coefficients between different nodes more comparable, we normalize all the chosen weights using the softmax function:

$$\alpha_{ij} = softmax_j = \frac{\exp(e_{ij})}{\sum_{k\in N_i} \exp(e_{ik})} \tag{12}$$

where i, j, and k stand for any given nodes, α_{ij} indicates the weight of the edge connecting nodes i and j, N_i denotes all neighboring nodes around node i. The LeakyReLU activation function is employed, and when a detailed expansion of the formula is performed, it can be represented as:

$$\alpha_{ij} = \frac{\exp\left(LeakyReLU(\vec{a}^T[W\vec{h}_i||W\vec{h}_j])\right)}{\sum_{k\in N_i} \exp\left(LeakyReLU(\vec{a}^T[W\vec{h}_i||W\vec{h}_k])\right)} \tag{13}$$

where k represents any neighboring node of i. Lastly, utilizing a multi-head attention mechanism, we either concatenate or average the aggregated features from each head to obtain $\vec{h}'_i$.

$$h_i' = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in N_i}\alpha_{ij}W^k h_j\right) \tag{14}$$

Where k stands for any head, i represents any node, and $\vec{h}_i'$ represents the new node feature, producing a new set of node features at this layer, $\vec{h}' = \{\vec{h}_1', \vec{h}_2', ...\vec{h}_N'\}, \vec{h}_i' \in R^{F'}$.

After the original EEG time series undergoes two GAT layers, a new set of node features with spatial information is outputted. These features are then put into the subsequent Att-Bi-GRU module for processing to yield the final classification.

Figure 3 elucidates how a channel's feature $\vec{h}_1$ is transformed into a new feature $\vec{h}_1'$ through the multi-head attention mechanism. Herein, arrows of three different colors represent three distinct attention heads. $\vec{\alpha}_i$ depict normalized attention coefficients calculated by various attention heads. Features are finally aggregated from each head and then concatenated or averaged to obtain $\vec{h}_1'$.

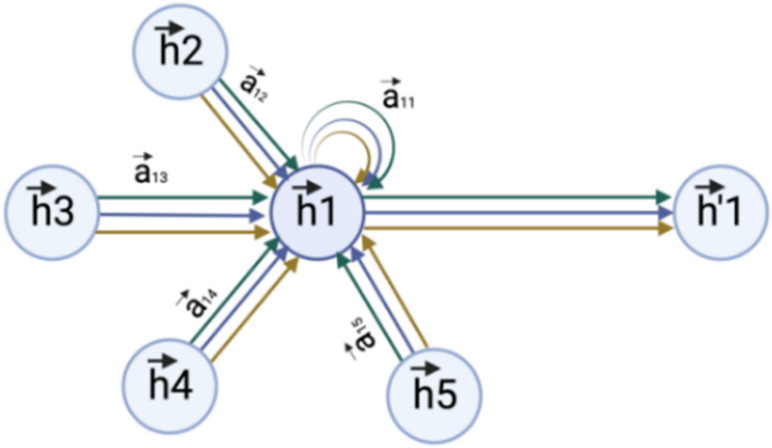


Figure 3. Multi-head attention mechanism in the GAT model

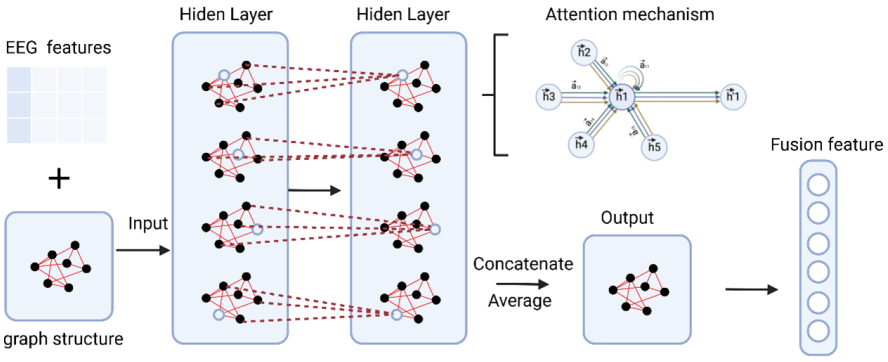


Figure 4. GAT network model diagram

The above description details the process of modeling the EEG signal using Graph Attention Network. Some advantages of using GAT can be summarized as follows:

1. Node Relationship Modeling: Effectively maps complex spatial dependencies in EEG signals as graph structure, optimizing spatial feature expressiveness.

2. Adaptive Attention Mechanism: Learns contextual node significance, enhancing focus on epileptic-related features.

3. Multi-head Attention Mechanism: Parallel computation of multiple attention weights, capturing diverse node relationships, beneficial for multi-band, multi-scale EEG analysis.

2) Attention-based Bidirectional Gated Recurrent Unit

The choice of a Bidirectional Gated Recurrent Unit (Bi-GRU) for EEG signal processing is pivotal due to its capacity to capture the full context of EEG signals, encompassing both forward and backward sequential relationships. The dual-directional design and gate mechanisms of the Bi-GRU facilitate pinpointing sequences linked with the onset and cessation of the pathological condition, or indicating its progression. Consequently, the bi-directionality of Bi-GRU ensures that the model comprehends the context from both directions, aiding in a comprehensive understanding of epilepsy features within EEG signals. Following the processing of each sliding window by GAT, the extracted features are relayed to the Bi-GRU module. The Bi-GRU module allows the model to additionally derive time-related information from the EEG data and produce novel features. These features are then processed through a fully connected layer, employing an activation function, for epilepsy detection.

Given that the Bi-GRU yields the "degree of influence" among the output information at each time step uniformly, we considered connecting an attention layer after the Bi-GRU layer. This attention layer allocates a weight for each time step, with the weight symbolizing the importance of that step in the epilepsy signal detection task. Let H be the matrix composed of output vectors, and T be the number of time steps. For the output $H = [\vec{h_1}, \vec{h_2}, \vec{h_3}, ...\vec{h_T}]$ of the Bi-GRU layer, the attention weights α can be calculated using the following formula:

$$M = tanh(H) \tag{15}$$

$$\alpha = softmax(w^T M) \tag{16}$$

$$r = H\alpha^T \tag{17}$$

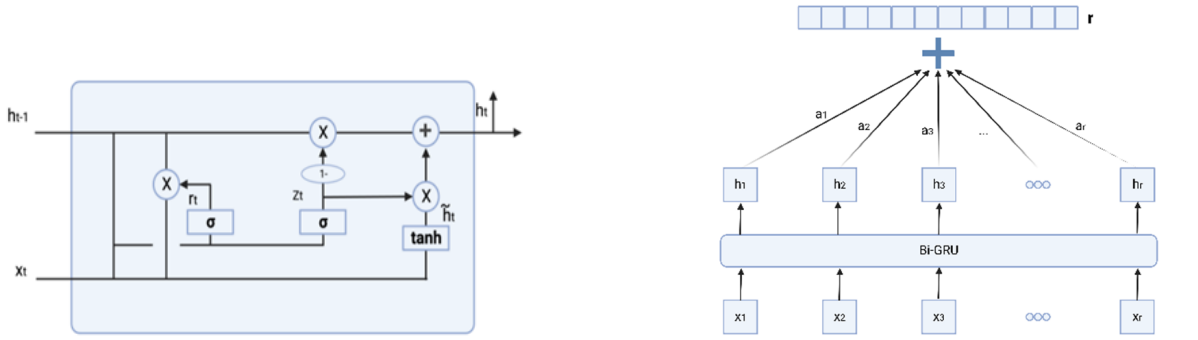
Where w is the training parameter vector, w is dotted with M, and combined with softmax to ensure all attention weights sum to 1. r, as the context vector, captures the weighted information of the entire input sequence.

$$h^* = tanh(r) \tag{18}$$

Then one can obtain h^* through the hyperbolic tangent function of r. The output of the hyperbolic tangent function is limited to the interval between -1 and 1. This is beneficial for stabilizing model training as it can prevent gradient explosion.

Subsequently, the context vector h^* is passed through a fully connected layer to convert its output dimension to 1, using the sigmoid activation function to determine the probability of the positive class. This probability is compared with a predefined threshold to determine the final classification.

3. **Experiments.**



(a) GRU unit internal structure (b) Bidirectional GRU model with Attention

Figure 5. Intrinsic architecture of GRU and comprehensive configuration of Att-Bi-GRU

3.1. **Experimental Preparation.**

1) Dataset

For our research, we chose the CHB-MIT dataset, an EEG dataset that is openly accessible and was jointly created by the Massachusetts Institute of Technology (MIT) and Boston Children's Hospital (CHB). This dataset comprises long-term electroencephalogram recordings of 24 patients diagnosed with epilepsy. Out of the 24 patients, datasets for 22 individuals comprises EEG recordings from both seizure and non-seizure periods, with the length of each recording varying from 20 minutes to several hours. Data collection was carried out through 23 channels in accordance with the international 10-20 system for electrode arrangement, augmented by an extra reference electrode. Each EEG recording was sampled at a rate of 256Hz. The CHB-MIT dataset displays a wide range of diversity, encompassing differences in patient age (from 1.5 to 22 years), gender, and epilepsy classification. Hence, any findings derived from this dataset possess significant generalizability.

2) Experiment Design

EEG recordings of 22 patients are labeled from CHB01 to CHB22. Each EEG recording captures information from 23 channels, named as follows: "FP1-F7", "F7-T7", "T7-P7", "P7-O1", "FP1-F3", "F3-C3", "C3-P3", "P3-O1", "FP2-F4", "F4-C4", "C4-P4", "P4-O2", "FP2-F8", "F8-T8", "T8-P8-0", "P8-O2", "FZ-CZ", "CZ-PZ", "P7-T7", "T7-FT9", "FT9-FT10", "FT10-T8", and "T8-P8-1".

3) Experimental Details

We constructed graphs for each patient's EEG data, considering the 23 channels as the graph's nodes. We chose the Wavelet Correlation method, instead of other approaches like Pearson's correlation coefficient, Euclidean Distance, and PLV, in gauging the degree of inter-channel correlation. Consequently, we created graphs encapsulating spatial relationships, which then served as the input for the front-end network DGAT.

As raw EEG data are typically lengthy, to capture spatial features more effectively, we utilized a sliding window of one-second duration to segment the 23 channels. Given the paucity of seizure windows in the raw data, we set the sliding window's overlap rate at 0.5. Our design involved a two-layer GAT structure, with the first layer encompassing four attention heads, processing each sliding window input. Multi-head attention mechanisms compute features between each channel and its neighbors, deriving normalized attention coefficients. These coefficients assign diverse weights to each channel, facilitating information aggregation across channels. By weighted integration of this information,

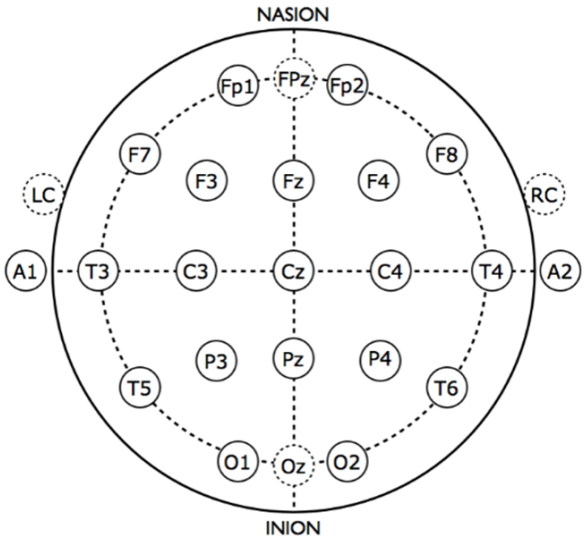


Figure 6. International 10-20 standard electrode

each channel's features are updated and subsequently fed into the second GAT layer, encompassing eight attention heads, to generate outputs.

As of post GAT module processing, features enriched with spatial information are input into the Att-Bi-GRU module to extract temporal features. Through Att-Bi-GRU's bidirectional feature capturing and attention-weighting mechanism, new features are extracted and finally undergo binary classification.

For data preprocessing, we applied ICA to remove artifacts and Z-score normalization to standardize data features, making them suitable for neural network training. ICA was chosen for its effectiveness in removing eye blink artifacts and other noise signals, while Z-score normalization ensured consistent data scaling.

The EEG data dimensions for each window are 23×256. The model was trained using a batch size of 64 for 100 epochs with a learning rate of 0.001. We chose the Adam optimizer for its stability and speed in converging, and focal loss was used as the loss function due to its suitability for sample imbalance problem.

The experimental code was developed using Python 3.8.0, built upon the PyTorch Geometric library, which is an extension of the deep learning framework PyTorch. The experiments were conducted on a platform running the Windows 10 operating system, equipped with a GeForce RTX 3090 GPU and an i5-12400 CPU, supported by 32GB of RAM.

4) Optimizer and Training Strategy

We chose the Adam optimizer for model parameter updates with an initial learning rate of 0.0001. To further enhance the model's generalization capabilities, we incorporated a learning rate decay strategy. In particular, should there be no enhancement in the model's accuracy on the validation set after 10 successive epochs, the learning rate will be adjusted to a tenth of its initial value. Hyperparameters β_1, β_2, and ϵ are set to 0.9, 0.999, and $1e{-}8$, respectively, signifying the exponential decay rates for the first and second-moment estimates, and a nominal constant for ensuring numerical stability. We employed the dropout technique in both GAT and Att-Bi-GRU with a rate of 0.5. Considering training stability and computational efficiency, we select a batch size of 32 for training. The early stopping strategy is activated when validation does not improve over 20 consecutive epochs. Furthermore, we applied a ten-fold cross-validation method to ensure the accuracy and trustworthiness of our results.

5) Evaluation Metrics

1. *Accuracy*: This represents the ratio of samples predicted correctly by the model to the total number of samples. It is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

where TP, TN, FP, FN denote True Positive, True Negative, False Positive, and False Negative, respectively.

2. *Sensitivity* (True Positive Rate, TPR): This represents the proportion of actual positive samples that are predicted correctly. It is calculated as:

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{20}$$

3. *Specificity* (True Negative Rate, TNR): Represents the proportion of actual negative samples that are predicted correctly. It is calculated as:

$$Specificity = \frac{TN}{(TN + FP)} \tag{21}$$

4. *F1-score*: Provides a comprehensive evaluation that simultaneously considers both precision and recall.

Precision, also termed as the positive predictive value, indicates the percentage of correctly predicted positive instances among all instances predicted as positive. The calculation formula is:

$$Precision = \frac{TP}{TP + FP} \tag{22}$$

Recall (Recall): Represents the proportion of actual positive samples that are predicted as positive. It is calculated as:

$$Recall = \frac{TP}{TP + FN} \tag{23}$$

The F_1-score is then calculated as:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{24}$$

3.2. **Channel Correlation Algorithm experiments.** The calculation of the graph's adjacency matrix is a pivotal step, as it determines the degree to which spatial correlations between different EEG signal channels are assessed.

We employed four methods to calculate different adjacency matrices for patients: Pearson correlation coefficient, Euclidean Distance, Wavelet Correlation, and phase-locking value. Through these adjacency matrices and node features, we constructed the graph structure of EEG data. The experiment was conducted with identical training setups and assessment criteria to guarantee the impartiality of the comparative analysis. The adjacency matrix calculated for patient CHB01 is illustrated in Figure 7.

The results of the experiments are shown in Table 1. Each method demonstrated notable efficacy. Particularly, the Wavelet Correlation algorithm yielded superior outcomes across all measured metrics when contrasted with conventional algorithms like Euclidean Distance and Pearson's correlation coefficient. Moreover, while the accuracy gain was somewhat minimal compared to PLV, there was a discernible enhancement in Specificity and F1-score, with the specificity increasing by 2.95% and the F1-score by 3.13% relative

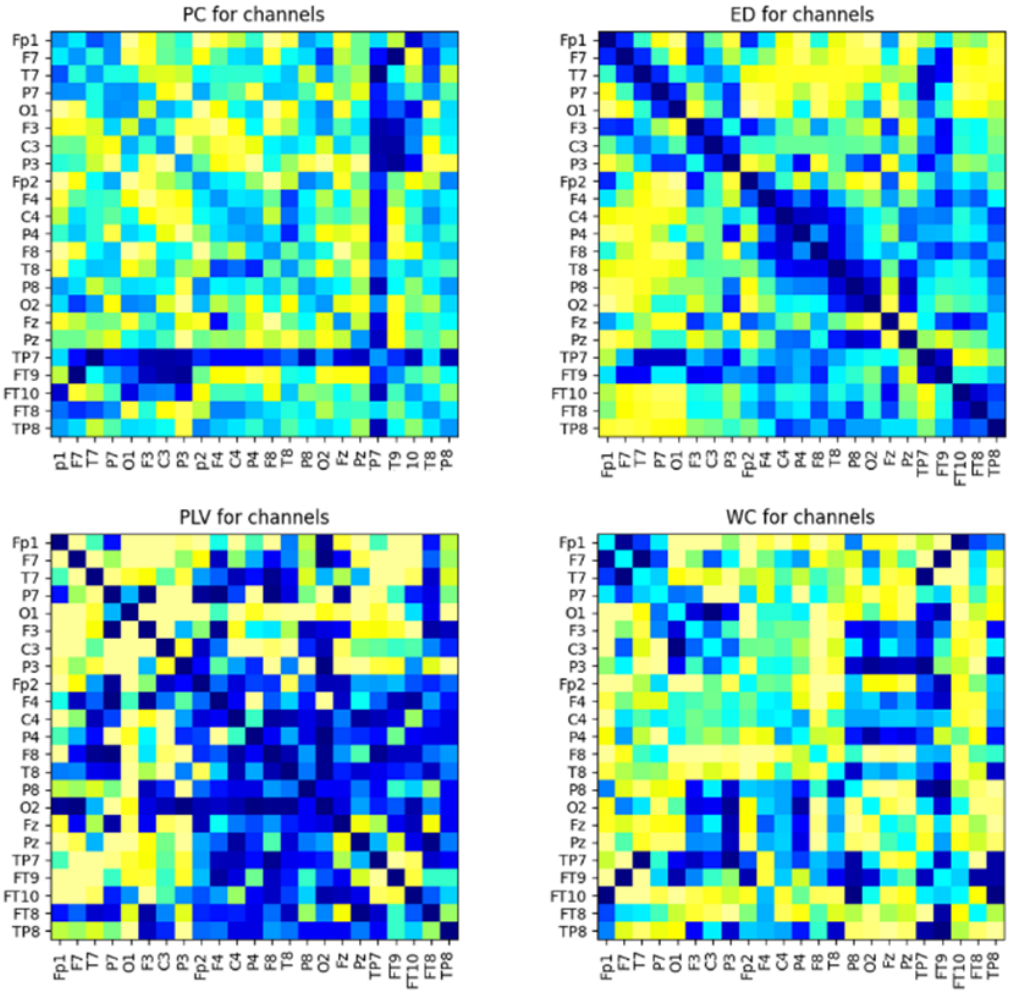


Figure 7. Average adjacency matrix calculated using different correlation algorithms for patient CHB01

Table 1. Results of comparative experiments on channel correlation methods

Method	**Accuracy**	**Sensitivity**	**Specificity**	**F1-score**
ED	96.69%	92.45%	94.46%	92.68%
PC	97.56%	92.77%	95.54%	93.84%
WC	99.21%	98.72%	99.50%	98.55%
PLV	98.28%	95.77%	96.55%	95.42%

to the PLV algorithm. A plausible explanation for this performance improvement is that wavelet analysis can capture signal variations across different frequency scales. This endows Wavelet Correlation with a significant advantage in capturing the intricate dynamics of EEG signals. Therefore, in the following series of experiments, we will choose WC as the correlation algorithm.

3.3. **Experiments of Specific Patients on the CHB-MIT.** We executed tailored experiments for each individual patient using the CHB-MIT dataset. In this approach, the EEG data from each patient were uniquely utilized to train and evaluate our model. We assessed the EEG data from all 22 patients using five key performance indicators, with detailed results presented in Table 2.

Table 2. Patient-specific experiments in seizure detection tasks using the FF-DGAT model

Patient	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**F1-score**
CHB01	99.81	100	99.66	99.20
CHB02	99.72	99.11	100	99.60
CHB03	98.89	98.55	100	98.66
CHB04	99.83	100	99	99.50
CHB05	98.81	96.55	100	98.25
CHB06	99.52	98.50	100	99.24
CHB07	100	99.93	100	100
CHB08	100	100	100	98.00
CHB09	94.87	95.17	100	96.12
CHB10	99.83	100	99	99.50
CHB11	99.2	100	95.2	97.47
CHB12	97.5	100	97.75	98.84
CHB13	96.38	93.91	91.49	94.93
CHB14	98.20	99.27	99.56	99.00
CHB15	99.38	99.44	100	99.80
CHB16	99.83	100	99	99.50
CHB17	100	98.24	100	98.78
CHB18	99.92	98.45	99.27	98.94
CHB19	98.80	96.55	100	98.24
CHB20	96.38	97.78	97.49	97.54
CHB21	100	100	100	99.66
CHB22	98.64	97.56	98.76	97.45
Mean±std	**99.21±1.02**	**98.72±0.98**	**99.50±1.10**	**98.55±1.12**

Most patients (from CHB01 to CHB22) achieved an Accuracy exceeding 98%, with a significant number reaching beyond 99%. This indicates the model's extraordinary precision in detecting epileptic seizures. Additionally, the model exhibited nearly or exactly 100% Sensitivity and Specificity across most patients, a testament to its efficacy in correctly identifying genuine epileptic episodes while ruling out non-seizure events. The F1 scores, serving as a balanced measure of precision and recall, were notably high for most patients, demonstrating the model's adeptness at maintaining a harmonious balance between these two critical metrics.

While the model's performance was predominantly exemplary, a few patients (like CHB09, CHB13, and CHB20) showed slightly lower performance metrics, probably because of variations among individuals or the intricate nature of certain scenarios. The mean values (Meanstd) reflected a remarkably stable performance across all patients, indicated by a low standard deviation. This uniformity highlights the model's robustness and its consistent performance across diverse patients.

3.4. **Comparative Experiments.**

1) Comparison with Machine Learning Methods

Table 3 demonstrates the superiority of our model in comparison to other models. Specifically, when juxtaposed with the traditional Wavelet + SVM method, our accuracy improved by 2.34%, while sensitivity and specificity improved by 5.73% and 1.37%, respectively. Relative to the Discrete Wavelet Transform (DWT) technique, our approach

demonstrated an increase of 10.02% in accuracy, 10.33% in sensitivity, and 9.88% in specificity. One salient advantage of our FF-DGAT model over these conventional methodologies is its ability to learn the spatiotemporal features from EEG data, capturing intricate associations between distinct brain regions and bypassing the complexities intrinsic to manual feature design.

Table 3. Comparative experiments

Author	**Method**	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
Chen	DWT	89.01	88.39	89.62
Janjarasjitt	Wavelet+SVM	96.87	92.99	98.13
Bizopoulos [17]	1D-CNN	85.30	91.55	90.31
Ahmedt	LSTM	95.54	91.83	90.50
Roy [22]	CNN-RNN	95.22	92.77	93.51
Chen [20]	a 3-layer GRU	96.67	94.68	95.88
Chen [20]	LSTM	96.82	92.31	93.67
Zhang [33]	Bi-GRU	98.49	93.89	98.49
LawHern [35]	EEGnet	90.54	90.16	92.85
Chen [29]	GCN	98.56	98.56	97.96
Zhao [34]	GAT	98.89	97.10	99.63
Covert [27]	T-GCN	98.69	98.21	98.14
Wei [28]	3D-CNN	93.83	94.15	90.20
OURS	**FF-DGAT**	**99.21**	**98.72**	**99.50**

2) Comparison with Deep Learning Approaches

In comparison to the CNN model introduced by Bizopoulos [17], FF-DGAT approach yielded a commendable 13.91% higher accuracy. With respect to Zhang L's Bi-GRU technique [33], while our method's accuracy marginally improved by 0.72%, there was a significant 4.83% surge in sensitivity. When set against Chen's three-layer GRU and LSTM methods [20], we marked 2.54% and 2.39% enhancement in accuracy. It is apparent that our proposed approach reliably surpasses established deep learning frameworks, including CNN and LSTM, across key performance indicators such as accuracy, sensitivity, and specificity.

3) Comparison with Deep Learning Methods Extracting Spatial Information

Our approach's superiority in spatial feature extraction was highlighted when compared against other deep learning methods emphasizing spatial information. Xiaoyan Wei's 3D-CNN strategy [28], although innovative in its consideration of inter-channel relationships by treating spatial coordinates as an additional dimension, did not match up in terms of accuracy. While Chen Xin's Graph Convolution Network effectively acknowledges spatial inter-channel relationships [29], it is observed that in various performance metrics, Chen Xin's model is slightly outpaced by FF-DGAT model. This edge in our model's performance is primarily attributed to its innovative adjacency matrix formulation, which transcends the traditional correlation algorithms employed in Chen's approach. Zhao's GAT model [34], specifically tailored for epilepsy detection, paralleled our model in accuracy and specificity. However, our approach demonstrated a slight edge in sensitivity. Notably, while Zhao's model ingeniously incorporated attention mechanisms, it did not adequately consider the temporal intricacies inherent in time series data and the extraction of frequency domain features.

In summary, the improvements and performance exhibited by the FF-DGAT method are commendable. The experimental results stand testament to the efficacy of our approach in the realm of epilepsy detection.

3.5. **Ablation Experiments.** To further substantiate the effectiveness of the FF-DGAT model, an ablation study was conducted. This study compared four different methodologies: the Bi-GRU, the Att-Bi-GRU, a combination of GNN and Att-Bi-GRU, a combination of GAT and Att-Bi-GRU, and our FF-DGAT model. Each method was tested on the CHB-MIT dataset, employing a uniform Wavelet Correlation algorithm for inter-channel correlation assessment and identical model parameter settings. The results are detailed in Table 4.

The standalone Bidirectional Gated Recurrent Unit (Bi-GRU), though effective to some extent, failed to fully consider the spatial attributes of EEG signals, thus limiting its accuracy in epilepsy detection. The Att-Bi-GRU showed improvement in processing temporal sequences but overlooked the importance of spatial relationships. The GNN+Att-Bi-GRU method, attempting to merge spatial and temporal information, fell short due to its lack of attention mechanisms within the graph neural network, resulting in inaccurate weight distribution among channels and inadequate integration of these two types of information. The GAT+Att-Bi-GRU [36] method effectively integrates spatial information and utilizes attention mechanisms to allocate weights to nodes. However, due to its end-to-end model architecture simplifying the feature extraction process, it does not operate on the frequency domain of the original EEG signals, a crucial step in epilepsy detection. Therefore, its performance is not as optimal as our dual GAT structure that simultaneously considers both the temporal and frequency aspects. In contrast, FF-DGAT model not only integrates temporal and frequency domain information of EEG signals but also enhances the utilization of inter-channel relationships using the graph attention network, thereby markedly excelling beyond alternative approaches in terms of accuracy, sensitivity, specificity, and F1-score.

Table 4. Ablation study

Method	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**F1-score**
Bi-GRU	92.57	93.21	94.58	98.03
Att-Bi-GRU	95.69	93.45	94.77	98.33
GNN+Att-Bi-GRU	97.49	93.89	97.45	98.32
GAT+Att-Bi-GRU	98.14	95.69	97.66	98.21
OURS	**99.21**	**98.72**	**99.50**	**98.55**

3.6. **Hyper-Parameters Experiments.** Tests were carried out to assess how the chosen thresholds for the wavelet coherence method affect the model's efficacy in generating the adjacency matrix. As depicted in Figure 8, optimal outcomes were obtained with the threshold at 0.5.

3.7. **Hyper-Parameters Experiments.** To evaluate the robustness of FF-DGAT, we assessed the influence of pivotal parameters on the model's accuracy. This included an examination of the number of GAT layers, the quantity of attention heads, and the dimensions of hidden units. Table 5 outlines the results.

Initially, we observed that the FF-DGAT model demonstrated a relatively low sensitivity to variations in the number of GAT layers and attention heads. Fine-tuning these parameters could marginally enhance the optimal outcomes. Specifically, the incorporation of additional GAT layers and attention heads appeared to exert a negligible effect on the model's accuracy. Notably, in our experimental setup, the employment of two GAT layers resulted in an accuracy improvement of 1.60%.

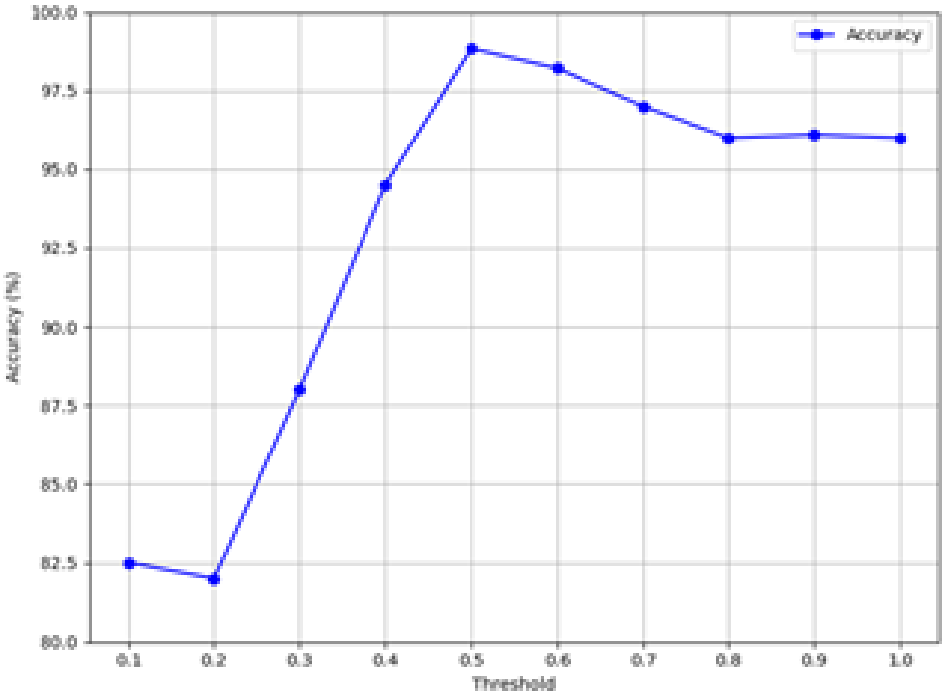


Figure 8. Effect of different thresholds on model accuracy when the threshold value is 0.1-1.0

Table 5. Impact of parameters on accuracy

Layers of Dual GAT	Heads of muti-attention	Hidden size	Accuracy%
1	4	64	97.32
2	4,4	64	98.93
3	4,4,4	64	98.96
2	4,4	64	98.93
2	4,8	64	98.97
2	8,8	64	99.01
2	4,4	2	53.52
		4	68.83
		8	78.81
		16	84.93
		32	98.32
		64	99.21
		128	99.53

In the context of multi-head attention layers, an increase in the number of heads was found to correlate with higher accuracy. However, this escalation was accompanied by diminishing returns regarding performance gains and a surge in computational demands. Balancing considerations of model efficacy and computational efficiency, we ultimately decided on the deployment of four attention heads in each GAT layer.

Furthermore, an initial augmentation in the number of hidden units was found to substantially enhance the model's performance, propelling it towards a more stable operational state. This can be interpreted as the insufficient number of hidden units failing to adequately extract information, resulting in information loss. Conversely, when the

hidden layer size reached an adequate threshold, the incremental benefits of further expansion were minimal, while the computational costs escalated exponentially. Considering the stochastic characteristics of the training phase, our experiments utilized a setup with 64 hidden units.

4. **Conclusion.** This study introduces an efficient automatic epilepsy detection framework, FF-DGAT, based on DGAT leveraging wavelet channel correlation and Att-Bi-GRU networks. Within this framework, DGAT serves as a front-end network, handling raw EEG signals and frequency domain information, to achieve a successful integration of multi-scale features for effective epilepsy detection. Building on this, the incorporation of the Att-Bi-GRU network further integrates temporal sequence information, culminating in a deep amalgamation of EEG features. It is this tri-dimensional integration of EEG features that endow the model with exceptional accuracy, specificity, and sensitivity in epilepsy detection. Testing on the CHB-MIT dataset has demonstrated that, to the best of our knowledge, FF-DGAT surpasses the performance of current deep learning models documented in the literature, showcasing its cutting-edge capability in the field of epilepsy detection.

Despite these advances, the present study is not without limitations. The computational complexity of our model is higher than simpler deep learning models due to the incorporation of dual graph attention networks and multi-scale feature extraction. Additionally, while our model achieves high performance on the CHB-MIT dataset, its performance on other datasets or in real-world scenarios where EEG data may be noisier or come from different demographics has not been thoroughly evaluated.

For future work, we plan to conduct a thorough analysis of subtle variations in EEG signals, and to improve feature extraction algorithms, such as extracting mutual information, Shannon entropy, Sample entropy, and others, to more accurately detect subtle epileptic characteristics. We also aim to enhance our model's ability to accurately classifying different seizure types.

Acknowledgment. This work is partially supported by the Key Research and Development Program of Yunnan Province, (No. 202102AA100021).

REFERENCES

[1] E. Pippa, E.-I. Zacharaki, I. Mporas, V. Tsirka, M.-P. Richardson, M. Koutroumanidis, and V. Megalooikonomou, "Improving classification of epileptic and non-epileptic EEG events by feature selection," *Neurocomputing*, vol. 171, pp. 576-585, 2016.

[2] Q. Yuan, W. Zhou, Y. Liu, and J. Wang, "Epileptic seizure detection with linear and nonlinear features," *Epilepsy & Behavior*, vol. 24, no. 4, pp. 415-421, 2012.

[3] S.-O. Wietstock, S.-L. Bonifacio, J.-E. Sullivan, K.-B. Nash, and H.-C. Glass, "Continuous Video Electroencephalographic (EEG) Monitoring for Electrographic Seizure Diagnosis in Neonates A Single-Center Study," *Journal of Child Neurology*, vol. 31, no. 3, pp. 328-332, 2016.

[4] E. Alickovic, J. Kevric, and A. Subasi, "Performance Evaluation of Empirical Mode Decomposition, Discrete Wavelet Transform, and Wavelet Packed Decomposition for Automated Epileptic Seizure Detection and Prediction," *Biomedical Signal Processing and Control*, vol. 39, pp. 94-102, 2018.

[5] M. Khodatars, A. Shoeibi, D. Sadeghi, N. Ghaasemi, M. Jafari, P. Moridian, A. Khadem, R. Alizadehsani, A. Zare, Y. Kong, A. Khosravi, S. Nahavandi, S. Hussain, U.-R. Acharya, and M. Berk, "Deep learning for neuroimaging-based diagnosis and rehabilitation of autism spectrum disorder: a review," *Computers in Biology and Medicine*, vol. 139, p. 104949, 2021.

[6] A. Craik, Y. He, and J.-L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *Journal of Neural Engineering*, vol. 16, no. 3, p. 031001, 2019.

[7] J. Thomas, J. Jin, P. Thangavel, E. Bagheri, R. Yuvaraj, J. Dauwels, R. Rathakrishnan, J.-J. Halford, S.-S. Cash, and B. Westover, "Automated detection of interictal epileptiform discharges from scalp electroencephalograms by convolutional neural networks," *International Journal of Neural Systems*, vol. 30, no. 11, p. 2050030, 2020.

[8] R.-T. Schirrmeister, J.-T. Springenberg, L.-D.-J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391-5420, 2017.

[9] Y. Li, Z. Yu, Y. Chen, C. Yang, Y. Li, X. Allen Li, and B. Li, "Automatic seizure detection using fully convolutional nested LSTM," *International Journal of Neural Systems*, vol. 30, no. 04, p. 2050019, 2020.

[10] D. K. Thara, B. G. PremaSudha, and F. Xiong, "Epileptic seizure detection and prediction using stacked bidirectional long short term memory," *Pattern Recognition Letters*, vol. 128, pp. 529-535, 2019.

[11] M. Ahmadlou and H. Adeli, "Complexity of weighted graph: A new technique to investigate structural complexity of brain activities with applications to aging and autism," *Neuroscience Letters*, vol. 650, pp. 103-108, 2017.

[12] J. DelEtoile and H. Adeli, "Graph theory and brain connectivity in Alzheimer's disease," *The Neuroscientist*, vol. 23, no. 6, pp. 616-626, 2017.

[13] M. Ahmadlou, H. Adeli, and A. Adeli, "Graph theoretical analysis of organization of functional brain networks in ADHD," *Clinical EEG and Neuroscience*, vol. 43, no. 1, pp. 5-13, 2012.

[14] M. Ahmadlou, H. Adeli, and A. Adeli, "New diagnostic EEG markers of the Alzheimer's disease using visibility graph," *Journal of Neural Transmission*, vol. 117, pp. 1099-1109, 2010.

[15] M. Ahmadlou, H. Adeli, and A. Adeli, "Improved visibility graph fractality with application for the diagnosis of autism spectrum disorder," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 20, pp. 4720-4726, 2012.

[16] O. Faust, Y. Hagiwara, T.-J. Hong, O.-S. Lih, and U.-R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Computer Methods and Programs in Biomedicine*, vol. 161, pp. 1-13, 2018.

[17] P. Bizopoulos, G. I. Lambrou, and D. Koutsouris, "Signal2image modules in deep neural networks for EEG classification," in *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 702-705, 2019.

[18] M. Golmohammadi, S. Ziyabari, V. Shah, S. L. de Diego, I. Obeid, and J. Picone, "Deep architectures for automated seizure detection in scalp EEGs," *arXiv preprint arXiv:1712.09776*, 2017.

[19] D. Ahmedt-Aristizabal, C. Fookes, K. Nguyen, and S. Sridharan, "Deep classification of epileptic signals," in *Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 332-335, 2018.

[20] X. Chen, J. Ji, T. Ji, and P. Li, "Cost-sensitive deep active learning for epileptic seizure detection," in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 226-235, 2018.

[21] M. Golmohammadi, S. Ziyabari, V. Shah, S. L. de Diego, I. Obeid, and J. Picone, "Deep Architectures for Automated Seizure Detection in Scalp EEGs," *arXiv: Learning*, 2017.

[22] S. Roy, I. Kiral-Kornek, and S. Harrer, "Deep learning enabled automatic abnormal EEG identification," in *Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2756-2759, 2018.

[23] D.-P. Dash, M.-H. Kolekar, and K. Jha, "Multi-channel EEG based automatic epileptic seizure detection using iterative filtering decomposition and Hidden Markov Model," *Computers in Biology and Medicine*, vol. 116, p. 103571, 2020.

[24] A. Bhattacharya, T. Baweja, and S. P. K. Karri, "Epileptic seizure prediction using deep transformer model," *International Journal of Neural Systems*, vol. 32, no. 02, p. 2150058, 2022.

[25] U.-R. Acharya, S.-L. Oh, Y. Hagiwara, J.-H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Computers in Biology and Medicine*, vol. 100, pp. 270-278, 2018.

[26] L. Vidyaratne, A. Glandon, M. Alam, and K. M. Iftekharuddin, "Deep recurrent neural network for seizure detection," in *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 1202-1207, 2016.

[27] I.-C. Covert, B. Krishnan, I. Najm, J. Zhan, M. Shore, J. Hixson, and M.-J. Po, "Temporal graph convolutional networks for automatic seizure detection," in *Proceedings of the Machine Learning for Healthcare Conference*, pp. 160-180, 2019.

[28] X. Wei, L. Zhou, Z. Chen, L. Zhang, and Y. Zhou, "Automatic seizure detection using three-dimensional CNN based on multi-channel EEG," *BMC Medical Informatics and Decision Making*, vol. 18, p. 71, 2018.

[29] X. Chen, Y. Zheng, Y. Niu, and C. Li, "Epilepsy classification for mining deeper relationships between EEG channels based on GCN," in *Proceedings of the 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, pp. 701-706, 2020.

[30] D. Zeng, K. Huang, C. Xu, H. Shen, and Z. Chen, "Hierarchy graph convolution network and tree classification for epileptic detection on electroencephalography signals," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 4, pp. 955-968, 2020.

[31] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *Stat*, vol. 1050, no. 20, p. 10-48550, 2017.

[32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[33] Y. Zhang, S. Yao, R. Yang, X. Liu, W. Qiu, L. Han, and W. Shang, "Epileptic seizure detection based on bidirectional gated recurrent unit network," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 135-145, 2022.

[34] Y. Zhao, G. Zhang, C. Dong, Q. Yuan, F. Xu, and Y. Zheng, "Graph attention network with focal loss for seizure detection on electroencephalography signals," *International Journal of Neural Systems*, vol. 31, no. 07, p. 2150027, 2021.

[35] V.-J. Lawhern, A.-J. Solon, N.-R. Waytowich, S.-M. Gordon, C.-P. Hung, and B.-J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 2018.

[36] Y. Wang, Y. Shi, Y. Cheng, Z. He, X. Wei, Z. Chen, and Y. Zhou, "A spatiotemporal graph attention network based on synchronization for epileptic seizure prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 2, pp. 900-911, 2022.