A Heterogeneous Feature Fusion Network for EEG Seizure Detection Integrating Spatio-Temporal Channel Networks and Vision Transformer

Zhiqiang Chang, Qiaoli Zhou*, Bin Guo, Tao Hong
School of Computer Science, Shenyang Aerospace University, Shenyang, China
E-mail: changzhiqiang@stu.sau.edu.cn, zhouql@sau.edu.cn*, guobin1@stu.sau.edu.cn,hongtao1@stu.sau.edu.cn

***Abstract*—Epileptic seizures pose a significant challenge to patient quality of life, highlighting the need for efficient and automated detection systems to support clinical decision-making. Electroencephalogram (EEG) signals, as a critical diagnostic modality, require models that can effectively capture temporal, spatial, and frequency-domain features, along with their complex interdependencies. However, existing methods often fall short in comprehensively modeling these multi-dimensional characteristics. To address this limitation, we propose a Heterogeneous Feature Fusion Network (HFFNet), which integrates a Spatio-Temporal Channel Network (STCN) and a Vision Transformer (ViT) pre-trained model for robust epilepsy detection. Specifically, raw EEG signals and their corresponding STFT-transformed spectrograms are used as dual-path inputs to extract complementary features. STCN utilizes Distributed Residual LSTM and Graph Neural Network layers to model spatio-temporal dependencies and inter-channel relationships, while ViT captures global frequency-domain representations from the spectrograms. The two subnetworks are independently trained to minimize mutual interference and are subsequently integrated through an attention-based feature fusion module. Experimental results on three public datasets—CHB-MIT, TUSZ, and Helsinki—demonstrate that HFFNet achieves state-of-the-art performance and significantly outperforms existing approaches.Notably, we adopt leave-one-subject-out (LOSO) validation for subject-independent evaluation, confirming the model's strong generalization capability in real-world clinical scenarios.**

***Index Terms*—EEG, Epilepsy detection, Heterogeneous Neural Network, Inter-channel relationship, Dual Input Path, Vision Transformer**

I. INTRODUCTION

Epilepsy is a chronic neurological disorder that affects individuals of all ages. Approximately 50 million people worldwide live with this condition, making it one of the most common neurological diseases. Studies have shown that with timely and appropriate treatment, up to 70% of patients can achieve long-term seizure control [1].

Electroencephalogram (EEG) is widely used for epilepsy diagnosis due to its non-invasive nature and low cost. Compared to imaging methods such as fMRI and MEG [2], EEG offers superior temporal resolution and directly reflects neuronal dynamics, making it a critical tool for seizure detection [3]. However, analyzing EEG signals remains challenging due to the complexity and variability of seizure manifestations. Seizures may be focal or generalized, with distinct spatial propagation characteristics [4]; abnormal waveforms such as spikes and sharp waves appear inconsistently across channels [5]; and seizure evolution spans interictal, preictal, and ictal phases, where preictal signals offer only a narrow prediction window [6]. Moreover, seizure unpredictability and strong inter-subject variability further complicate automatic detection [7].

To address these challenges, there is a growing need for advanced recognition frameworks that can effectively capture spatiotemporal patterns, model inter-channel dependencies, and exploit complementary time–frequency information. Such frameworks should leverage multi-view EEG features to support robust automated seizure identification and provide reliable assistance for clinical decision-making.

The organization of this paper is as follows. Section II introduces the datasets, preprocessing methods, and the proposed HFFNet framework. Section III describes the experimental setup and implementation details. Section IV presents and discusses the experimental results. Finally, Section V concludes the paper and outlines future research directions.

A. Related Work

Deep learning (DL) has substantially advanced EEG-based seizure detection, outperforming traditional machine learning methods that rely on handcrafted features [8]. Enabled by large datasets and increased computational power, DL has become the dominant paradigm for EEG classification.

Several DL architectures have been applied to EEG analysis. Zhou et al. [9] introduced a CNN that processes raw EEG signals directly. While effective in capturing spatial features, CNNs lack the capacity to model long-range temporal dependencies. To address this, LSTM networks were introduced: Ahmedt-Aristizabal et al. [10] proposed an LSTM-based model to learn temporal patterns in EEG signals, and Abdelhameed et al. [11] further improved performance using Bi-LSTM, which processes data in both directions. However, RNN-based models still suffer from limited memory and inefficiency when modeling long EEG sequences.

More recently, attention-based models, particularly Transformers [12], have attracted increasing interest. These architectures apply self-attention mechanisms to dynamically weight input components, enabling the modeling of global dependencies. Transformer-based methods have demonstrated competitive performance in EEG-based seizure detection [13],

[14], and offer a promising alternative to purely convolutional or recurrent architectures.

Transfer learning has also gained traction as a means of addressing data scarcity and generalization issues. CNNs pretrained on natural images, such as AlexNet, ResNet, or VGG, can be adapted to EEG tasks by fine-tuning output layers or reusing feature extractors [15]. Raghu et al. [16] retrained the final layers of multiple pretrained CNNs for seizure detection and achieved superior results. Similarly, Thara et al. [17] and Nogay et al. [18] demonstrated that adapted CNNs outperform conventional approaches. Nevertheless, CNN-based transfer learning remains limited in modeling temporal dynamics [19], motivating the adoption of Transformer-based models and more expressive temporal encoders.

Another promising direction involves multi-feature fusion, which combines time-domain, frequency-domain, and time–frequency features such as power spectral density (PSD), wavelet transform (DWT), STFT, and nonlinear descriptors like entropy and recurrence plots [20], [21]. Kong et al. [20] integrated DWT-based and Welch-based features, while Sun et al. [21] combined Hilbert spectra with recurrence plots. Beyond feature-level fusion, hybrid models have emerged: Radman et al. [22] fused features via Dempster–Shafer theory and employed ensemble trees for classification, and Ali et al. [23] proposed ConTraNet, a hybrid CNN–Transformer model that jointly captures local and global patterns, especially in low-data scenarios. Despite these efforts, many fusion strategies rely on hand-engineered features or loosely coupled modules, and often lack a unified way to jointly model heterogeneous temporal and time–frequency representations.

Finally, the evaluation strategy plays a critical role in EEG-based seizure detection research [24]. Due to high inter-subject variability, the validation protocol significantly impacts reported generalization. Three major strategies are used: intra-subject, inter-subject (or cross-subject), and subject-independent validations. Among them, leave-one-subject-out (LOSO) is widely recognized as the most rigorous method [25], as it excludes subjects from training and evaluates on entirely unseen individuals, offering a realistic assessment of clinical applicability [26]. However, many existing works still rely on less stringent protocols, which can overestimate performance and hinder fair comparison across methods.

B. Limitations

Despite notable progress in EEG-based seizure detection, several limitations remain:

• **Inadequate spatiotemporal and multi-view modeling:** Most existing methods still struggle to capture hierarchical seizure dynamics across time and channels. CNNs mainly focus on local patterns with limited global context, LSTMs are inefficient for long sequences, and GNN-based approaches may yield weak inter-channel representations when temporal dynamics are not sufficiently modeled beforehand. Moreover, multi-feature fusion is often implemented by concatenation or averaging, which fails to explicitly learn interactions among spatial, temporal, and time–frequency cues.

• **Limited robustness and generalization:** Model performance is often sensitive to dataset-specific factors (e.g., electrode layouts, sampling rates, and recording conditions) and may degrade under subject-independent or cross-dataset settings. In addition, lenient validation protocols (e.g., intra-subject or random splits) can overestimate real-world performance. Hence, architectures that explicitly improve generalization under heterogeneous EEG conditions are still required.

C. Paper contribution

This work proposes a Heterogeneous Feature Fusion Network (HFFNet) for epileptic seizure detection, consisting of a Spatio-Temporal Channel Network (STCN) and a Time-Frequency Transfer Learning Network (TFTLN). STCN combines Distributed Residual LSTM (DRLSTM) and a Graph Neural Network (GNN) to jointly model per-channel temporal dynamics and inter-channel spatial dependencies. TFTLN adopts a pretrained Vision Transformer (ViT) to extract global time–frequency representations from STFT spectrograms. The extracted features are integrated via an attention-based fusion module. Our main contributions are as follows:

- **STCN:** We design a hybrid subnetwork that integrates DRLSTM and GNN. DRLSTM models each EEG channel independently to preserve distinct temporal patterns and reduce inter-channel interference, while residual connections stabilize long-sequence learning. The GNN further captures dataset-specific inter-channel dependencies through electrode adjacency modeling.
- **TFTLN:** We incorporate a pretrained ViT backbone to model global time–frequency dependencies from STFT spectrograms. Transfer learning enhances representation quality under limited supervision, improving robustness and generalization on unseen data.
- **Dual-path feature extraction:** We introduce a dual-path architecture that jointly exploits raw EEG and time–frequency representations, enabling complementary temporal–spatial and spectral feature learning in a modular and scalable manner.
- **Attention-based feature fusion:** We independently train STCN and TFTLN with temporary task heads, and then fuse their extracted features using a self-attention module, which encourages specialization and yields more informative joint representations for robust seizure detection across datasets.

II. Materials and Methods

This section provides a description of the publicly accessible datasets used in this study. Next, we introduce the method of data preprocessing, followed by a detailed explanation of the Spatio-Temporal Channel Networks (STCN) and the Time-Frequency Transfer Learning Network (TFTLN) utilized in the dual-path architecture. Finally, we elaborate on the feature fusion strategy employed to integrate the outputs from both subnetworks. Fig. 1 illustrates the overall block diagram of our proposed method.

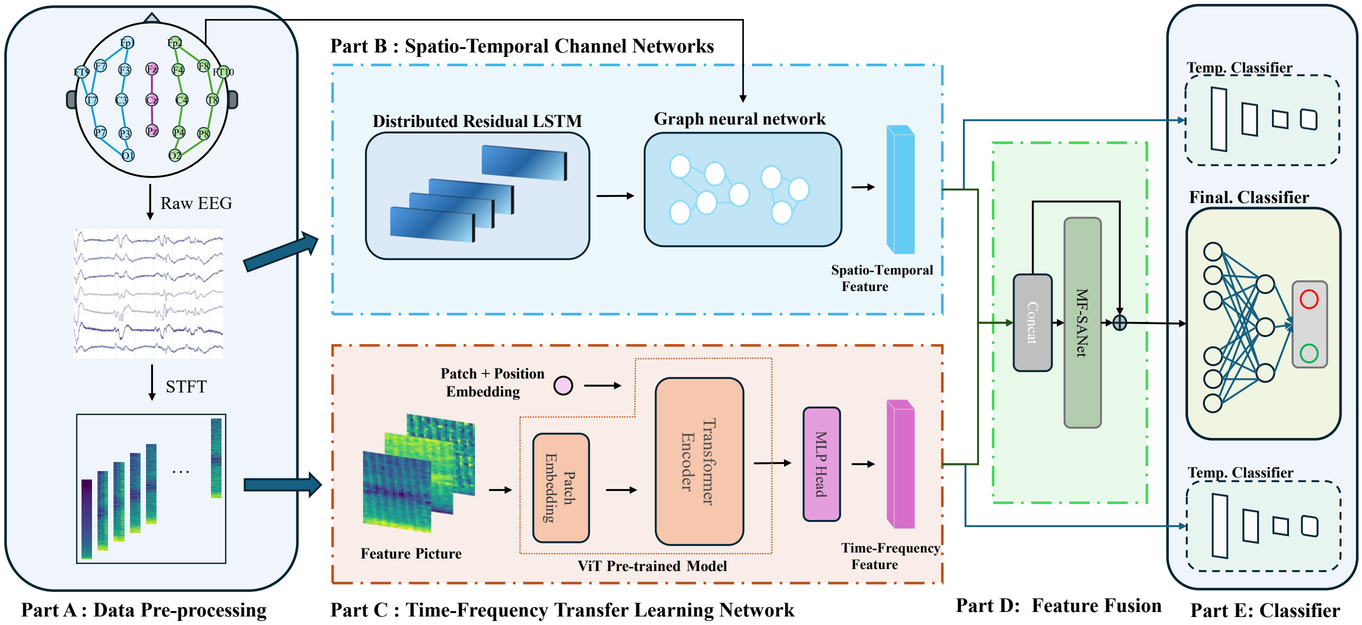

Fig. 1: The overall block diagram of the proposed method for epilepsy detection.

TABLE I: Dataset Characteristics

Dataset	No. of Patients	Recording Length per Segment	Number of Events/Segments	Total Segments	Sampling Freq. (Hz)
CHB-MIT Scalp EEG	22	Typically 1 h	198 events	664	256
Helsinki Univ. NICU	79	74 min (IQR: 64–96)	$\sim$460 events per expert	87	256
TUSZ (v1.5.2)	300+	$\leq$ 1 h (pruned)	3050 events	5612	$\geq$ 250

The HFFNet architecture consists of five main components, as shown in Fig. 1: a data processing module (Part A), STCN (Part B), the TFTLN (Part C), a feature fusion module (Part D), and a classification module (Part E). The data processing module includes two procedures: first, the raw EEG data is directly fed into STCN; second, the STFT is applied to the raw EEG signals, followed by linear interpolation, to generate spectrogram images. The STCN framework integrates DRLSTM and GNN to effectively capture spatio-temporal dependencies across EEG channels. Meanwhile, TFTLN utilizes a ViT pre-trained on ImageNet to extract high-level spectral representations from EEG spectrograms, thereby enhancing the understanding of frequency-domain characteristics. The extracted features from STCN and TFTLN are then aggregated by the feature fusion module, which is based on attention mechanisms. This module optimizes the combined representation by emphasizing the most relevant and complementary features from both domains. Finally, the classification module utilizes the fused features to generate the final classification results.

A. Dataset

We evaluate HFFNet on three public EEG datasets: CHB-MIT [27], Helsinki NICU [28], and TUSZ v1.5.2 [29]. Their statistics are summarized in Table I.

B. Data Pre-processing

As shown in Fig. 2, we adopt a unified dual-stream pre-processing pipeline to generate both time-domain segments for STCN (Part B) and STFT spectrograms for TFTLN (Part C). All recordings are standardized into fixed-length windows, with samples labeled as seizure/non-seizure based on provided annotations. To mitigate class imbalance, seizure segments are augmented using overlap sampling, while non-seizure segments are downsampled.

For TFTLN (Part C), each segment is transformed by STFT using a Hann window ($N_w = 512$) and 50% overlap ($N_{\text{step}} = 256$), producing a compact time-frequency representation. The spectrogram is then reshaped and interpolated into a $32 \times 32 \times Ch$ tensor and projected to a $32 \times 32 \times 3$ pseudo-RGB input via a 2D convolution, enabling compatibility with the ViT backbone.

C. Spatio-Temporal Channel Networks (STCN)

To effectively learn deep representations from multi-channel EEG waveform data, directly concatenating the raw inputs from all channels is insufficient to preserve the unique characteristics of each channel. Inspired by advancements in distributed deep learning and graph-based techniques, we propose the use of DRLSTM and GNN within the STCN module. Part B illustrates the framework of STCN.

1) Distributed Residual LSTM: To effectively model EEG signals, it is crucial to extract informative temporal features

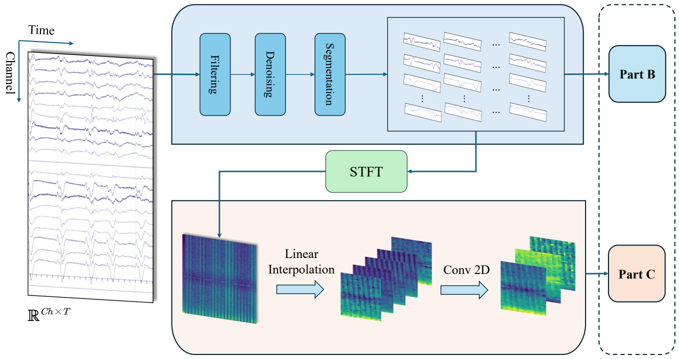

Fig. 2: The pipeline of Data Pre-processing

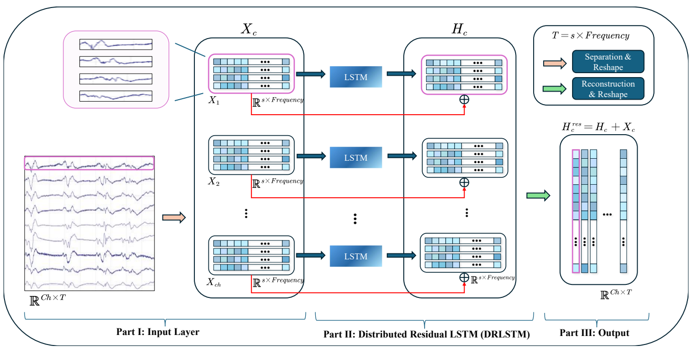

Fig. 3: The detailed architecture of DRLSTM

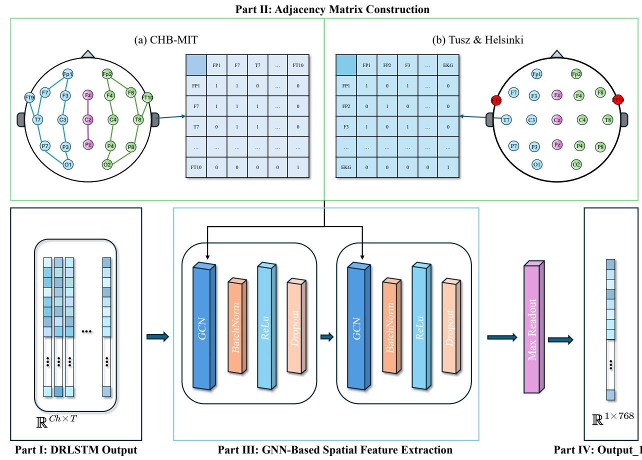

Fig. 4: The detailed architecture of GNN

while minimizing interference between different channels. To address this, we introduce a DRLSTM model that applies residual learning to mitigate vanishing gradients and enhance feature extraction from multi-channel data.

As shown in Figure 3, the DRLSTM model processes multi-channel EEG signals by splitting the raw data into independent LSTM layers, with the number of layers corresponding to the number of EEG channels. Each LSTM layer operates independently on a single-channel input, ensuring that temporal characteristics are preserved while preventing mutual interference.

In Part I, the raw EEG signal is split into multiple single-channel signals, each with shape $s \times$ Frequency, where s represents the time slice length (4 seconds) and Frequency is the sampling rate (256 Hz).

In Part II, single-channel segments are fed into independent LSTM layers to extract temporal dependencies. To enhance long-term dependency modeling, residual connections are introduced, summing the output of each LSTM layer with its input:

$$H_c^{\text{res}} = H_c + X_c \tag{1}$$

where H_c is the hidden state output of the c^{th} EEG channel, and X_c is the corresponding input feature.

In Part III, the outputs from all LSTM layers are merged to reconstruct a multi-channel representation, maintaining the shape $Ch \times T$. This refined temporal feature representation is passed to the subsequent GNN for spatial feature extraction.

By employing this distributed processing strategy, DRLSTM captures temporal dependencies within individual channels while enabling subsequent spatial feature extraction.

2) Graph Neural Network: Seizure detection relies not only on temporal dependencies but also on spatial relationships among EEG channels, which reflect functional interactions between brain regions. While traditional methods use channel-specific attention mechanisms, they often fail to capture global spatial dependencies. To address this, we employ Graph Neural Networks (GNNs) to model inter-channel relationships, treating electrodes as nodes and their connections as edges for structured spatial feature learning.

The feature propagation mechanism follows the standard Graph Convolutional Network (GCN) formulation:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{2}$$

Here, $H^{(l)} \in \mathbb{R}^{N\times d}$ denotes the node feature matrix at the l-th GCN layer, where N is the number of EEG channels (i.e., graph nodes) and d is the input feature dimension. The input feature $H^{(0)}$ is derived from the DRLSTM output with a dimension of $N \times T$.

The GNN module, positioned after the DRLSTM layer (Fig. 4), models the spatial dependencies among EEG channels. Part I of Fig. 4 shows the output of DRLSTM as a $Ch\times T$ feature matrix.The adjacency matrix, constructed based on the channel connectivity strategies of the datasets (Part II of Fig. 4), defines spatial relationships among EEG electrodes.

GNN uses two GCN layers (Part III of Fig. 4) to iteratively refine spatial representations. The first layer captures local spatial dependencies with the adjacency matrix $\tilde{A}$, and the second expands the feature dimension to 768 to model higher-order interactions. Each GCN layer is followed by BatchNorm, ReLU, and Dropout modules to enhance training stability.

To aggregate the spatial features from all nodes, a Max Readout operation selects the most discriminative features

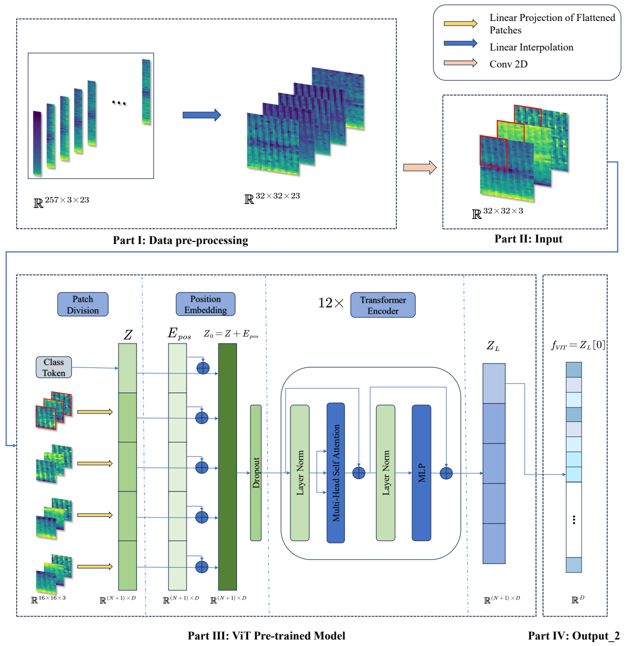

Fig. 5: The framework of TFTLN.

across the EEG channels. The final 1×768 feature vector combines the temporal features from DRLSTM and spatial features from GNN, serving as input to the classifier for seizure detection.

D. Time-Frequency Transfer Learning Network

Epileptic seizures exhibit complex dynamics across time, frequency, and spatial domains, which are difficult to capture directly from raw EEG. We transform each EEG segment into a spectrogram using the STFT and adopt a pretrained Vision Transformer (ViT) to model global dependencies in the time–frequency domain. The ViT-B16, pretrained on large-scale image datasets, has its classification head removed, and the encoder is reused as a feature extractor, enabling effective transfer learning with limited labeled EEG data.

As shown in Fig. 5, the STFT–ViT pipeline includes four stages: data preprocessing (Part I), ViT-compatible transformation (Part II), ViT feature extraction (Part III), and final feature representation (Part IV). In Parts I–II, the spectrogram is resized by linear interpolation and passed through a 2D convolution to obtain an input tensor of size $32 \times 32 \times 3$, encoding time–frequency information with compressed multi-channel EEG.

In Part III, the ViT-B16 encoder processes the transformed spectrogram. The input $\mathbf{X} \in \mathbb{R}^{32\times32\times3}$ is divided into $N = 4$ non-overlapping patches of size 16×16, each flattened and projected into a D-dimensional token ($D = 768$). A learnable class token $\mathbf{z}_{\text{cls}}$ is prepended, and positional encodings $\mathbf{E}_{\text{pos}}$ are added, forming the input sequence $\mathbf{Z}_0 \in \mathbb{R}^{5\times768}$.

This sequence is then fed into 12 standard Transformer encoder layers, each consisting of Layer Normalization, Multi-Head Self-Attention (MSA), and a Feed-Forward Network

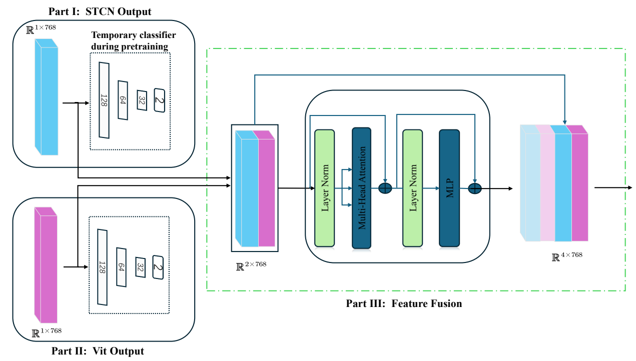

Fig. 6: The detailed architecture of FeatureFusion

with residual connections and dropout, capturing long-range dependencies across the time–frequency spectrum.

In Part IV, the encoder outputs a sequence $\mathbf{Z}_L \in \mathbb{R}^{5\times768}$. The final class token is taken as the spectrogram-level representation $\mathbf{f}_{\text{ViT}} \in \mathbb{R}^{768}$, which serves as input to downstream seizure detection modules. Leveraging the pretrained ViT-B16 encoder, this TFTLN models global spectral–temporal structure, improving generalization to unseen subjects in EEG-based seizure detection.

E. Feature Fusion

In epileptic seizure detection, leveraging multi-modal features from both temporal and time-frequency domains is crucial for capturing complex EEG dynamics. Traditional fusion methods often suffer from feature redundancy or interference. To address this, we propose a dual-path architecture with two independently trained sub-networks, fused using an attention-based mechanism for effective cross-modal integration (Fig. 6).

The fusion module receives inputs from two sub-networks: Part I (STCN) models temporal and spatial features, while Part II (TFTLN) extracts time-frequency representations using ViT. Both sub-networks produce 1×768 feature vectors, trained with temporary classification heads and discarded after training for fusion.

In the fusion stage (Part III in Fig. 6), the feature vectors are concatenated into a 2×768 matrix, passed into a Multi-Head Self-Attention (MHSA) module with 8 attention heads to capture cross-modal dependencies. The output is normalized, processed by a two-layer FFN with ReLU activation, and passed through a classification head for seizure detection.

III. Experimental Setup

We conducted experiments to evaluate the effectiveness of HFFNet on three public EEG datasets: CHB-MIT, Helsinki, and TUSZ. For CHB-MIT and Helsinki, we employed LOSO validation to ensure strict subject-independent evaluation, while TUSZ was evaluated using ten-fold cross-validation. The training hyperparameters included separate learning rates for STCN and TFTLN (0.00006 and 0.00003, respectively) and a dropout rate of 0.3. Both branches were trained for 70 epochs

TABLE II: Performance on CHB-MIT Dataset Using LOSO Validation

ID	Accuracy	Sensitivity	Specificity	ID	Accuracy	Sensitivity	Specificity
1	99.02%	99.15%	98.86%	**14**	93.91%	100.00%	87.81%
2	97.62%	98.27%	96.99%	**15**	96.10%	97.53%	94.76%
3	95.07%	97.12%	93.06%	**17**	95.58%	97.52%	89.22%
4	91.91%	92.56%	91.27%	**18**	93.44%	91.50%	95.37%
5	96.99%	95.61%	98.37%	**19**	96.27%	96.55%	96.00%
6	95.30%	94.30%	96.20%	**20**	93.46%	90.53%	96.39%
7	97.14%	98.69%	95.76%	**21**	92.53%	93.68%	91.42%
8	94.12%	95.78%	92.45%	**22**	99.04%	99.90%	97.11%
9	97.39%	97.33%	97.45%	**23**	97.60%	97.66%	97.55%
10	98.38%	99.27%	97.49%	**24**	94.82%	98.29%	91.35%
11	96.30%	95.80%	96.73%	**Avg.**	**95.81%**	**95.95%**	**95.32%**

TABLE III: Performance Comparison with Previous Works Using LOSO Validation on CHB-MIT

Authors	Method	Sensitivity (%)	Specificity (%)	Accuracy (%)
Jiang et al. [30]	KNN	–	–	74.03
Zhou et al. [31]	SOF	85.00	81.73	–
Emran et al. [32]	Random Forest	75.34	–	–
Lawhern et al. [33]	EEGNet-8,2	–	–	88.41
Wei et al. [34]	CNN	72.11	**95.89**	84.00
Hossain et al. [35]	CNN	90.00	91.65	**98.05**
Zhao et al. [36]	Adversarial Deep Learning	77.42	–	–
Jana et al. [37]	CapsNet	55.63	56.98	56.07
Li et al. [38]	Adaptive F-G TM	84.75	92.31	91.92
Ali et al. [39]	CNN + FC + RF + SVM	90.76	–	–
Mohammad et al. [40]	SPERTL (1D ResNet)	up to 88	–	–
Chio et al. [41]	DANN	71.09±26.17	85.75±14.49	–
Zhang et al. [42]	DGCL	74.10	72.36	73.67
Ours	**Proposed Method**	**95.95**	95.32	95.81

with a batch size of 256 on CHB-MIT and Helsinki, and for 100 epochs with a batch size of 256 on TUSZ. We report Accuracy, Sensitivity (Sn), Specificity (Sp), and AUC to assess seizure detection performance under class imbalance.

IV. Results

We evaluated HFFNet through comparative experiments against representative state-of-the-art approaches and conducted ablation studies to analyze the contributions of key components.

A. Comparative experiment

This subsection summarizes the experimental results and compares HFFNet with representative baselines under a binary seizure detection setting. Results are reported in Tables II, IV, and VI. CHB-MIT and Helsinki are evaluated using LOSO validation (Tables II and VI), while TUSZ is assessed with ten-fold cross-validation (Table IV).

On CHB-MIT, HFFNet achieves 95.81% accuracy, 95.95% sensitivity, and 95.32% specificity, demonstrating strong cross-subject generalization under LOSO. On TUSZ, HFFNet reaches 99.00% accuracy, 97.21% sensitivity, and 99.14% specificity, indicating stable performance under cross-validation. On Helsinki neonatal EEG, HFFNet obtains 73.48% sensitivity, 79.06% specificity, and 82.48% AUC, showing competitive robustness despite artifacts and short seizure durations. Notably, several subjects (e.g., ID 34) exhibit near-saturated performance, reflecting the potential of the model under favorable signal quality.

CHB-MIT Dataset: HFFNet improves sensitivity to 95.95%, exceeding the best comparable reported sensitivity by approximately 5 percentage points, while maintaining balanced specificity (95.32%). Although the accuracy (95.81%) is below the reported peak accuracy of 98.05% by Hossain et al., HFFNet provides stronger sensitivity under strict patient-independent evaluation.

TABLE IV: Experimental Results Using Ten-Fold Cross-Validation on TUSZ

Fold	Accuracy	Sensitivity	Specificity	Fold	Accuracy	Sensitivity	Specificity
1	98.89%	98.68%	98.91%	6	98.76%	98.21%	98.80%
2	99.26%	100.00%	99.20%	7	98.64%	93.06%	99.19%
3	99.14%	98.25%	99.20%	8	99.63%	100.00%	99.60%
4	98.89%	94.74%	99.20%	9	99.01%	100.00%	98.93%
5	98.76%	97.01%	98.92%	10	99.01%	92.16%	99.47%
Avg.	**99.00%**	**97.21%**	**99.14%**				

TABLE V: Performance comparison with previous works using ten-fold cross-validation on TUSZ.

Authors	Method	Sensitivity (%)	Specificity (%)	Accuracy (%)
Henni et al. [43]	Random Forest	–	–	86.80
Zhang et al. [44]	Random Forest	73.01	93.92	–
Zhang et al. [45]	DWT-Net	59.07	97.05	–
Thuwajit et al. [46]	EEGWaveNet	59.21±22.23	75.30±13.97	–
Khan et al. [47]	Modified CNN	95.05–97.07	94.56–94.91	94.98–95.79
Statsenko et al. [48]	CNN+LSTM	87.70	91.16	–
Wu et al. [49]	DTGCN	–	–	87.30
Hu et al. [50]	IGGCN	91.80	–	91.80
Xie et al. [51]	MRCNN	–	–	95.73
Ours	**Proposed Method**	**97.21**	**99.14**	**99.00**

TUSZ Dataset: HFFNet substantially improves accuracy over Henni et al. (86.80%), achieving 99.00%. It also attains 97.21% sensitivity, slightly higher than the previously reported upper bound (97.07%), and reaches 99.14% specificity, surpassing Zhang et al.'s reported 97.05%.

Helsinki Dataset: Compared with the LOSO baseline by Frassineti et al., HFFNet improves sensitivity from about 63% to 73.48% (+10.48%) and increases AUC from 81 to 82.48 (+1.48), demonstrating better subject-independent generalization.

B. Ablation studies

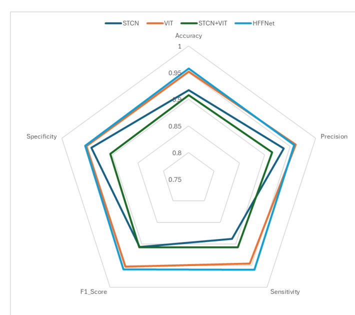

Fig. 7: The results of the ablation experiments on CHB-MIT

We conducted an ablation study to assess the contribution of each component in HFFNet. The STCN-only model achieved 91.69% accuracy and 94.16% specificity, but obtained a lower sensitivity of 88.80%. The TFTLN branch (pretrained on ImageNet) performed better than STCN, reaching 95.18% accuracy, 94.52% sensitivity, and 96.11% precision, indicating stronger time–frequency representation learning from spectrograms.

Directly combining the two branches (STCN+TFTLN) led to a performance drop, with 90.70% sensitivity and 90.41%

TABLE VI: Performance on Helsinki Dataset

ID	Sensitivity	Specificity	AUC	ID	Sensitivity	Specificity	AUC
1	90.80%	56.35%	78.57%	39	83.69%	86.41%	91.48%
4	87.97%	73.11%	84.59%	40	56.16%	93.15%	90.20%
5	90.08%	64.46%	85.81%	41	68.75%	81.25%	80.12%
7	72.20%	80.34%	83.26%	44	90.48%	96.60%	98.25%
9	55.04%	88.76%	78.39%	47	90.58%	94.76%	95.94%
11	70.27%	72.97%	78.23%	50	58.54%	59.30%	65.01%
13	87.54%	87.70%	94.85%	51	73.61%	63.89%	77.06%
14	93.53%	88.97%	97.29%	52	95.00%	66.25%	90.55%
15	82.44%	75.61%	86.13%	62	97.61%	93.85%	97.25%
16	95.81%	68.59%	86.41%	63	47.11%	58.36%	54.06%
17	78.57%	85.71%	81.12%	66	67.83%	81.32%	81.84%
19	41.16%	90.47%	74.81%	67	80.80%	79.57%	87.09%
20	31.56%	81.90%	58.02%	69	63.71%	91.94%	86.49%
21	88.89%	88.89%	96.14%	73	30.20%	95.69%	79.11%
22	59.12%	67.57%	71.00%	75	55.46%	79.48%	76.36%
25	51.35%	81.08%	65.45%	76	71.58%	40.98%	57.75%
31	78.88%	86.96%	89.29%	77	40.32%	82.21%	70.78%
34	99.78%	100.00%	99.78%	78	69.57%	73.57%	75.77%
36	90.09%	95.95%	98.06%	79	90.24%	79.27%	92.10%
38	89.43%	50.26%	82.20%	**Avg.**	**73.48%**	**79.06%**	**82.48%**

TABLE VII: Performance comparison with previous works using LOSO validation on the Helsinki neonatal EEG dataset.

Authors	Validation	Sensitivity (%)	Specificity (%)	AUC
Tapani et al. [52]	Subject-Independent	–	–	95.5
O'Shea et al. [53]	Subject-Independent	–	–	95.6
Frassineti et al. [54]	LOSO	63	83	81
Daly et al. [55]	Subject-Independent	–	–	96.4
Nagarajan et al. [56]	Inter-subject	87.0	–	–
Ours	**LOSO**	**73.48**	**79.06**	**82.48**

specificity, suggesting feature overlap and training interference. In contrast, the full HFFNet with attention-based fusion achieved the best overall results (95.81% accuracy, 95.95% sensitivity, and 95.75% precision), validating the effectiveness of the fusion strategy.

V. Discussion

This study proposes the STCN-STFTLN dual-path neural network for epileptic seizure detection using EEG data. The framework includes data preprocessing, temporal-spatial feature extraction, ViT-based time-frequency learning, and attention-guided feature fusion. The STCN path captures temporal and spatial dependencies using DRLSTM and GNN, while the ViT path processes STFT-transformed EEG signals to capture frequency-domain features. The outputs are fused through multi-head self-attention, enabling dynamic interaction between features.

Experimental results show that the proposed method outperforms existing approaches on multiple datasets. Ablation studies confirm the importance of multi-feature input, with performance drops observed when either STCN or ViT was removed. Incorporating DRLSTM into STCN also improved accuracy over GNN alone. However, limitations such as the fixed STFT window size and the lack of domain-specific knowledge in pretrained ViT models remain. Future work will focus on adaptive time-frequency representations and EEG-specific pretrained Transformers for better feature extraction.

References

[1] W. H. Organization, "Epilepsy," 2024. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs999/en/

[2] F. Emily, A. Fazila, and W. Jessica, "Implementation of 2helps2b seizure risk score: A high-value approach to seizure detection in the icu," *Critical Care Medicine*, vol. 51, 2023.

[3] Epilepsy.com, "Eeg (electroencephalogram)," n.d. [Online]. Available: https://www.epilepsy.com/diagnosis/eeg

[4] T. J. H. University, "Types of Seizures," 2023. [Online]. Available: https://www.hopkinsmedicine.org/health/conditions-and-diseases/epilepsy/types-of-seizures

[5] C. Li, W. Zhou, and G. Liu, "Seizure onset detection using empirical mode decomposition and common spatial pattern," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 458–467, 2021.

[6] L. Ilias, D. Askounis, and J. Psarras, "Multimodal detection of epilepsy with deep neural networks," *Expert Systems with Applications*, vol. 213, p. 119010, 2023.

[7] J. Cao, Y. Ding, D. Hu, X. Lai, S. Wang, and Y. Wang, "Epileptic state classification by fusing hand-crafted and deep learning eeg features," *Circuits and Systems II: Express Briefs*, vol. 68, no. 4, pp. 1542–1546, 2021.

[8] 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.

[9] M. Zhou, C. Tian, R. Cao, B. Wang, Y. Niu, T. Hu, H. Guo, and J. Xiang, "Epileptic seizure detection based on EEG signals and CNN," *Frontiers in neuroinformatics*, vol. 12, p. 95, 2018.

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

[11] A. M. Abdelhameed, H. G. Daoud, and M. Bayoumi, "Deep convolutional bidirectional LSTM recurrent neural network for epileptic seizure detection," in *2018 16th IEEE International New Circuits and Systems Conference (NEWCAS)*. IEEE, 2018, pp. 139–143.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[13] Y. Ma, C. Liu, M. S. Ma, Y. Yang, N. D. Truong, K. Kothur, A. Nikpour, and O. Kavehei, "TSD: Transformers for Seizure Detection," *bioRxiv*, pp. 2023–01, 2023.

[14] S. Rukhsar and A. K. Tiwari, "Lightweight Convolution Transformer for Cross-patient Seizure Detection in Multi-channel EEG Signals," *arXiv preprint arXiv:2305.04325*, 2023.

[15] A. M. Azab, H. Ahmadi, L. Mihaylova, and M. Arvaneh, "Dynamic time warping-based transfer learning for improving common spatial patterns in brain–computer interface," *Journal of neural engineering*, vol. 17, no. 1, p. 016061, 2020.

[16] S. Raghu, N. Sriraam, Y. Temel, S. V. Rao, and P. L. Kubben, "Eeg based multi-class seizure type classification using convolutional neural network and transfer learning," *Neural Networks*, vol. 124, pp. 202–212, 2020.

[17] C. Tan, F. Sun, and W. Zhang, "Deep transfer learning for eeg-based brain computer interface," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2018, pp. 916–920.

[18] H. S. Nogay and H. Adeli, "Detection of epileptic seizure using pretrained deep convolutional neural network and transfer learning," *European neurology*, vol. 83, no. 6, pp. 602–614, 2021.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[20] G. Kong, S. Ma, W. Zhao, H. Wang, Q. Fu, and J. Wang, "A novel method for optimizing epilepsy detection features through multi-domain feature fusion and selection," *Frontiers in Computational Neuroscience*, vol. Volume 18 - 2024, 2024. [Online]. Available: https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2024.1416838

[21] C. Sun, C. Xu, H. Li, H. Bo, L. Ma, and H. Li, "A novel multi-feature fusion attention neural network for the recognition of epileptic eeg signals," *Frontiers in Computational Neuroscience*, vol. Volume 18 - 2024, 2024. [Online]. Available: https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2024.1393122

[22] M. Radman, M. Moradi, A. Chaibakhsh, M. Kordestani, and M. Saif, "Multi-feature fusion approach for epileptic seizure detection from eeg signals," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3533–3543, 2021.

[23] O. Ali, M. S. ur Rehman, T. Glasmachers, I. Iossifidis, and C. Klaes, "Contranet: A hybrid network for improving the classification of eeg and emg signals with limited training data," *Computers in*

Biology and Medicine, vol. 168, p. 107649, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482523011149

[24] J. Dan, U. Pale, A. Amirshahi, W. Cappelletti, T. M. Ingolfsson, X. Wang, A. Cossettini, A. Bernini, L. Benini, S. Beniczky, D. Atienza, and P. Ryvlin, "¡scp¿szcore¡/scp¿: Seizure community open-source research evaluation framework for the validation of ¡scp¿electroencephalography¡/scp¿-based automated seizure detection algorithms," *Epilepsia*, Sep. 2024. [Online]. Available: http://dx.doi.org/10.1111/epi.18113

[25] S. Kunjan, T. S. Grummett, K. J. Pope, D. M. W. Powers, S. P. Fitzgibbon, T. Bastiampillai, M. Battersby, and T. W. Lewis, "The necessity of leave one subject out (loso) cross validation for eeg disease diagnosis," in *Brain Informatics: 14th International Conference, BI 2021, Virtual Event, September 17–19, 2021, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 2021, p. 558–567. [Online]. Available: https://doi.org/10.1007/978-3-030-86993-9_50

[26] F. D. Pup, A. Zanola, L. F. Tshimanga, A. Bertoldo, L. Finos, and M. Atzori, "The role of data partitioning on the performance of eeg-based deep learning models in supervised cross-subject analysis: a preliminary study," 2025. [Online]. Available: https://arxiv.org/abs/2505.13021

[27] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[28] N. Stevenson, K. Tapani, L. Lauronen, and S. Vanhatalo, "A dataset of neonatal eeg recordings with seizure annotations," *Scientific Data*, vol. 6, p. 190039, 2019. [Online]. Available: https://doi.org/10.1038/sdata.2019.39

[29] A. Harati, S. Lopez, I. Obeid, J. Picone, M. Jacobson, and S. Tobochnik, "The tuh eeg corpus: A big data resource for automated eeg interpretation," in *2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 2014, pp. 1–5.

[30] X. Jiang, K. Xu, and W. Chen, "Transfer component analysis to reduce individual difference of eeg characteristics for automated seizure detection," in *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2019, pp. 1–4.

[31] J. Zhou, L. Liu, Y. Leng, Y. Yang, B. Gao, Z. Jiang, W. Nie, and Q. Yuan, "Both cross-patient and patient-specific seizure detection based on self-organizing fuzzy logic," *International Journal of Neural Systems*, vol. 32, no. 06, p. 2250017, 2022, pMID: 35306966. [Online]. Available: https://doi.org/10.1142/S0129065722500174

[32] A. Emran, A. Maia, and K. Chandan, "Epileptic seizure detection using chb-mit dataset: The overlooked perspectives," *R. Soc. Open Sci.*, vol. 11, p. 230601, 2024.

[33] 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, jul 2018. [Online]. Available: https://dx.doi.org/10.1088/1741-2552/aace8c

[34] Z. Wei, J. Zou, J. Zhang, and J. Xu, "Automatic epileptic eeg detection using convolutional neural network with improvements in time-domain," *Biomedical Signal Processing and Control*, vol. 53, p. 101551, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809419301259

[35] M. S. Hossain, S. U. Amin, M. Alsulaiman, and G. Muhammad, "Applying deep learning for epilepsy seizure detection and brain mapping visualization," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 15, no. 1s, Feb. 2019. [Online]. Available: https://doi.org/10.1145/3241056

[36] Y. Zhao, G. Zhang, Y. Zhang, T. Xiao, Z. Wang, F. Xu, and Y. Zheng, "Multi-view cross-subject seizure detection with information bottleneck attribution," *Journal of Neural Engineering*, vol. 19, no. 4, p. 046011, jul 2022. [Online]. Available: https://dx.doi.org/10.1088/1741-2552/ac7d0d

[37] G. C. Jana, K. Swami, and A. Agrawal, "Capsule neural network based approach for subject specific and cross-subjects seizure detection from eeg signals," *Multimedia Tools and Applications*, vol. 82, no. 23, pp. 35 221–35 252, September 2023. [Online]. Available: https://doi.org/10.1007/s11042-023-14995-w

[38] Y. Li, H. Li, Z. Zhong, and J. Zhou, "A real-time cross-patient seizure detection algorithm based on adaptive template matching," in *2023 6th International Conference on Electronics Technology (ICET)*, 2023, pp. 1122–1126.

[39] A. Esmaeilpour, S. S. Tabarestani, and A. Niazi, "Deep learning-based seizure prediction using eeg signals: A comparative analysis of classification methods on the chb-mit dataset," *Engineering Reports*, vol. 6, no. 11, p. e12918, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/eng2.12918

[40] U. Mohammad and F. Saeed, "Robustness of ml-based seizure prediction using noisy eeg data from limited channels," in *2024 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)*, 2024, pp. 620–626.

[41] W. Choi, J.-M. Kim, H. Nam, and T.-E. Kam, "Curriculum learning with sampling scheduler for imbalanced eeg-based seizure detection," in *2025 13th International Conference on Brain-Computer Interface (BCI)*, 2025, pp. 1–5.

[42] Z. Zhang and D. Sun, "Dynamic graph contrastive learning for eeg-based epileptic seizure detection," in *2025 IEEE 8th Information Technology and Mechatronics Engineering Conference (ITOEC)*, vol. 8, 2025, pp. 883–887.

[43] K. Henni, L. Abou-Abbas, I. Jmal, A. Mitiche, and N. Mezghani, "Imbalance-aware machine learning for epileptic seizure detection," in *2024 IEEE 12th International Symposium on Signal, Image, Video and Communications (ISIVC)*, 2024, pp. 1–5.

[44] D. Zhang and H. Wang, "Seizure detection based on multilayer brain network features," in *2025 8th International Conference on Advanced Algorithms and Control Engineering (ICAACE)*, 2025, pp. 1924–1928.

[45] Z. Zhang, Y. Ren, N. Sabor, J. Pan, X. Luo, Y. Li, Y. Chen, and G. Wang, "Dwt-net: Seizure detection system with structured eeg montage and multiple feature extractor in convolution neural network," *Journal of Sensors*, vol. 2020, no. 1, p. 3083910, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1155/2020/3083910

[46] P. Thuwajit, P. Rangpong, P. Sawangjai, P. Autthasan, R. Chaisaen, N. Banluesombatkul, P. Boonchit, N. Tatsaringkansakul, T. Sudhawiyangkul, and T. Wilaiprasitporn, "Eegwavenet: Multiscale cnn-based spatiotemporal feature extraction for eeg seizure detection," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 8, pp. 5547–5557, 2022.

[47] I. D. Khan, O. Farooq, and Y. U. Khan, "Automatic seizure detection using modified cnn architecture and activation layer," *Journal of Physics: Conference Series*, vol. 2318, no. 1, p. 012013, aug 2022. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/2318/1/012013

[48] Y. Statsenko, V. Babushkin, T. Talako, T. Kurbatova, D. Smetanina, G. L. Simiyu, T. Habuza, F. Ismail, T. M. Almansoori, K. N.-V. Gorkom, M. Szólics, A. Hassan, and M. Ljubisavljevic, "Automatic detection and classification of epileptic seizures from eeg data: Finding optimal acquisition settings and testing interpretable machine learning approach," *Biomedicines*, vol. 11, no. 9, 2023. [Online]. Available: https://www.mdpi.com/2227-9059/11/9/2370

[49] G. Wu, K. Yu, H. Zhou, X. Wu, and S. Su, "Time-series anomaly detection based on dynamic temporal graph convolutional network for epilepsy diagnosis," *Bioengineering*, vol. 11, no. 1, 2024. [Online]. Available: https://www.mdpi.com/2306-5354/11/1/53

[50] Y. Hu, J. Liu, R. Sun, Y. Yu, and Y. Sui, "Classification of epileptic seizures in eeg data based on iterative gated graph convolution network," *Frontiers in Computational Neuroscience*, vol. 18, p. 1454529, 2024. [Online]. Available: https://doi.org/10.3389/fncom.2024.1454529

[51] T. Xie, W. Zhao, Y. Liu, and S. Xiao, "Rdpnet: A multi-scale residual dilated pyramid network with entropy-based feature fusion for epileptic eeg classification," *Entropy*, vol. 27, no. 8, 2025. [Online]. Available: https://www.mdpi.com/1099-4300/27/8/830

[52] K. T. Tapani, S. Vanhatalo, and N. J. Stevenson, "Time-varying eeg correlations improve automated neonatal seizure detection," *International journal of neural systems*, vol. 29, no. 04, p. 1850030, 2019.

[53] A. O'Shea, G. Lightbody, G. Boylan, and A. Temko, "Neonatal seizure detection from raw multi-channel eeg using a fully convolutional architecture," *Neural Networks*, vol. 123, pp. 12–25, 2020.

[54] L. Frassineti, D. Ermini, R. Fabbri, and C. Manfredi, "Neonatal seizures detection using stationary wavelet transform and deep neural networks: Preliminary results," in *2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON)*, 2020, pp. 344–349.

[55] A. Daly, A. O'Shea, G. Lightbody, and A. Temko, "Towards deeper neural networks for neonatal seizure detection," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 920–923.

[56] V. Nagarajan, A. Muralidharan, D. Sriraman, and P. K. S, "Scalable machine learning architecture for neonatal seizure detection on ultra-edge devices," 2021. [Online]. Available: https://arxiv.org/abs/2111.15569