INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES
BULGARIAN ACADEMY OF SCIENCES
CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **25**, No 4
Sofia • 2025 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2025-0041

NeuroFusionNet: A Multi-Modal Graph Transformer with Contrastive Alignment and Evidential Uncertainty for Epileptic Seizure Detection

Riyazulla Rahman Jabiulla[1], *Afroz Pasha*[1], *Pinnepalli Sadhashiviah Prasad*[2], *Mohan Devollu Narasimhamurthy*[3], *Vidya Virupaksha*[4], *Manjula Hebbala Munithimmaiah*[5]

[1]*School of Computer Science & Information Science, Presidency University, Bengaluru, India*
[2]*Department of Computer Science and Engineering, JAIN (Deemed-to-be University), Bengaluru, India*
[3]*Department of Information Science and Engineering, SJC Institute of Technology, Chickballapur, India*
[4]*Department of Computer Science and Design, SJC Institute of Technology, Chickballapur, India*
[5]*School of Computer Science and Engineering, Presidency University, Bengaluru, India*
E-mails: riyaz@presidencyuniversity.in afrozpasha@presidencyuniversity.in prasad.ps@jainuniversity.ac.in Mohandn028@gmail.com Vid16ud@gmail.com manjulahm@presidencyuniversity.in

***Abstract*:** *Reliable epileptic seizure detection remains challenging due to the heterogeneity of modalities and poor interpretability in existing models. To address these issues, this research proposes NeuroFusionNet, a unified multi-modal framework that jointly leverages Electro-Encephalo-Gram (EEG) and functional Magnetic Resonance Imaging (fMRI) signals through modality-specific graph encoders and a Cross-Modal Graph Transformer (CMGT). The CMGT architecture captures both temporal and spatial-functional dynamics, enabling robust feature learning across modalities. Additionally, a modality-wise contrastive alignment objective is employed to ensure latent consistency, then an evidential uncertainty head is also incorporated, which assists in estimating clinical reliability for calibrated confidence. Hence, the model demonstrates strong generalization across CHB-MIT, resting-state (rs)-fMRI from UW–Madison, and 7 T fMRI datasets. Finally, the proposed NeuroFusionNet achieved higher results with 99.22% accuracy, 99.89% precision, and 99.85% recall, outperforming the existing TriSeizureDualNet model. These results determine that the proposed NeuroFusionNet is interpretable and trustworthy for seizure detection.*

***Keywords*:** *Cross-Modal Graph Transformer (CMGT), Electro-Encephalo-Gram (EEG), Epileptic seizure detection, Functional Magnetic Resonance Imaging (fMRI), NeuroFusionNet.*

1. Introduction

Epilepsy is a chronic neurological disorder characterized by recurrent unprovoked seizures due to abnormal electrical activity in the brain [1]. These seizures have a multitude of manifestations, ranging from a transient lapse of awareness to significant

body convulsions, and have profound effects on an individual's behaviors, cognition, and consciousness [2]. Globally, epilepsy affects more than 50 million individuals across all age groups and geographical regions [3]. Epilepsy has several etiologies, including genetic mutations, traumatic brain injuries, and idiopathic mechanisms. Consequently, the cause of epilepsy may complicate the diagnosis and treatment of this disorder [4]. These challenges highlight the importance of developing effective detection systems that are reliable, interpretable, and generalizable [5].

The detection of the epileptogenic zone is a challenging process because of the complexity of focal epilepsy and the partial nature of individual imaging modalities [6]. Specifically, in structural Magnetic Resonance Imaging (MRI), subtle lesions such as focal cortical dysplasia remain undetected, especially in patients with MRI-negative findings, thereby underscoring the need for more sensitive diagnostic modalities. [7]. In addition, the complexity of high-dimensional Electro-Encephalo-Graph (EEG) data leads to more classification problems because irrelevant and redundant features reduce performance accuracy and increase computational cost [8]. Hence, using patient-independent seizure detection strategies can improve generalizability; however, they are adversely affected by inter-individual differences and the non-stationary characteristics of EEG signals, which hinder generalization performance [9]. Although Deep Learning (DL) deployed within IoT-cloud frameworks enables real-time monitoring, optimal feature selection remains a problem that constrains both classification performance and system efficiency [10]. Epileptic seizure prediction has gained considerable research attention because early forecasting can prevent potential serious incidents and ensure patient welfare through timely intervention [11]. Although various data-driven approaches have been explored, they were not able to capture the complex spatiotemporal dependencies in EEG signals, highlighting the need for more expressive DL architectures [12]. Additionally, recent studies have emphasized hybrid models that integrate temporal and spatial feature extraction to provide seizure detection models with the most robust and generalizable representations as possible [13]. Further, research has been conducted on secure processing for sensitive data, such as blockchain, that assists in enhancing the clinical applicability by resolving the issues of trust and privacy concerns that are present in sensitive medical situations [14]. Furthermore, these advancements collectively emphasize the broader requirement for intelligent, secure, and multi-modal architectures that can change the process of traditional seizure diagnosis from retrospective case-based development toward patient-centered monitoring systems [15]. Hence, the objective of this research is to develop a robust, interpretable, and generalizable seizure detection framework that integrates multimodal data and resolves the issue of dataset variability, optimizes model efficiency, and provides clinically deployable solutions for epilepsy detection.

Therefore, this research proposes NeuroFusionNet that incorporates EEG and fMRI modalities with uncertainty-aware decision-making for robust, generalizable, and interpretable classification of epileptic events.

The key contributions of this research are:

- The proposed NeuroFusionNet model incorporates a multi-modal graph-based dual-stream architecture that models EEG and functional Magnetic Resonance Imaging (fMRI) data as independent graphs, which assists in preserving the modality-specific temporal and spatial-functional dynamics for seizure characterization.
- A Cross-Modal Graph Transformer (CMGT) is introduced to jointly learn intra and inter-modality dependencies by utilizing self and cross-attention mechanisms, which allows effective information exchange between heterogeneous neural modalities.
- Modality-wise contrastive alignment is employed to resolve the issue of distributional differences across EEG and fMRI. Thus, this operation aligns semantically similar representations across modalities without requiring strict spatial or anatomical correspondence.
- Uncertainty-aware classification through evidential deep learning is performed with an evidential output head that predicts class probabilities along with epistemic uncertainty that enables the model to quantify its confidence and enhance reliability in clinical decision-making.

The overall research is organized as follows: Section 2 demonstrates the related works, and Section 3 describes the proposed NeuroFusionNet methodology. Section 4 illustrates the experimental results, and the discussion and conclusion of this research are given in Section 5.

2. Related works

Recent research on epileptic seizure detection has advanced substantially across Deep Learning (DL), hybrid optimization frameworks, and high-resolution neuroimaging. This section demonstrates existing research based on the respective methodological architecture.

Recent innovations have introduced transformer-based networks as they can model long-range dependencies across temporal, spatial, A, and spectral dimensions in seizure-related signals. Hence, one notable contribution recommended a Signal Embedding Temporal-Spatial-Spectral Transformer (SE-TSS-Transformer), which was designed for SEEG signals that focused on comprehensively capturing seizure propagation features across multiple scales [16]. Consequently, the model utilized an embedding module for dimensionality reduction and a unified multiscale transformer encoder for joint domain learning. Thus, this approach's strength lies in the global feature relevance learned through transformers. However, due to the dependence on a specialized SEEG dataset, this approach obstructed reproducibility across broader clinical contexts.

Similarly, TriSeizureDualNet employed a multimodal transformer-based approach that combined EEG and fMRI inputs [17]. This architecture utilized a triple-stream skipped feature extractor and a Dual Parallel Attention Transformer (DPAT), which enabled context-aware fusion of heterogeneous modalities. Further, the model incorporated Hummingbird Optimization for feature selection and classified seizure types into focal, general, and unknown onsets. Nonetheless, the dependency on dual-

modal data, such as CHB-MIT for EEG and UW-Madison for fMRI, limited the applicability in single-modality clinical settings.

Further, CNN-based frameworks were dominated due to the strong feature learning capabilities from raw EEG data. Hence, a notable design, named ultra-lightweight 3D-CNN, utilized temporal-spatial separable convolutions and depth-wise separable layers to reduce parameter complexity without sacrificing accuracy [18]. Therefore, this operation made it suitable for wearable or edge-based applications. Despite its practicality, the model's generalizability across different EEG sources or seizure types remained unverified. Another approach employed a dual-stream CNN trained on phase and power spectrograms derived through the S-transform [19]. Subsequently, this model highlighted the significance of phase synchronization as a discriminative feature during seizures. However, channel variability and seizure class imbalance affected performance robustness.

Hybrid architectures that integrated Deep Neural Networks (DNNs) and optimization algorithms have demonstrated their potential in seizure detection. Specifically, one such model combined a deep Residual Network (ResNet), which was optimized using the Walrus Optimizer for hyperparameter tuning, and then incorporated Variational Mode Extraction (VME) with regression to remove ocular artifacts [20]. However, the method's efficiency relied heavily on meta-optimization tuning that led to additional computational overhead. Additionally, to address the demand for robust decision-making, a study utilized an Evidential Multi-view Learning (EML) framework that explicitly modeled prediction confidence in EEG-based seizure detection [21]. Subsequently, by utilizing specific DNNs and Evidential Neural Networks, the model generated reliability scores for each classification decision. Further, upon validation using the CHB-MIT dataset, the study demonstrated that the EML frame outperformed several baseline models, thereby providing uncertainty quantification. However, it remained a challenge to implement multi-view models, and this approach relied upon handcrafted feature views, which affected the deployment and real-time decision support.

From a neuroimaging research perspective, a hybrid framework was proposed to evaluate independent acquisition of EEG and resting-state fMRI (rs-fMRI) data for the prediction of an epileptogenic network, which was the combination of deep learning and edge computing [22]. Specifically, this system performed unsupervised feature extraction for the prediction of preictal states, which enabled the framework to be implemented into real-time deployment through edge devices. The suggested system was infrastructure-agnostic for both invasive and non-invasive modalities and provided additional emphasis to differentiate true preictal states through convolutional structures. But implementation remained limited by the asynchronous nature of data acquisition and the dependence on edge-computing infrastructures. Another rs-fMRI-focused study conducted a direct comparison between 7 T (Tesla) and 3 T field strengths for lateralizing Seizure-Onset Zones (SOZ) in temporal lobe epilepsy [23]. Specifically, the 7 T imaging illustrated the substantially improved hippocampal-DMN-DMN connectivity contrast and resulted in a superior AUC for 3 that facilitated the clinical value of ultra-high-field imaging. However, it was less

accessible, as 7 T scanners were not widely available, and motion artifacts during rs-fMRI limited reliability in some patients.

In addition, recent neuromorphic approaches have demonstrated that Spiking Neural Networks (SNNs) effectively model the spatio-temporal dynamics of brain activity when compared to conventional deep learning methods. In particular, the NeuCube EvoSpike architecture, which is a biologically inspired framework, was introduced [24]. This architecture incorporated Spike-Timing-Dependent Plasticity (STDP) to capture and visualize dynamic neural pathways from EEG and fMRI data. Subsequently, based on this approach, an MRI-structured SNN (MRI-SNN) was designed, which assisted in the spatial alignment of EEG and MRI signals [25]. Specifically, MRI-SNN enabled personalized modeling and enhanced interpretability through reservoir-based spike learning. Consequently, another approach was developed using a hybrid Bi-LSTM-Reservoir SNN-NeuCube model for differentiating epilepsy and migraine while identifying spatial biomarkers including T6, F7, and C4 [26]. Hence, these approaches emphasized SNNs as a biologically plausible and interpretable paradigm for multimodal brain data fusion, which were capable of capturing fine-grained temporal dependencies and spatial correlations, thereby enhancing transformer-based architectures for neurodiagnostic applications.

The existing research in epileptic seizure detection demonstrated various unresolved gaps. Models relied only on time-domain augmentation and lacked integration of intracranial and extracranial data, which limited generalization. Further, cross-dataset variability and model interpretability were also underexplored, which facilitated the minimization of clinical trust. Furthermore, sensitivity to hyperparameters and optimized EEG channel selection were typically ignored, which affected performance and deployment efficiency. Additionally, a few systems incorporated predictive neurostimulation or patient-specific clinical records, which obstructed personalized care. Hence, in rs-fMRI-based approaches, the lack of a large, multi-institutional 7 T dataset constrained the broader validation of seizure-onset zone localization methods.

Model selection. To address the limitations associated with multimodal integration, dataset variability, and clinical trustworthiness in seizure detection, NeuroFusionNet is proposed, which is a modality-agnostic, graph-based transformer framework. Unlike traditional CNN or view-specific fusion models, NeuroFusionNet dynamically constructs modality-specific graphs from EEG and fMRI inputs, enabling flexible, high-order spatiotemporal representation learning. A cross-modal graph transformer captures shared dynamics between electrophysiological and hemodynamic signals, while an evidential output head quantifies prediction uncertainty that assists in enhancing clinical reliability. Subsequently, multiresolution encoding and contrastive alignment are incorporated, which ensured robustness across seizure types and the dataset. Hence, this model is selected for its scalability, interpretability, and superior adaptability across real-world neurophysiological inputs.

3. Methodology

The proposed NeuroFusionNet is a strong, modality-independent framework designed for seizure detection using EEG (CHB-MIT) rs-fMRI and 7T fMRI data, as demonstrated in Fig. 1. Particularly, this framework includes a preprocessing module, which is used to extract relevant temporal and spatial features. Subsequently, the MultiResolution Temporal Encoder (MRTE) is employed to capture the dynamics of the seizure at various time scales and generate a dynamic graph across modalities. Further, the Cross-Modal Graph Transformer acquires the representation across modalities to learn spatial and functional dependencies in both EEG and fMRI using the self and cross-attention layers. Furthermore, the contrastive learning operates on heterogeneous embeddings, and the evidential output head provides unique seizure predictions with their uncertainty. Finally, the dropout of modalities helps maintain robustness with missing input during training.

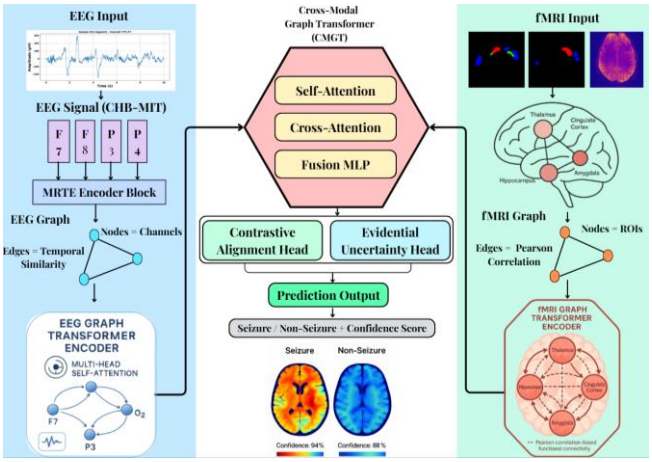

Fig. 1. Architecture diagram of proposed NeuroFusionNet

3.1. Data acquisition

To optimize modality diversity and structural robustness, the proposed NeuroFusionNet approach is implemented with data from three complementary and publicly available neurophysiological databases, as summarized in Table 1, such as the CHB-MIT scalp EEG dataset, the UW-Madison rs-fMRI epilepsy dataset, and the 7 T ultra-high-field fMRI database from the Human Connectome Project (HCP).

Table 1. Dataset description of EEG (CHB-MIT) rs-fMRI and 7 T fMRI

Dataset	Modality	Resolution	Patients	Key features	Label type
CHB-MIT	EEG	256 Hz (scalp)	24	Multichannel EEG, Seizure onsets, pediatric-specific patterns	Seizure intervals
UW-Madison rs-fMRI	fMRI	3 T, 2 s TR	>40	Connectivity matrices, resting-state BOLD, epilepsy metadata	SOZ + clinical labels
HCP 7T fMRI	fMRI	1.6 mm^3	>100	Subcortical ROIs, ultrahigh resolution, MNI space alignment	Healthy, no seizures

3.1.1. CHB-MIT

The CHB-MIT Scalp EEG dataset is an appropriate source of high-frequency electrophysiological signals, as illustrated in Fig. 2. Particularly, the dataset includes 24 pediatric subjects with clinically intractable epilepsy and multiple continuous EEG monitoring sessions lasting hours. The EEG recordings are in accordance with the international 10-20 electrode placement system, with a sampling frequency of 256 Hz, and a total of 23 scalp channels. Seizure events have been precisely annotated with onset and termination time stamps. Thus, the data is acquired through PhysioNet and is publicly available on Kaggle.

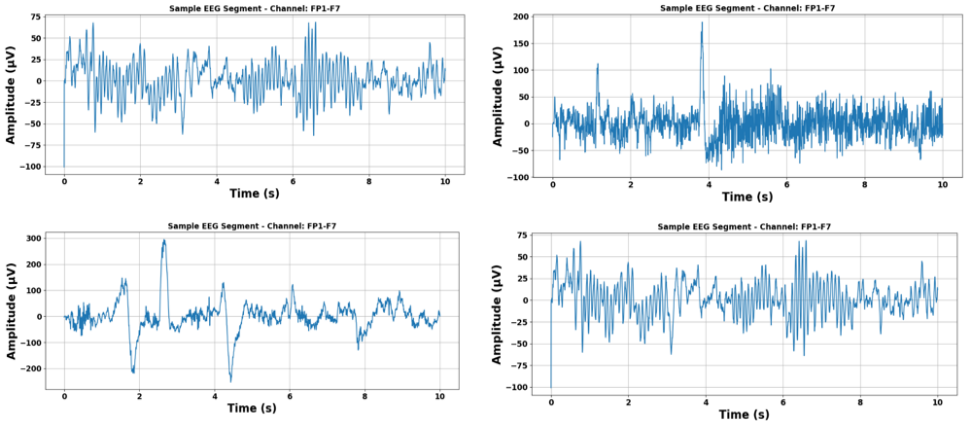

Fig. 2. Sample images of CHB-MIT Dataset across various channels

3.1.2. UW-Madison

The UW-Madison rs-fMRI epilepsy dataset adds the EEG data with non-invasive measures of seizure-related functional connectivity patterns, as shown in Fig. 3. The dataset contains resting-state BOLD fMRI scans from patients diagnosed with focal and generalized epilepsy using a 3 T MRI scanner. Each scan is associated with clinical site metadata (e.g., seizure onset zone, classification of epilepsy condition, anatomical ROIs, etc.). The BOLD time series were pre-processed using fMRIPrep, spatially normalized to the MNI template for imaging, and used for ROI-level Pearson correlations to convert the fMRI data into a functional connectivity matrix. The importance of this dataset is based on unquantified disruptions in functional networks on a large scale that are often used to model seizure propagation and seizure localization.

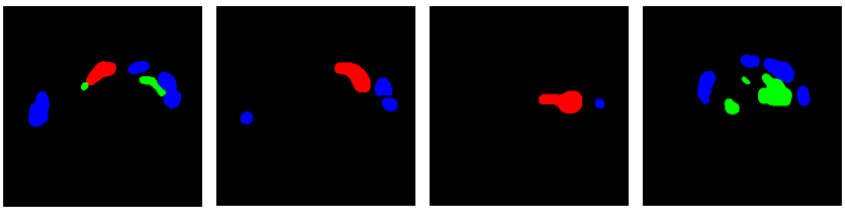

Fig. 3. Sample Images of UW-Madison data across various cases

3.1.3. 7 T fMRI data

To increase the resolution of the neurovascular events that are modeled, the model also uses the 7 T fMRI data from the HCP. The HCP dataset provides ultra-high-field (7 T) BOLD fMRI data as shown in Fig. 4, with an isotropic spatial resolution of 1.6 mm^3, which allows fine-grained anatomical mapping of both cortical and subcortical structures. The HCP project data set is predominantly healthy subjects, but its fine-grained anatomical coverage is utilized by the model to increase the structural generalization of the model, as well as help train the graph construction module with subcortical ROIs, which are relevant to epilepsy (i.e., hippocampus, thalamus, amygdala). The HCP data set is available through ConnectomeDB and contains rich demographic and neurocognitive metadata to facilitate some domain adaptation across clinical and non-clinical dataset sources.

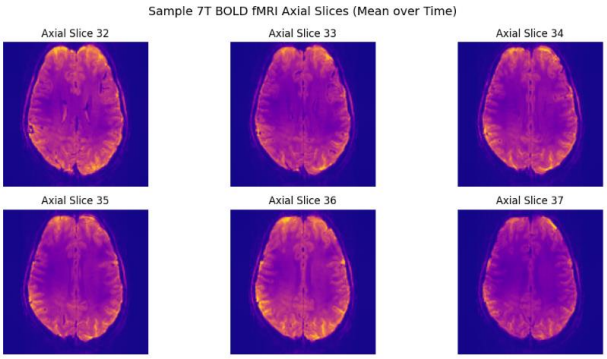

Fig. 4. Sample images of 7T BOLD fMRI axis slices

Hence, from the integration of the three datasets, this framework acquires temporal precision from EEG data, spatial richness from rs-fMRI connectivity, and ultra-high-resolution 7 T fMRI anatomy.

3.2. Pre-processing

To ensure robust multi-modal feature integration and reduce the effect of non-neural artifacts, each modality undergoes modality-specific further preprocessing after data acquisition.

3.2.1. EEG Signal Pre-processing

For the EEG signals collected from the CHB-MIT dataset, the preprocessing involves various techniques such as the elimination of artifacts, filtering based on the relevant band, and time–frequency decomposition to obtain discriminative features for seizure detection. Further, the raw EEG signal $x(t)$ from each channel is filtered using a bandpass filter from 0.5 - 40 Hz, with a target of the relevant frequency bands ranging from delta to beta, as this range includes information related to the epileptiform activity. The localized frequency content is obtained using the Short-Time Fourier Transform (STFT) as expressed in the equation

(1) $$X(t,f) = \sum_{\tau=-\infty}^{\infty} x(\tau) \cdot w(\tau - t) \cdot e^{-j2\pi f\tau},$$

where, and $X(t, f)$ signifies the complex-valued spectrogram, the input EEE signal is denoted as $x(\tau)$, the Gaussian window centred at time t is defined as $w(\tau - t)$, and f signifies the frequency component. The resulting clean EEG segments are given as an input to the Multiresolution Temporal Encoder (MRTE), a lightweight attention-based temporal module designed to extract multi-scale, frequency-aware embeddings. Further, this representation is normalized and reduced into a magnitude spectrogram $| X(t, f) |$, which is an input into the MRTE module. In addition, several entropy-based features are extracted per time window, such as spectral entropy H_{s} and permutation entropy $H_{\mathrm{p,}}$ which assist in capturing the complexity of the neural activity as demonstrated in the equation

(2) $$H_{\mathrm{s}}(t) = -\sum_{i=1}^{N} P_i(t) \cdot \log P_i(t).$$

Here, the normalized power of the frequency bin i at time t and N signifies the total number of bins.

3.2.2. rs-fMRI and fMRI signal processing

To maintain the integrity of hemodynamic signals and reduce sources of non-neural variability across subjects and sites, all fMRI and rs-fMRI volumes are preprocessed using a fixed preprocessing pipeline. For the UW-Madison rs-fMRI data set and the HCP 7 T fMRI dataset, a hybrid of fMRIPrep with the HCP minimal preprocessing workflows is adopted, as it follows many of the guidelines defined by the neuroimaging community. Hence, the preprocessing steps include the following:

The first step involves rigid-body realignment to adjust intra-session head motion. Particularly, each 3D volume is aligned to a reference 3D volume, usually the middle volume, based on six motion parameters that include 3 translations and 3 rotations, which are estimated using least-squares fitting. Further, this aligns the 2D slices that comprise each 3D volume in time to compensate for temporal acquisition differences across slices due to different slice orders, followed by interpolation to estimate the voxel-wise time series relative to a reference slice. Subsequently, structural images are co-registered to the MNI152 standard template using linear and non-linear transformations. Next, the functional images are warped into the same standardized MNI152 anatomical space. Thus, this allows for consistent parcellation across subjects and enables comparison across subjects. Consequently, a 3D isotropic Gaussian kernel with a 6 mm Full-Width at Half-Maximum (FWHM) is applied to the functional volumes. In particular, the spatial smoothing enhances the Signal-to-Noise Ratio (SNR) and compensates for inter-subject anatomical variability, especially in higher-resolution datasets such as HCP 7T. In addition, the voxel-wise time series are temporally bandpass filtered in the 0.01 - 0.1 Hz range to capture and remove slow-varying and neuronally driven fluctuations. Then, bandpass filtering is applied voxel-wise to remove low-frequency scanner drift and high-frequency physiological noise. Thus, the bandpass filtered time series helps acquire an interpretable BOLD signal.

Further, functional time series are extracted for each Region of Interest (ROI) using a standard brain atlas (e.g., Harvard-Oxford or AAL) after preprocessing is completed. Hence, the mean time series for ROI r, which is denoted as $s_r(t)$, is

computed by averaging BOLD intensities across all the voxels V_r defined by that ROI using the equation,

$$s_r(t) = \frac{1}{|V_r|}\sum_{(x,y,z)\in V_r} B(x,y,z,t), \quad (3)$$

where the preprocessed BOLD signal at spatial location (x, y, z) is defined as $B(x, y, z, t)$ and time t, and $|V_r|$ signifies the number of voxels in ROI r.

3.2.3. Extracted feature representation

As a final step of the pre-processing pipeline, the clean EEG and fMRI time series are passed through a unified MRTE model to generate rich and temporally embedded representations. This model uses dilated convolutions and attention blocks to preserve local and long-range signal characteristics across modalities. After the preprocessing, a set of discriminative features is extracted from each modality to establish the node-level input representations for the graph construction and transformer modeling.

EEG Features from CHB-MIT. For each EEG segment corresponding to a particular channel c, derived features include time–frequency decomposition, entropy measures, and statistics as discussed in Section 3.2.1. In application, each segment is initially transformed to a spectrogram through an STFT. Thus, features are then derived from energy in the standard clinical bands of delta, theta, alpha, and beta.

The resultant information-theoretic measures, such as spectral entropy, permutation entropy, and statistical features like mean, variance, skewness, and kurtosis, are calculated for each window. Feature vectors are concatenated and pooled across time steps to reduce the dimension of the analysis into a fixed-dimensional embedding vector $h_c^{\mathrm{EEG}} \in \mathbb{R}^d$ for each EEG channel c, with d denoting an equal feature dimension across modalities.

As discussed in Section 3.2.2, for each (ROI) r, the BOLD time series is processed to obtain several descriptive statistics and connectivity features, including the mean signal intensity, temporal variability, and first-order derivatives of BOLD. Once the full time series is computed for each ROI, it is fed through the MRTE encoder, which captures dynamics across short and longer temporal scales. This produces a compact multiscale embedding $h_r^{\mathrm{fMRI}} \in \mathbb{R}^d$, where $r \in \{1, 2, \ldots, R\}$ represents indices for ROIs. Finally, the extracted features from each modality are preserved as modality-specific node feature matrices, denoted as $H^{\mathrm{EEG}} \in \mathbb{R}^{N_e \times d}$ for EEG channels and $H^{\mathrm{fMRI}} \in \mathbb{R}^{N_f \times d}$ for fMRI/rs-fMRI ROIs. Here, the number of EEG electrodes and fMRI brain regions is represented as N_e and N_f, respectively, and d denotes the shared feature dimension after temporal encoding. Thus, these matrices are embedded as input for distinct graph construction pipelines.

3.3. Graph construction

To facilitate structured learning from heterogeneous neurophysiological data, the extracted feature representation from EEG and fMRI modalities is transformed into graph-structured data. Then, each data segment is encoded as a graph $\mathcal{G}^{(s)} = (\mathcal{V}, \mathcal{E}, A^{(s)}, H^{(s)})$ where s denotes the segment index.

3.3.1. Graph components

Nodes $\mathcal{V}$. Each node represents an EEG channel (for CHB-MIT) or an fMRI ROI. The node feature matrix $H^{(s)} \in \mathbb{R}^{N\times d}$ includes the per-node embeddings $h_i^{(s)} \in \mathbb{R}^{N\times d}$ created from MRTE or statistical descriptors.

Edges $\mathcal{E}$. Edges connect to reflect the strong interaction, or similarity, between nodes. The weighted adjacency matrix $A^{(s)} \in \mathbb{R}^{N\times N}$ defines the weight of the edges for each graph at a segment s.

3.3.2. EEG graph construction

The graphs constructed for the EEG are based on pairwise similarity of time-series features between EEG channels. Thus, the node embeddings and for channels i and j are defined as $h_i^{(s)}$ and $h_j^{(s)}$, respectively, the edge weight $a_{ij}^{(s)}$ is computed using a Gaussian kernel as demonstrated in the equation

$$a_{ij}^{(s)} = \exp\left(\frac{\left\|h_i^{(s)} - h_j^{(s)}\right\|_2^2}{\sigma^2}\right), \tag{4}$$

where the bandwidth hyperparameter used for controlling sensitivity to feature distance is denoted as $\sigma \in \mathbb{R}^+$ and $\|\cdot\|$ signifies the Euclidean norm. Hence, this means that if two EEG channels record similar activity, which the model links more strongly, that imitates how brain areas communicate during seizures.

3.3.3. fMRI graph construction

The graph topology for fMRI and rs-fMRI modalities is defined using FC between ROIs, which are derived from their BOLD time series $s_i(t)$. Subsequently, the edge weight between each pair (i and j) of ROI time series $s_i(t)$ and $s_j(t)$ is computed based on the Pearson Correlation Coefficient (ρ) using the equation

$$a_{ij}^{(s)} = \begin{cases} \rho_{ij} & \text{if } |\rho_{ij}| > \theta, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where, $\rho_{ij} = \frac{\sum_t (s_i(t)-\bar{s}_i)(s_j(t)-\bar{s}_j)}{\sqrt{\sum_t (s_i(t)-\bar{s}_i)^2}\sqrt{\sum_t (s_j(t)-\bar{s}_j)^2}}$.

Here, the temporal means of ROIs i and j are defined as $\bar{s}_i$ and $\bar{s}_{j,}$ respectively, and $\theta \in [0, 1]$ signifies the sparsity threshold.

3.3.4. Cross-modal graph alignment

To achieve multi-modal fusion in the neural modeling framework, separate graphs are constructed per modality while ensuring node correspondence by aligning either spatial locations or functional similarity, and the node embeddings $H^{\text{EEG}} \in \mathbb{R}^{N_e\times d}$ and $H^{\text{fMRI}} \in \mathbb{R}^{N_f\times d}$ are then mapped into a common latent space before fusion.

3.3.5. Dynamic graph adaptation

Instead of constructing a fixed global graph to represent topological brain dynamics across subjects, NeuroFusionNet constructs dynamic graphs per segment to capture the seizure-related transformations within the brain dynamics. For each segment s, a

graph $\mathcal{G}^{(s)}$ is constructed independently, temporally specific to the segment. The adjacency relationship on the dynamic graphs is evaluated through a similarity function $\phi(\cdot)$ of the node embeddings as expressed in the equation

$$A_{ij}^{(s)} = \phi(h_i^{(s)}, h_j^{(s)}) = \cos\left(h_i^{(s)}, h_j^{(s)}\right) = \frac{h_i^{(s)} \cdot h_j^{(s)}}{\left\|h_i^{(s)}\right\| \left\|h_j^{(s)}\right\|}. \quad (6)$$

Hence, this allows the model to capture context-specific seizure propagation patterns, which are crucial for generalization across recording sessions and patients. Specifically, encoding EEG and fMRI data into graphs specific to each modality allows the model to maintain each signal's intrinsic spatial/functional structure. Despite the model's representation of spatial graphs based on EEG and fMRI modalities, the temporal evolution of neural activity is represented through the MRTE modeling structure, regularized continuous dynamics of graph embedding, and continuous updating of the graph topology across time windows. Thus, this enables the proposed NeuroFusionNet model to capture time-dependent variations in neural activity. Specifically, this operation is similar to the temporal spike encoding mechanism in SNNs, thereby assisting in effectively tracing how seizures evolve and propagate across brain regions over time.

3.4. Cross-modal graph transformer

After constructing separate EEG and fMRI graphs per segment, a Cross-Modal Graph Transformer (CMGT) is incorporated to learn inter- and intra-modal interactions and integrate corresponding information from both modalities. The CMGT model is designed to learn from modality-specific brain graphs jointly by modeling the intra- and inter-modality dependencies. Further, each modality, EEG and fMRI, is initially encoded with a two-layer graph transformer with four head multi-head self-attention and a hidden dimension of 128, which assists in contextualizing nodes within each graph. Furthermore, a lightweight cross-attention module, which aligned complementary features between modalities in the form of bidirectional attention, enabling cross-modal interactive processing. Thus, the resulting derived representations are concatenated and passed through a fusion MLP that returns the effective shared segment-level embedding.

3.4.1. Dual branch graph encoding

Each modality-specific graph $\mathcal{G}_{\text{EEG}}^{(s)} = (\mathcal{V},\ \mathcal{E},\ A^{(s)},\ H^{(s)})$ and $\mathcal{G}_{\text{fMRI}}^{(s)} = (\mathcal{V},\ \mathcal{E},\ A^{(s)},\ H^{(s)})$ is independently processed by a graph transformer encoder that results in enhanced node representations using following equations:

$$Z^{\text{EEG}} = \text{GTransformer}\left(H^{\text{EEG}},\ A^{\text{EEG}}\right), \quad (7)$$

$$Z^{\text{fMRI}} = \text{GTransformer}(H^{\text{fMRI}},\ A^{\text{fMRI}}). \quad (8)$$

Further, each encoder applies stacked multi-head self-attention layers over the nodes, where attention weights are computed using

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}(a^T[Wh_i \| Wh_j])\right)}{\sum_{k \in \mathcal{N}(i)} \exp(\text{LeakyReLU}(a^T[Wh_i \| Wh_k]))}, \quad (9)$$

where the learnable projection matrix is defined as $W \in \mathbb{R}^{d \times d}$, $a \in \mathbb{R}^{2d}$ signifies the attention score vector, $\|$ symbolises vector concatenation, and $\mathcal{N}(i)$ represents the

neighborhood of node i. Hence, this means each brain node learns from its own neighborhood as well as weighs signals differently based on attention that signifies the most seizure-relevant nodes.

3.4.2. Cross-modal attention fusion

Further, to connect the data between two graphs, which allows individual nodes in a specific modality to attend to nodes in the other modality. Specifically, modality-aware cross-attention is computed through the following equations:

(10) $$Z^{\mathrm{EEG}\leftarrow\mathrm{fMRI}} = \mathrm{Attention}(Q^{\mathrm{EEG}},\ K^{\mathrm{fMRI}},\ V^{\mathrm{fMRI}}),$$

(11) $$Z^{\mathrm{fMRI}\leftarrow\mathrm{EEG}} = \mathrm{Attention}(Q^{\mathrm{fMRI}},\ K^{\mathrm{EEG}},\ V^{\mathrm{EEG}}).$$

Here, the Query, Key, and Value projections are denoted as Q, K, and V, respectively, and further each projection is assessed as $Q = ZW^{Q}$, $K = ZW^{K}$, and $V = ZW^{V}$ with learnable parameters W^{Q}, W^{K}, $W^{V} \in \mathbb{R}^{d\times d}$. Hence, the outputs of both self and cross model attention branches are concatenated and passed through a fusion head as expressed in the equation,

(12) $$Z^{\mathrm{fused}} = \mathrm{Fusion}\left(Z^{\mathrm{EEG}},\ Z^{\mathrm{fMRI}},\ Z^{\mathrm{EEG}\leftarrow\mathrm{fMRI}},\ Z^{\mathrm{fMRI}\leftarrow\mathrm{EEG}}\right).$$

Furthermore, to derive a unified segment-level embedding, modality-specific global average pooling (GAP) is applied across the graph nodes of the EEG and fMRI branches, respectively. Then, the pooled representations are concatenated and projected using a shared linear fusion layer that assists in enabling the integration of complementary temporal and spatial features into a compact representation, as represented in the equation

(13) $$Z^{\mathrm{segment}} = \mathrm{MLP}\left(\mathrm{GAP}\left(Z^{\mathrm{fused}}\right)\right).$$

Hence, this embedding is further passed to the downstream classification and contrastive heads. Thus, cross-modal attention fuses EEG and fMRI so that spatial and temporal features align.

3.5. Contrastive embedding alignment and uncertainty-aware classification

To ensure that the fused embeddings from EEG and fMRI modalities capture shared patterns that are related to seizure during the event, while recollecting unique modality-specific features. Therefore, a modality-wise contrastive learning objective is defined and contextualizes this learning in an evidential deep learning classifier that assists in modeling predictive uncertainty, which is essential for clinical applications.

3.5.1. Modality-wise contrastive learning

While EEG and fMRI learn from different graphs, they both encode semantically aligned neural phenomena at the time of seizures. However, due to variability in acquisition, signal scale, and physiological characteristics, direct feature fusion leads to sub-optimal alignment. Therefore, to resolve this issue, a modality-wise contrastive loss is employed to mitigate representational disparity between EEG and fMRI feature spaces that encourages temporally aligned segments sourced from the same subject and time window, which assists in producing convergent embeddings

while simultaneously applying separation from embeddings of unaligned or cross-subject pairs:

- $Z^{\mathrm{EEG}} \in \mathbb{R}^d$: Pooled embedding from the EEG graph;
- $Z^{\mathrm{fMRI}} \in \mathbb{R}^d$: Pooled embedding from the fMRI graph.

Thus, a symmetric contrastive loss is formulated using InfoNCE as expressed in the equation

$$\mathcal{L}_{\mathrm{con}} = -\log \frac{\exp\left(\frac{\mathrm{sim}\left(Z^{\mathrm{EEG}}, Z^{\mathrm{fMRI}}\right)}{\tau}\right)}{\sum_{k=1}^{K}(\mathrm{sim}(Z^{\mathrm{EEG}},\ Z_k^{\mathrm{fMRI}}/\tau)}. \tag{14}$$

Here, sim(·) signifies the cosine similarity $\mathrm{sim}(a, b)$. The Temperature hyperparameter is defined as τ, and Z_k^{fMRI} signifies the negative sample drawn from other segments. Hence, this loss is applied bidirectionally to maintain cross-modal consistency, which assists the model to learn latent seizure signatures that are common in both the EEG and fMRI domains. Henceforth, contrastive learning ensures that EEG and fMRI views of the same seizure look similar in the learned space, while different patients or time windows remain distinct.

3.5.2. Evidential uncertainty modeling

While diagnosing epilepsy and monitoring seizures, it is essential to not only accurately classify neural segments but also quantify the model's confidence in the given prediction. For seizure detection, the problem is framed as a supervised classification task in which features that are collected from temporally segmented EEG and fMRI data have been aligned and then classified to be in one of two classes: seizure (ictal) and non-seizure (interictal or preictal). The decision is made based on patterns learned from modality-specific graphs and their fused representation Z^{segment}, which includes both spatial (fMRI), temporal (EEG), and related dynamics from the CMGT model.

To make this prediction interpretable and clinically viable, an Evidential Deep Learning (EDL) classifier is incorporated, which predicts a Dirichlet distribution over classes instead of deterministic softmax outputs. For instance, EDL operates by only passing the fused embedding $Z^{\mathrm{segment}} \in \mathbb{R}^d$ through a fully connected evidential output head which provides an estimation of the non-negative evidence vector $e = [e_1, e_2, \ldots, e_C] \in \mathbb{R}_+^C$ where C, $C = 2$, in binary classification. Thus, this evidence is then utilized to parameterize the Dirichlet distribution as expressed in the equation

$$\alpha = e + 1. \tag{15}$$

From that equation, the expected class probabilities are evaluated through the equation:

$$\hat{p}_i = \frac{\alpha_i}{\sum_{j=1}^{C} \alpha_j}. \tag{16}$$

Further, the associated uncertainty mass is measured using the equation

$$u = \frac{C}{\sum_{j=1}^{C} \alpha_j}, \tag{17}$$

where, a higher value of $u \in [0, 1]$ denotes the lower model confidence that resembles the uncertainty in the decision due to ambiguous or insufficient evidence. Unlike a standard classifier, which just outputs Yes or No, the evidential modeling

provides both a prediction and a confidence score, thereby ensuring the system avoids overconfident errors.

Training of this module is accomplished with an evidential loss $\mathcal{L}_{\text{EDL}}$ that penalizes overconfident errors and rewards calibrated predictions. Specifically, this loss is comprised of the expected negative log-likelihood of the true class label taken under the Dirichlet Posterior together with a Kullback-Leibler divergence that regularizes the predicted Dirichlet distribution toward a uniform prior as demonstrated in the equation

(18) $$\mathcal{L}_{\text{EDL}} = \mathbb{E}_{\text{Dir}(\alpha)}[-\log p(y)] + \lambda \cdot \text{KL}[\text{Dir}(\alpha) \| \text{Dir}(1)].$$

Hence, this framework ensures that the seizure predictions are made fully only when the model is sufficiently confident and the fused embeddings from EEG and fMRI convey consistent and convergent evidence of ictal activity. The model is trained in a joint optimization framework where both the modality-wise contrastive loss and the evidential classification loss are integrated. Specifically, the total training objective is defined through the equation

(19) $$L_{\text{total}} = \lambda_1 L_{\text{contrastive}} + \lambda_2 L_{\text{evidential}}.$$

Here, λ_1 and λ_2 balance the contribution of alignment and classification. This ensures that the learned embeddings remain semantically aligned across modalities, simultaneously being optimized for reliable and uncertainty-aware seizure prediction. Thus, both losses operate jointly in a single end-to-end training process rather than being used separately.

The final output of this model includes both:

- $\hat{p}_{\text{seizure}} \in [0, 1]$ signifies the predicted probability of seizure,
- $u \in [0, 1]$ defines the associated uncertainty score.

Thus, these two outputs jointly guide downstream action in a clinical setting, which enables decisions to be thresholded or flagged based on both seizure likelihood and prediction confidence. Moreover, contrastive learning and uncertainty modeling together ensure high accuracy and clinical trustworthiness as it learns robust shared features and provides a signal when it is uncertain.

4. Experimental results

All experiments are conducted on a system with an NVIDIA RTX 3090 GPU (24 GB VRAM), Intel Core i9-13900K CPU, and 128 GB RAM. Further, modality-specific data were used for the management of data through MNE-Python EEG and NiLearn for fMRI. The model is trained using the Adam optimizer with a learning rate of 1×10^{-4}, a cosine annealing schedule, and early stopping criteria on validation loss. Thus, all experiments are repeated over five random seed splits, and average metrics are reported. The proposed NeuroFusionNet is evaluated in terms of accuracy, precision, and recall as formulated in the following equations:

(20) $$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}},$$

(21) $$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}},$$

(22) $$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}.$$

where TP denotes True Positives, FP represents False Positives, FN signifies False Negatives, TN denotes True Negatives.

4.1. Performance analysis

To examine the robustness of the proposed NeuroFusionNet, performance against baseline classifiers typically used in recent epilepsy seizure detection is conducted. The baseline classifiers include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Gated Recurrent Units (GRUs), GCN, GAT, and DGCNN, which are designed to exploit spatial or temporal signatures essential in EEG and fMRI signals.

Table 2. Performance analysis of benchmark models with NeuroFusionNet

Model	Accuracy (%)	Precision (%)	Recall (%)
CNN	96.42	95.83	95.97
RNN	94.75	94.28	92.54
GRU	95.66	95.12	94.80
GCN	97.88	98.01	96.74
GAT	98.25	98.42	97.83
DGCNN	98.56	98.87	98.01
Proposed NeuroFusionNet	99.22	99.89	99.85

As shown in Table 2, conventional CNN and RNN models exhibited good performance, and Graph-based models demonstrated improved performance over the baseline models, particularly in the case of GAT and DGCNN, which can model relationships in non-Euclidean space, wherein, in the case of DGCNN, an accuracy of 98.56% was achieved. Thus, the proposed NeuroFusionNet effectively outperformed all other models with an accuracy of 99.22%, precision of 99.89% and recall of 99.85%.

4.2. Comparative analysis

To further assess the performance of NeuroFusionNet, the comparative analysis against multiple existing seizure detection models is performed as listed in Table 3. The models are comprised of deep learning models TriSeizureDualNet [17], ST-CNN [19], and Optimized ResNet-18 [20], as well as hybrid and evidential frameworks LSTM-SVM [22] and EML [21]. Thus, all comparative results are sourced from published benchmarks and re-evaluated under comparable experimental settings to ensure fair evaluation.

Table 3. Comparative Analysis of the proposed NeuroFusionNet with existing models

Model	Accuracy (%)	Precision (%)	Recall (%)
TriSeizureDualNet [17]	99.00	99.245	99.14
ST-CNN [19]	92.00	92.23	79.59
Optimized ResNet-18 [20]	97.64	97.90	98.01
EML [21]	99.05	99.64	97.96
LSTM-SVM [22]	98.00	97.00	96.00
Proposed NeuroFusionNet	99.22	99.89	99.85

The proposed NeuroFusionNet outperforms the existing models with an accuracy of 99.22%, precision of 99.89%, and recall of 99.85% and even outperforms

the recent evidential model EML [21]. Thus, the high precision and recall of the proposed NeuroFusionNet demonstrate high resistance to false positives and false negatives, which is important for clinical seizure detection.

4.3. Ablation study

To evaluate the effectiveness of each component within the NeuroFusionNet framework, a series of ablation experiments is conducted, particularly on the architectural components, including CMGT, Contrastive Learning (CL), and EOH. Moreover, these experiments assess how different architectural modules contribute to overall seizure detection performance, as demonstrated in Fig. 5.

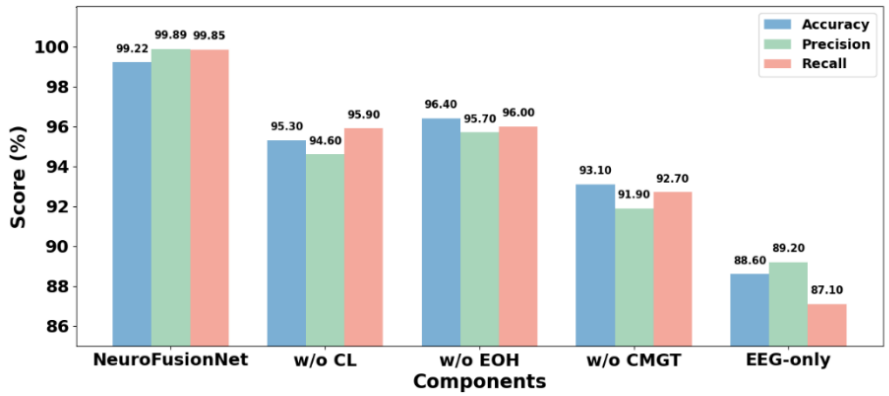

Fig. 5. Comparative performance analysis in terms of Accuracy, Precision, and Recall across ablation variants of NeuroFusionNet

As demonstrated in Fig. 5, the largest performance drop (−6.1% accuracy) occurs when the CMGT module is excluded from the model. Thus, it is inferred that these results empirically validate the architectural integrity of NeuroFusionNet, where interpretability and reliability are essential for robust seizure detection.

4.4. Statistical analysis

To validate the statistical significance of performance differences between NeuroFusionNet and other approaches, a paired one-tailed t-test is performed, where both models use the same data folds:

- Null Hypothesis H_0: There is no significant difference in model performance, H_0: $\mu_{\mathrm{NF}} \leq \mu_{\mathrm{GCN}}$.
- Alternative Hypothesis H_1: NeuroFusionNet has a significantly greater model performance than baseline, H_1: $\mu_{\mathrm{NF}} > \mu_{\mathrm{GCN}}$,

where μ_{NF} and μ_{GCN} are the mean accuracy scores from NeuroFusionNet and the GCN model, respectively. Further, they established the normality of the accuracy differences through the Shapiro-Wilk test before conducting the t-test. Hence, this determines that the distributional assumptions are satisfied. Finally, to complement the t-test results and address possible deviations from normality, a Wilcoxon signed-rank test is conducted as a non-parametric alternative to the paired t-test. The results of the statistical tests are illustrated in Table 4.

Table 4. Statistical analysis of the proposed NeuroFusionNet

Test	Statistic	p-value	Conclusion
Paired t-test (one-tailed)	$t = 37.19$	< 0.00001	Reject H_0, significant
Wilcoxon signed-rank	$W = 55$	< 0.00001	Reject H_0, significant

From Table 4, it is inferred that the mean accuracy of NeuroFusionNet is significantly higher (99.23% ± 0.06%) compared to the GCN (94.75% ± 0.29%).

4.5. Discussion

The proposed NeuroFusionNet introduces a substantial advancement over existing seizure detection models through the novel integration of modality-specific graph encoders, cross-modal attention, and evidential uncertainty estimation. When compared to the traditional CNN, RNN, or unimodal architectures, the proposed NeuroFusionNet effectively captures both the temporal dynamics of EEG and the spatial-functional patterns of fMRI using graph-based representations. Thus, the incorporation of a cross-modal graph transformer enables the model to learn rich interdependencies across modalities while modality-wise contrastive learning ensures the alignment of semantically similar patterns without applying structural integration. Furthermore, the evidential output head presents calibrated uncertainty scores, which enhance clinical interpretability and decision reliability. Hence, this combination of multi-modal fusion, graph-based modeling, and uncertainty-aware prediction determines NeuroFusionNet as a robust and clinically significant framework for seizure detection.

5. Conclusion

This research presents NeuroFusionNet, a multi-modal graph-based framework that addresses key limitations in current epileptic seizure detection systems. By fusing EEG and fMRI signals through modality-specific graph encoders and a cross-modal transformer, the proposed NeuroFusionNet model overcomes the constraints of time-domain learning and supports integration of both intracranial and extracranial modalities. Furthermore, modality-wise contrastive learning minimizes cross-dataset variability, and the evidential uncertainty head enhances interpretability and fosters clinical trust. When compared to the existing models, NeuroFusionNet is robust to hyperparameter sensitivity and supports flexible modality input, thereby enhancing its generalization and deployment potential. The proposed NeuroFusionNet outperforms the existing models with an accuracy of 99.22%, precision of 99.89% and 99.85% of recall. Finally, the proposed model advances seizure detection by resulting in a reliable, generalizable, and clinically aligned solution. In the future, the proposed NeuroFusionNet will be utilized in real-time clinical monitoring systems that facilitate initial alerts during an epileptic event and assist neurologists in localizing the seizure-onset zones. Furthermore, the uncertainty-aware outputs are helpful in building hospital decision-support platforms and conducting neuroprosthetic research. Thus, develops the proposed NeuroFusionNet framework's applications beyond epilepsy to other neurological disorders associated with spatio-temporal abnormalities.

References

1. Rehab, N., Y. Siwar, Z. Mourad. Machine Learning for Epilepsy: A Comprehensive Exploration of Novel EEG and MRI Techniques for Seizure Diagnosis. – Journal of Medical and Biological Engineering, Vol. **44**, 2024, pp. 317-336.
2. Lin, P. T., J. H. Sie, H. J. Lee, C. C. Chou, Y. C. Shih, C. Chen, F. H. Lin, W. J. Kuo, H. M. Khoo, H. Y. Yu. Detection of Epileptogenic Zones in People with Epilepsy Using Optimized EEG-fMRI. – Epilepsy & Behavior, Vol. **164**, 2025, 110257.
3. Aravind Britto, K. R., A. B., S. Srinivasan, S. K. Mathivanan, M. Venkatesan, B. A. Malar, S. Mallik, H. Qin. A Multi-Dimensional Hybrid CNN-BiLSTM Framework for Epileptic Seizure Detection Using Electroencephalogram Signal Scrutiny. – Systems and Soft Computing, Vol. **5**, 2023, 200062.
4. Nandakumar, N., D. Hsu, R. Ahmed, A. Venkataraman. DeepEZ: A Graph Convolutional Network for Automated Epileptogenic Zone Localization from Resting-State fMRI Connectivity. – IEEE Transactions on Biomedical Engineering, Vol. **70**, 2022, No 1, pp. 216-227.
5. Werdiningsih, I., I. Puspitasari, R. Hendradi. Recognizing Daily Activities of Children with Autism Spectrum Disorder Using a Convolutional Neural Network Based on Image Enhancement. – Cybernetics and Information Technologies, Vol. **25**, 2025, No 1, pp. 78-96.
6. Karimi-Rouzbahani, H., S. Vogrin, M. Cao, C. Plummer, A. McGonigal. Multimodal and Quantitative Analysis of the Epileptogenic Zone Network in the Pre-Surgical Evaluation of Drug-Resistant Focal Epilepsy. – Neurophysiologie Clinique, Vol. **54**, 2024, No 6, 103021.
7. Berger, M., R. Licandro, K. H. Nenning, G. Langs, S. B. Bonelli. Artificial Intelligence Applied to Epilepsy Imaging: Current Status and Future Perspectives. – Revue Neurologique, Vol. **181**, 2025, No 5, pp. 420-424.
8. Sadiq, M., M. N. Kadhim, D. Al-Shammary, M. Milanova. Novel EEG Feature Selection Based on Hellinger Distance for Epileptic Seizure Detection. – Smart Health, Vol. **35**, 2025, 100536.
9. Ejaz, S., U. Noor, Z. Rashid. Visualizing Interesting Patterns in Cyber Threat Intelligence Using Machine Learning Techniques. – Cybernetics and Information Technologies, Vol. **22**, 2022, No 2, pp. 96-113.
10. Hussein, A. M., S. A. Alomari, M. H. Almomani, R. A. Zitar, K. Saleem, A. Smerat, S. Nusier, L. Abualigah. A Smart IoT-Cloud Framework with Adaptive Deep Learning for Real-Time Epileptic Seizure Detection. – Circuits, Systems, and Signal Processing, Vol. **44**, 2025, No 3, pp. 2113-2144.
11. Xu, T., Y. Wu, Y. Tang, W. Zhang, Z. Cui. Dynamic Functional Connectivity Neural Network for Epileptic Seizure Prediction Using Multi-Channel EEG Signal. – IEEE Signal Processing Letters, Vol. **31**, 2024, pp. 1499-1503.
12. Li, Z., B. Chen, N. Zhu, W. Li, T. Liu, L. Guo, J. Han, T. Zhang, Z. Yan. Epileptic Seizure Detection in SEEG Signals Using a Signal Embedding Temporal-Spatial-Spectral Transformer Model. – IEEE Transactions on Instrumentation and Measurement, Vol. **74**, 2025, 4001111.
13. Mokhiamar, M. O., A. Mahmoud, M. I. Eldagla, Y. Aribi, A. M. Anter. Blockchain-Integrated Multi-Modal LSTM-CNN Fusion for High-Precision Epileptic Seizure Detection from EEG Signals. – Knowledge-Based Systems, Vol. **323**, 2025, 113703.
14. Karthik, S. A., K. N. Bharath, B. R. Ramji, K. Puttegowda, B. Aruna, D. S. Kumar. Enhanced EEG Signal Processing for Accurate Epileptic Seizure Detection. – SN Computer Science, Vol. **6**, 2025, 608.
15. Cao, X., S. Zheng, J. Zhang, W. Chen, G. Du. A Hybrid CNN-Bi-LSTM Model with Feature Fusion for Accurate Epilepsy Seizure Detection. – BMC Medical Informatics and Decision Making, Vol. **25**, 2025, 6.

16. F u, R., B. Z h a n g, B. X u e, D. W a n g. A Spatiotemporal Posterior Graph Convolutional Neural Network Based on Multihead Attention with Squeeze-and-Excitation Module for Patient-Specific Epileptic Seizure Prediction. – IEEE Transactions on Instrumentation and Measurement, Vol. **74**, 2025, 2517713.
17. S u n k a r a, M., S. R. R e e j a. Tri-SeizureDualNet: A Novel Multimodal Brain Seizure Detection Using a Triple Stream Skipped Feature Extraction Module Entrenched Dual Parallel Attention Transformer. – Biomedical Signal Processing and Control, Vol. **88**, 2024, 105593.
18. X i e, C., S. Q i u, C. Z h o u, X. S o n g, J. Y a n g, J. H u a n g, W. W a n g, H. J i a o. Automatic Epileptic Seizure Detection with an Ultra-Lightweight 3D-CNN Model. – Biomedical Signal Processing and Control, Vol. **110**, 2025, 108260.
19. L i u, Y., G. L i u, S. W u, C. T i n. Phase Spectrogram of EEG from S-Transform Enhances Epileptic Seizure Detection. – Expert Systems with Applications, Vol. **262**, 2025, 125621.
20. S i l p a, B., M. K. H o t a. An Efficient Epileptic Seizure Detection Framework Based on Optimized Deep Residual Network. – Neural Computing and Applications, Vol. **37**, 2025, pp. 16849-16870.
21. L i u, Y., C. X u, Z. W e n, Y. D o n g. Trust EEG Epileptic Seizure Detection via Evidential Multi-View Learning. – Information Sciences, Vol. **694**, 2025, 121699.
22. H o s s e i n i, M. P., T. X. T r a n, D. P o m p i l i, K. E l i s e v i c h, H. S o l t a n i a n-Z a d e h. Multimodal Data Analysis of Epileptic EEG and rs-fMRI via Deep Learning and Edge Computing. – Artificial Intelligence in Medicine, Vol. **104**, 2020, 101813.
23. L u c a s, A., E. J. C o r n b l a t h, N. S i n h a, L. C a c i a g l i, P. H a d a r, A. T r a n q u i l l e, J. M. S t e i n, S. D a s, K. A. D a v i s. Seizure-Onset Zone Lateralization in Temporal Lobe Epilepsy Using 7 T rs-fMRI: Direct Comparison with 3 T rs-fMRI. – Epilepsia, Vol. **66**, 2022, No 9, pp. 34440-3452.
24. K a s a b o v, N. Neucube Evospike Architecture for Spatio-Temporal Modelling and Pattern Recognition of Brain Signals. – In: Proc. of IAPR Workshop on Artificial Neural Networks in Pattern Recognition Berlin, Heidelberg: Springer Berlin Heidelberg, September 2012, pp. 225-243.
25. S a e e d i n i a, S. A., M. R. J a h e d-M o t l a g h, A. T a f a k h o r i, N. K a s a b o v. Design of MRI Structured Spiking Neural Networks and Learning Algorithms for Personalized Modelling, Analysis, and Prediction of EEG Signals. – Scientific Reports, Vol. **11**, 2021, No 1, 12064.
26. S a e e d i n i a, S. A., M. R. J a h e d-M o t l a g h, A. T a f a k h o r i, N. K. K a s a b o v. Diagnostic Biomarker Discovery from Brain EEG Data Using LSTM, Reservoir-SNN, and NeuCube Methods in a Pilot Study Comparing Epilepsy and Migraine. – Scientific Reports, Vol. **14**, 2024, No 1, 10667.

Received: 25.09.2025, First revision: 05.11.2025, Accepted: 10.11.2025