Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 2, April 2026, pp: 447-472 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

A Multimodal Explainable-AI Approach for Deep-Learning-based Epileptic Seizure Detection

Ashwini Patil, and Megharani Patil

Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

Corresponding author: Ashwini Patil (e-mail: patil.ashwini.03@gmail.com), **Author(s) Email**: Megharani Patil (e-mail:megharani.patil@thakureducation.org)

Abstract Epilepsy carries a high risk of sudden death and increased premature mortality, highlighting the importance of automatic seizure detection to support faster diagnosis and treatment. The opacity of existing deep learning models limits their real-world application in diagnosing epileptic seizures, underscoring the need for more transparent and explainable systems. Limited research studies are available on Explainable Artificial Intelligence (XAI)-based epileptic seizure detection, and these studies provide only a visual explanation for the model's behaviour. Additionally, these studies lack validation of the XAI outputs using quantitative measures. Thus, this research aims to develop an explainable epileptic seizure detection model to address the limitations of existing black-box deep learning approaches. It proposes a novel Hybrid Transformer-DenseNet121-XAI (HTD-MXAI) integrated model for detecting epileptic seizures from EEG data. The proposed model leverages advanced deep learning architectures, namely the Transformer and DenseNet121, for automatic feature extraction, while simultaneously extracting handcrafted features from the time, frequency, and spatial domains. The XAI techniques, such as Attention Weights, Saliency Maps, and SHapley Additive eXplanations (SHAP), are integrated with the proposed model to provide multimodal explainability for the model's decision-making process. The results demonstrate that the proposed model outperforms state-of-the-art models for seizure detection. It achieves an overall (aggregated across subjects) accuracy of 99.14%, Sensitivity of 98.49%, and Specificity of 99.68% when applied to the CHB-MIT dataset. The Faithfulness score of 40.94% and completeness score of 1.00 indicate that the explanations provided by the XAI method for the model's prediction are highly reliable. In conclusion, the proposed model offers a promising solution to the constraints, including the interpretability of black box models, limited multimodal explainability, and the validation of XAI techniques in the context of epileptic seizure detection.

Keywords: Epileptic Seizure Detection; Multimodal Explainable AI; Transformer; DenseNet121; Transfer Learning; SHAP

I. Introduction

Epilepsy is a disorder of the nervous system that affects both children and adults, marked by the repeated and untriggered occurrence of seizures [1]. Epileptic seizures result from sudden disruptions in the Electroencephalogram (EEG) brain signals, affecting groups of brain cells and leading to involuntary movements, sensory disturbances, mood changes, cognitive impairment, and potential loss of consciousness, posing significant risks to individuals [2]. The risks associated with epileptic seizures, such as sudden unexpected death, injuries, unconsciousness, cognitive, and memory issues, are major concerns [3]. EEG is vital for capturing brainwave signals and recording brain activity. It plays a significant role in diagnosing epilepsy by providing helpful information regarding the changes in brain activity during seizures [4]. However, manually reviewing these long-hour recordings to identify seizures is time-consuming and prone to error. The development of an automated system has emerged as a potential solution to support clinicians in the early diagnosis of epilepsy [5]. Detecting epileptic seizures offers valuable insights for early diagnosis, influencing treatment decisions. Recent approaches leverage Machine Learning (ML) and Deep Learning (DL) models to address the limitations of existing EEG signal processing methods. Deep learning enhances the analysis of EEG signals, benefiting research areas such as epilepsy, movement disorders, memory, depression, schizophrenia, and sleep [6]. Owing to the vast volumes of data and the implementation of deep learning, opaque models have effectively addressed challenges in real-world, life-threatening scenarios [7].

Many researchers have applied Convolutional Neural Networks (CNNs) [8-10] and CNN-based pre-trained models with transfer learning [11-15] for detecting epileptic seizures. The transfer learning approach,

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

which utilizes pre-trained models, reduces the training time for complex deep learning models and enhances their performance. CNNs can efficiently extract spatial features but cannot model long-range dependencies in sequential EEG time-series data [16]. To mitigate the limitations of CNNs, many existing studies have employed Recurrent Neural Networks (RNNs) [17] and their variants, such as Long Short-Term Memory (LSTM) [18] and Gated Recurrent Unit (GRU) [19], for seizure detection. Although RNNs and their variants like LSTM and GRU are effective at modeling sequential input data, they struggle to capture long-range dependencies due to issues such as vanishing gradients and increased training time [20]. Advanced deep learning models, such as the Transformer, have been developed to address these limitations. Transformers leverage self-attention mechanisms to process sequential data in parallel, enabling them to learn long-range relationships efficiently. With these attention-based capabilities, the Transformer model excels in sequence-related tasks, demonstrating strong performance in analyzing EEG signals for Epileptic Seizure Detection [21-23]. Additionally, numerous researchers have employed hybrid models that combine the strengths of various architectures to enhance overall performance [2, 24-26].

Although many advanced deep learning models have been applied in ongoing research on epileptic seizure detection, their black-box nature of these models hinders their adoption in the real world [27]. To overcome this challenge, Explainable Artificial Intelligence (XAI) has emerged as a pioneering technology for interpreting the behavior of complex deep learning models. This enables clinicians to trust the model's decision-making process and promotes its adoption in critical domains, such as healthcare [28]. Few research studies have utilized XAI for epileptic seizure detection with deep learning models. These studies lack user-understandable explanations for the models' behavior and standard evaluation metrics for measuring the performance of XAI techniques. Recent research on explainable epileptic seizure detection spans a variety of deep learning approaches, yet key challenges persist. Deep learning models that utilize connectivity features, attention, CNNs, and Bi-LSTM architectures [29], as well asattention-based CNNs for spatial channel relevance [30] enhance performance but offer limited interpretability and incur high computational costs. Several studies integrate visualization-based XAI methods, such as gradient ascent and SHAP [31], SHAP-based hybrid CNN models [32], LRP with Bi-LSTM [33], Grad-CAM and attention visualization in CNNs and ViT models [34], and SHAP-supported tree ensembles [35-37], to highlight important features or channels. However, explanations often remain qualitative, dataset-specific, or weakly aligned with clinically meaningful ictal patterns. Siamese CNNs with SHAP and LIME [38], bagged-tree models with SHAP [39], and neonatal-focused CNN-Graph Attention models with modified Grad-CAM [40] further emphasize interpretability but are constrained by dataset size, generalizability, and real-time applicability. Additional work explores explainable GNNs [41], interpretable SVM-based pipelines using clinically relevant features and t-SNE visualization [42], transfer-learning models with LRP [43], and classical classifiers combined with LIME and SHAP [44]; however, most are limited by qualitative evaluation and restricted datasets. Overall, despite these advances, significant gaps remain in achieving deeper interpretability, comprehensive multimodal explanations, and robust, quantitative evaluation metrics for XAI techniques in seizure detection. A comparative analysis of several state-of-the-art Epileptic Seizure detection research based on Explainable AI is given in Table 1.

A. Motivation and Problem Formulation

From the existing study on XAI-based Epileptic Seizure Detection, it is determined that there are still several research gaps that exist in the early diagnosis of epileptic seizures.

1. Lack of Interpretability in Seizure Detection Models

Although many ML and DL methods have been used for seizure detection, their decision processes remain largely opaque, and only a limited number of studies have applied XAI in this domain. To address this gap, this study proposes an explainable model for epileptic seizure detection.

$$g(x) \approx f(x) \qquad (1)$$

$$\underset{g \in G}{\operatorname{argmin}}\, L\left(f(x), g(x)\right) + \Omega(g) \qquad (2)$$

Eq. (1) [45] formulates an explainable surrogate model $g(x)$ that mimics $f(x)$ faithfully, where x represents input EEG features. As described in Eq. (2) [45], the aim is to minimize the difference between the $f(x)$, which is a complex, deep-learning model and $g(x)$ which is a simple, explainable model. The loss function L measures how different $g(x)$ from $f(x)$ and how faithful the explanation is to the original model. G represents a class of simple interpretable models and $\Omega(g)$ represents a complexity penalty that tries to keep g simple and interpretable.

2. Limited Multimodal Explainability in EEG-based AI Systems

Understanding the epileptic behaviors and state transitions from the visual explanation alone affects the early diagnosis of Epilepsy. Hence, providing a multimodal explanation to doctors and patients with the hybrid feature extraction is essential to improve the model's reliability and efficiency.

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

Table 1. Comparative Analysis of Explainable-AI-based Studies on Epileptic Seizure Detection

Year	Approach	Dataset	XAI Technique		Performance Metrics for XAI	
			Visualization	Textual Explanation	Qualitative	Quantitative
2020 [29]	CNN, Bi-LSTM	CHB-MIT	Feature Relevance	×	Correlation between Feature relevance and scientific understanding	×
2020 [30]	CNN with Attention	TUH	Attention topography	×	Correlation between attention weights and EEG channels	×
2021 [31]	CNN	REPO 2 MSE cohort,	Gradient Ascent and SHAP	×	Frequency components and spatial-temporal distribution	×
2022 [32]	1D-CNN, 3D-CNN	Helsinki University Hospital	SHAP	×	User Feedback, Explanation Patterns	User Study, t-test, Likert Scale Ratings
2022 [33]	Bi-LSTM	Bonn	LRP	×	Visual Inspection	×
2023 [34]	ResNet18, LeNet-5, VGG-11, Vision Transformer	iEEG data (Juntendo University Hospital)	Grad-CAM, Attention Layer	×	Visual Inspection	×
2023 [35]	Ensemble-based CatBoost	SeizIt1 and SeizIt2	SHAP	×	Feature Importance	×
2023 [36]	DT, kNN, LR, NB, RF, XGBoost, and SVM	University of Beirut Medical Centre	SHAP	×	Feature Importance	×
2024 [37]	LSTM	CHB-MIT	SHAP	×	Optimal Channel Combination Determination	×
2024 [38]	Siamese CNN (Wang_1d)	CHB-MIT, Siena, TUSZ	SHAP, LIME	×	×	Spearman's rank correlation coefficient
2024 [39]	Bagged Tree-based classifier (BTBC)	Bonn	SHAP	×	Feature Importance	×
2024 [40]	Multilayer Perceptron (MLP)	Helsinki University Hospital	Grad-CAM	×	Inspecting the relevance of each channel and time window	×
2024 [41]	Attention-based Graph Neural Network (GNN)	CHB-MIT	Attention Mechanism, GNN Explainer	×	Interpretation of EI and FI Scores	Edge Importance, Feature Importance Score

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeemi.v8i2.1380

2024 [42]	SVM with Gaussian Kernel	New Delhi, Bonn	t-SNE	×	Interpretable features	Coefficient of Variation, Statistical Significance Testing
2024 [43]	VGG16	Bonn	LRP	×	Interpretable features	×
2024 [44]	SVM, KNN, RF	CHB-MIT	SHAP and LIME	×	Interpretation of feature importance	×
2025 Proposed	**Hybrid Transformer-DenseNet121-ANN**	**CHB-MIT**	**Attention weights, Saliency Maps, SHAP**	✓	**Feature Importance**	**Faithfulness-40.94% Completeness 1.00**

We considered a surrogate-based XAI approach in which the black-box model $f(x)$ is locally approximated using interpretable surrogate models.

$$\underset{g_v g_t g_q}{\operatorname{argmin}} \quad \sum_{e\in\{v,t,q\}} L\left(f(x), g_e(x)\right) + \Omega(g_e)) \qquad (3)$$

Eq. (3) extends the surrogate-model optimization framework introduced in [45] multiple surrogate models. Eq. (3) formulates the multimodal explainable model, providing multiple forms of explainability, such as visual, textual, and quantitative, for a model's prediction. x represents the input EEG features, $f(x)$ denotes a DL-based model predicting seizure or normal class, $g_v, g_{t,} g_q$ represents visual, textual, and quantitative explanations, respectively. 'e' iterates over explanation forms.

3. Challenges in Validating XAI Techniques

Explainable AI is an emerging field of research. There is no standardized approach to applying its evaluation metrics, and existing studies lack quantitative validation of the XAI techniques.

B. Key Contributions

To address these issues, we have designed and developed a hybrid, multimodal explainable model for the early and effective detection of epileptic seizures. To improve clinical interpretability, a multimodal explainability approach is employed, closely aligned with the roles of the underlying model components and the types of explanations they naturally support. In particular, attention weights from the Transformer are used to convey both visual and quantitative information about the temporal relevance of raw EEG segments. The attention weights are directly interpretable from the model without needing external tools. Saliency maps applied to the DenseNet121-based spectrogram encoder visually emphasize time-frequency regions that are most influential in seizure-related feature learning, and SHAP is employed with the ANN classifier to provide visual and textual explanations of feature contributions based on additive attribution principles. SHAP provides both global and local explanations, quantifying the direction (positive/negative) of each feature's impact. Together, these techniques provide complementary perspectives on model behavior, capturing temporal relevance, spatial patterns, and the importance of descriptive features. This enables a more comprehensive and clinically meaningful interpretation of seizure predictions than relying on a single explainability method, such as LIME or Grad-CAM. We prefer SHAP over LIME due to its additive attribution guarantees and support for both local and aggregated explanations, which are important for clinical reliability, and we adopt saliency maps instead of Grad-CAM to obtain high-resolution, input-level visualization of distributed seizure-related time-frequency patterns in EEG spectrograms. Thus, a unique combination of the Hybrid deep learning model and XAI techniques has great potential to enhance transparency and foster clinicians' trust in using complex deep learning models in clinical practice for early and effective diagnosis of epileptic seizures.

The main contributions of this research work are as follows:

1. The study explores intra-modal multimodality, leveraging complementary representations derived from a single modality, EEG. Distinct from inter-modal multimodal approaches, which fuse heterogeneous data sources (e.g., EEG with fMRI or ECG), intra-modal multimodal learning combines raw EEG and their time-frequency representations (spectrograms) to jointly capture complementary temporal and spectral seizure dynamics without increasing sensing complexity.
2. The study utilizes a combination of automated and handcrafted features to enhance the seizure detection accuracy and model interpretability.
3. The proposed model employs two different deep learning models: a Transformer to extract

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

sequential, temporal, and global contextual features, and DenseNet121 to extract Spatial, Spectral, and local visual patterns, providing a rich feature vector to discriminate between seizure and normal signals.

4. By integrating different XAI techniques, such as Attention weights, Saliency maps, and SHAP with the proposed model, the study provides multimodal explanations, such as visual, textual, and quantitative, for the model's decision-making process.

This study is structured as follows: Section II describes the proposed methodology. Section III presents the results of the proposed model. Section IV interprets the findings, compares the results with existing studies, and discusses the limitations and related future work. Finally, Section V concludes the research.

II. Method

This research aims to design and develop a multimodal, explainable deep neural network model to enhance interpretability and facilitate early, efficient diagnosis of epileptic seizures. It proposes a Hybrid

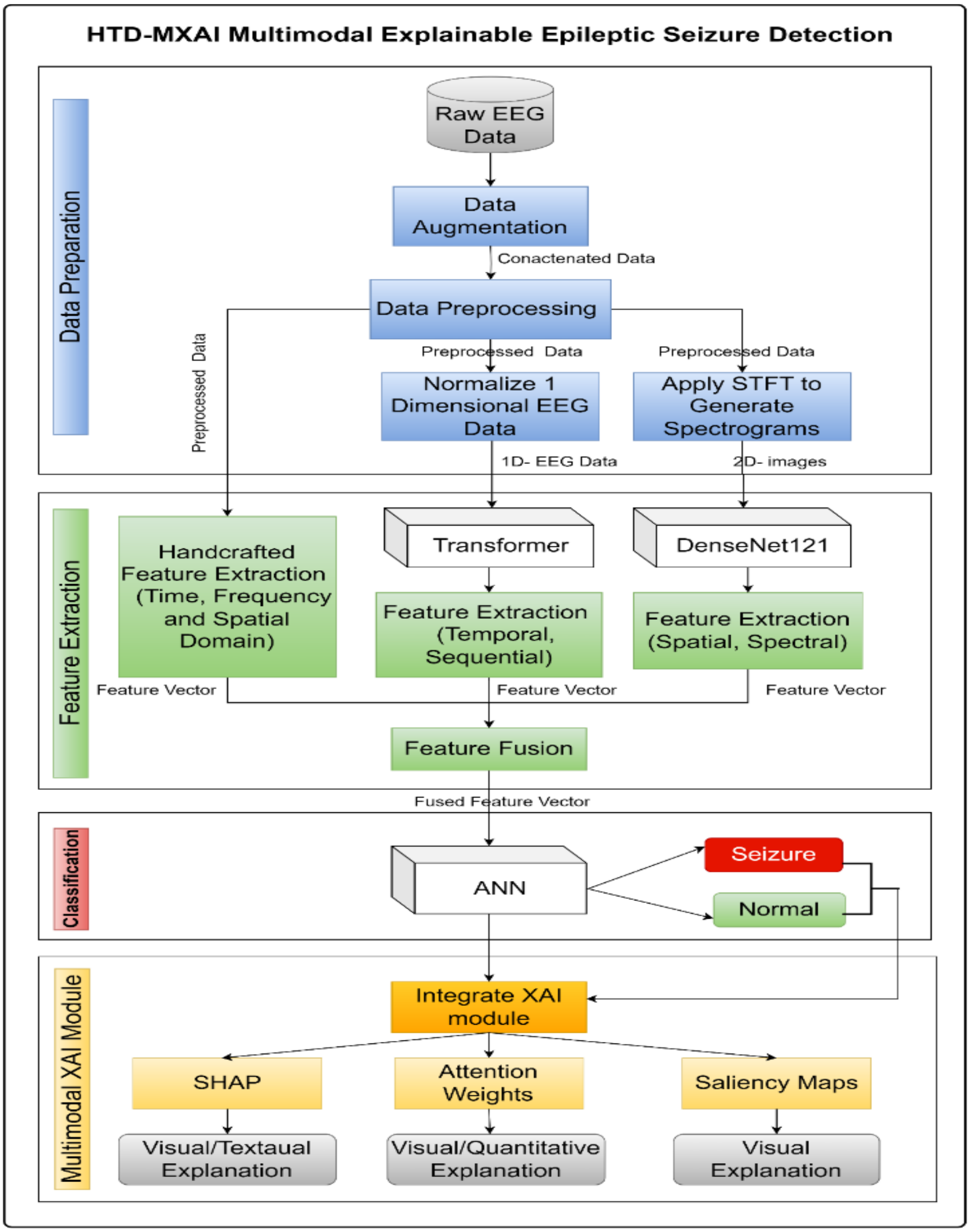


Fig.1. Proposed HTD-MXAI Epileptic Seizure Detection Model

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i2.1380

Transformer-DenseNet121-XAI (HTD-MXAI) integrated model for seizure detection using EEG data. The model uses intra-modal multi-modality input by combining 1-dimensional raw EEG time-series signals and 2-dimensional time-frequency images, along with handcrafted features extracted from the time, frequency, and spatial domains. In this work, a transformer model is applied to raw EEG data, whereas DenseNet121 is applied to spectrograms of EEG signals obtained via the Short-Time Fourier Transform (STFT). DenseNet121 learns local deep features from the time-frequency representations, whereas the transformer captures long-range temporal dependencies that CNNs typically struggle with due to their convolutional structure. These automatically learned features, together with handcrafted features, are fused and classified using an ANN. To address the limitations of prior work that focuses primarily on visual explanations, the proposed model delivers multimodal interpretability, visual, quantitative, and textual, providing deeper insight into model behavior and enhancing clinical transparency and trust. The design of the proposed HTD-MXAI consists of several steps, such as Data Preparation, Feature Extraction, Seizure classification, and integration of the XAI Module, as shown in Fig.1.

A. Dataset

The research uses the CHB-MIT scalp EEG dataset, available publicly via PhysioNet [46]. The dataset consists of recordings from 23 patients aged 1.5 to 22 years, grouped into 24 cases. The data are sampled at 256Hz, collected from 23 channels, and obtained from the Children's Hospital Boston, with each patient's seizure and non-seizure files. The continuous monitoring of brain activity across multiple days and varying conditions for each patient enhances the dataset's utility for developing robust and generalizable epileptic seizure detection models. The high sampling rate of 256 Hz preserves the temporal resolution necessary for accurate analysis of epileptic EEG patterns. Furthermore, the extensive scale of the dataset, comprising approximately 686 EEG recordings, enables the development and training of deep learning models tailored to epileptic seizure detection.

B. Data Preparation

1. Data Augmentation

Although brain signal data is recorded over long durations, seizure recordings are only available for very short periods, lasting for a minute or even seconds, compared to non-seizure recordings. To address class imbalance between seizure and normal classes, the proposed model employs a Generative Adversarial Network (GAN) architecture. GAN generates synthetic seizure samples by adding random noise. The GAN comprises a Generator G_θ, and a Discriminator D_ϕ . Both were implemented as fully connected neural networks operating on EEG-derived feature vectors. The Generator learns a mapping $G_\theta: \mathbb{R}^z \rightarrow \mathbb{R}^d$, where $z \sim \mathcal{N}(0, I)$ denotes a latent noise vector and d is the dimensionality of the EEG feature space, producing synthetic samples, $\hat{x} = G_\theta(z)$. The discriminator learns a mapping $D_\phi: \mathbb{R}^d \rightarrow [0,1]$, outputting the probability that a given input sample is real. Architecturally, the generator consists of three hidden layers with ReLU activation followed by a Tanh-activated output layer, while the discriminator employs multiple ReLU-activated hidden layers and a sigmoid output neuron. The networks are trained adversarially using the Generator and Discriminator loss functions as defined in Eq. (4) [47] and Eq. (5) [47].

Algorithm 1 Pseudocode for GAN Training for Generating Synthetic EEG Data

```
Input  : Real EEG Seizure Samples X
Output:  Trained Generator $G_\theta$
(1)   Initialize Generator $G_\theta$ and Discriminator $D_\phi$
(2)   Set hyperparameters for the Adam Optimizer, lr=0.0002, and $\beta_1 = 0.5$
(3)   for epoch=1 to n do
(4)       Sample minibatch x from X
(5)       Sample latent noise $z \sim \mathcal{N}(0, I)$
(6)       Generate synthetic samples $\hat{x} = G_\theta(z)$
            //Train Discriminator//
(7)       Compute $L_D$ using Eq. (5)
(8)       Update Discriminator parameters $\phi$
            //Train Generator//
(9)       Sample new noise $z \sim \mathcal{N}(0, I)$
(10)      Compute $L_G$ using Eq. (4)
(11)      Update Generator parameters $\theta$
(12)  endfor
(13)  Return Trained Generator $G_\theta$
```

The loss function of a Generator (G) is described as follows.

$$L_G = -\frac{1}{n}\sum_{i=1}^{n} \log\left(D(G(s_i))\right) \quad (4)$$

Where n denotes the number of samples (batch size), s_i denotes random noise input to Generator G, $G(s_i)$ denotes a synthetically generated sample.

The loss function of a Discriminator D is defined as follows.

$$L_D = -\frac{1}{n}\sum_{i=1}^{n}(\log D(x_i) + \log\left(1 - D(G(s_i))\right)) \quad (5)$$

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i2.1380

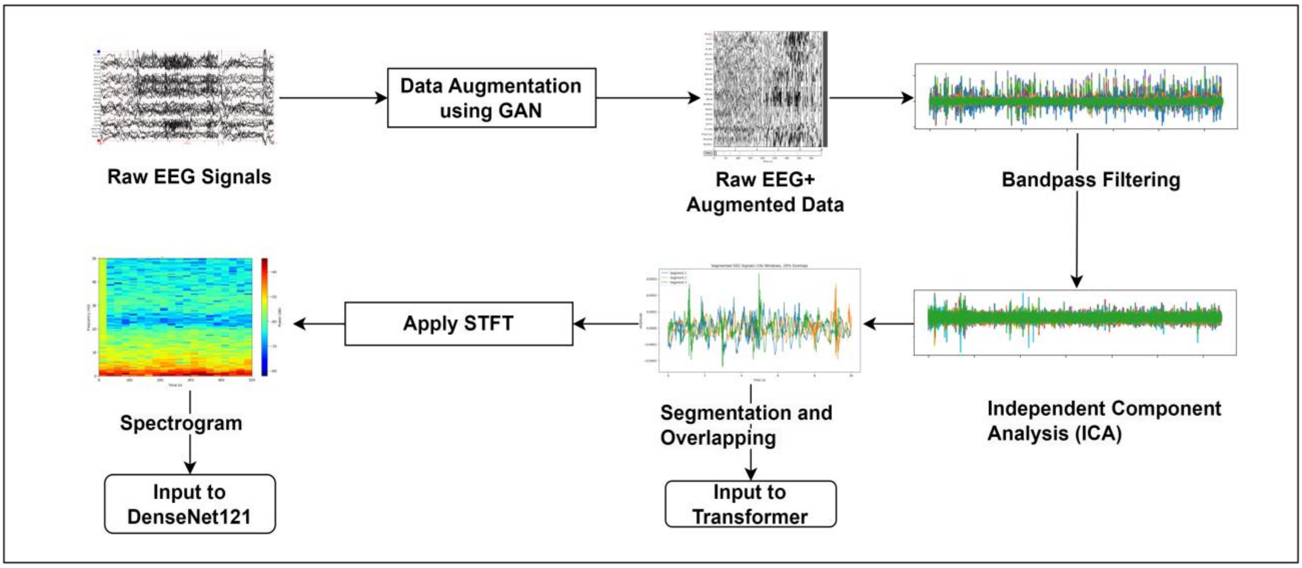


Fig.2. Data Preparation

Where, x_i represents a real EEG seizure sample from the dataset, D (x_i) represents the output of the discriminator, i.e., probability of realness, $D\big(G(s_i)\big)$ represents the discriminator's output for the generated sample. The Adam optimizer is applied to perform the optimization with a learning rate of 0.0002 and $\beta_1 = 0.5$, over 500 epochs with a batch size of 4. Algorithm 1 describes the pseudocode of the GAN training loop. It details the data augmentation process employed by the proposed model, which uses a GAN. The pseudocode outlines the functionality of the Generator and Discriminator in detail.

2. Data Preprocessing

Preprocessing techniques, such as Independent Component Analysis (ICA) and Bandpass Filtering, are applied to remove artifacts and noise from raw EEG data. The pre-processed data are then segmented into 10-second lengths to divide the continuous, long signals into smaller time windows. Also, the 20% overlap is used to maintain temporal continuity. Z-score normalization is used to center EEG signals at zero and scale their variance to 1. It helps the model to prevent feature dominance and learn effectively. Furthermore, this one-dimensional, pre-processed, and segmented data is fed as input to the transformer.

3. Short-Time Fourier Transform (STFT)

To prepare the input for CNN-based pre-trained models, the Short-Time Fourier Transform (STFT) is applied to the pre-processed data, converting the time-domain signal into the frequency domain. It generates 2-dimensional time-frequency images from the raw EEG signal, called Spectrograms. These 2D spectrograms are fed as input to the Pretrained model, DenseNet121. The STFT is mathematically defined in Eq. (6) [11], where x(t) represents the EEG signal as a function of time 't', m represents the time shift of the window, $w(t-m)$ represents a window function centered at time m, $e^{-j2\pi ft}$ represents a complex exponential denoting a sinusoidal wave at frequency 'f'.

$$STFT\{x(t)\}(m,f) = \int x(t)\, w(t-m) e^{-j2\pi ft} dt \quad (6)$$

In this study, a Hann window of length 256 samples (1 s at a sampling frequency of 256 Hz) with 50% overlap (128 samples) is used to compute the STFT. Fig. 2 describes the detailed process of data preparation shown above.

C. Automated Feature Extraction

The study employs two different deep learning architectures for automatically extracting features from EEG data: the Transformer model and DenseNet121, applied to different input forms of EEG data.

1. Transformer

The proposed framework adopts a Transformer-based encoder to analyze one-dimensional, pre-processed EEG signals, enabling the model to capture their underlying temporal and sequential dynamics. The architecture is composed of several stacked encoder blocks, each incorporating multi-head self-attention, a position-wise feed-forward network, residual pathways, and layer normalization. A Global Average Pooling (GAP) layer is applied at the final stage to aggregate the learned representations. Let $X \in r\mathbb{R}^{T\times C}$ denote a pre-processed one-dimensional EEG segment, where T represents the length of the segment and C denotes the number of EEG channels. The input sequence is first linearly projected into a higher-dimensional embedding space of dimension d_{model} before being fed into the Transformer encoder. Since the Transformer

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i2.1380

architecture does not inherently encode temporal order, a positional encoding is added to the input embeddings to preserve sequential information. The positional encoding matrix $P \in \mathbb{R}^{T\times d_{\text{model}}}$ is defined as in Eq. (7) [48]:

$$P_{(t,2i)} = \sin\left(\frac{t}{10000^{2i/d_{\text{model}}}}\right),$$
$$P_{(t,2i+1)} = \cos\left(\frac{t}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \quad (7)$$

The final input to the Transformer encoder is computed as in Eq. (8) [48]:

$$Z = XW_e + P \quad (8)$$

where $W_e \in \mathbb{R}^{C\times d_{\text{model}}}$ is the learnable embedding matrix.

Each Transformer encoder block consists of a multi-head self-attention mechanism. Attention is computed through scaled dot-product operations using three projected vector sets-Query (Q), Key (K), and Value (V), and also the model learns three distinct weight matrices: the Query Weights (W^Q), the Key Weights (W^K) and the Value Weights (W^V) which are used to compute attention scores and contextual representations. For each attention head $i \in \{1,\dots,h\}$, the input sequence Z is linearly projected into Query, Key, and Value representations as described in Eq. (9) [48].

$$Q_i = Z.W_i^Q, \quad K_i = Z.W_i^k, \quad V_i = X_v.W_i^v \quad (9)$$

The attention weights are computed by measuring the similarity between the query and key vectors. The attention output for head 'i' is formulated as in Eq. (10) [48], where d_k denotes the dimension of the key vectors

$$head_i = Softmax\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right)V_i \quad (10)$$

The attention weights from multiple heads are concatenated using Eq. (11) [48], and the final projection matrix, W^o is applied.

$$MHAttn(X) = Concat[head_1, \dots, head_h].W^o \quad (11)$$

where $W^O \in \mathbb{R}^{d_{\text{model}}\times d_{\text{model}}}$ is the output projection matrix.

The attention output is passed through a position-wise feed-forward network (FFN), applied independently to each time step as described in Eq. (12) [48]:

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (12)$$

Where $W_1 \in \mathbb{R}^{d_{\text{model}}\times d_{ff}}$, $W_2 \in \mathbb{R}^{d_{ff}\times d_{\text{model}}}$, and b_1, b_2 are learnable bias parameters. Residual connections and layer normalization are applied after both the attention and feed-forward sublayers to stabilize the training process. Each block processes the input sequence and extracts the high-level sequential, temporal, and brainwave frequency band-related features. These features are then refined using a Global Average Pooling operation applied along the temporal dimension. The feature vector is extracted from this layer and utilized for further processing. The final output of the stacked Transformer encoder layers is denoted as $H \in \mathbb{R}^{T\times d_{\text{model}}}$

$$F_{temporal,Global} = \frac{1}{T}\sum_{t=1}^{T} H_t \quad (13)$$

Eq. (13) represents the standard temporal global average pooling operation used to aggregate sequence-level representations in transformer-based models [48]. In this work, a fixed-length feature vector of 2560 dimensions is produced for each EEG segment.

2. DenseNet121

The DenseNet121 processes 2D spectrograms using convolutional layers, extracting deep, hierarchical, spatial features and frequency-dependent patterns. In the modified architecture, the final classification layer of DenseNet121 is replaced with a convolution layer, Global Average Pooling (GAP), and a Fully Connected (Dense) layer. Let the input provided to the DenseNet121 be defined as $X \in \mathbb{R}^{224\times224\times3}$. Eq. (14) [49] formulates the output feature map at l^{th} layer. The l^{th} layer receives the output feature maps from all preceding layers, where, $[x_0, x_1, \dots, x_{l-1}]$ denotes the concatenation of all output feature maps from preceding layers. T_l represents a non-linear transformation at the layer, l comprising batch normalization, RELU activation function, and convolution operations.

$$x_l = T_l([x_0, x_1, \dots, x_{l-1}]) \quad (14).$$

The DenseNet121 outputs a feature embedding, a high-dimensional vector, defined as follows in Eq. (15) [49].

$$F_{DenseNet} \in \mathbb{R}^{7\times7\times1024} \quad (15)$$

After that, the convolutional layer Conv2D with kernel K is applied to $F_{DenseNet}$, as described in Eq. (16). It refines the extracted features by capturing additional spatial patterns further and reducing feature map dimensions from (7,7,1024) to (5,5,128), following the standard convolution activation formulation used in DenseNet-based architectures [49].

$$F_{conv} = RELU\,(F_{DenseNet} * K + b) \quad (16)$$

GAP has been widely adopted in CNNs to convert feature maps into compact vectors [50]. The feature vector is refined through Global Average Pooling (GAP), converting the (5,5,128) feature map into a 128-dimensional feature vector by averaging across the spatial dimensions, as defined in Eq. (17) [50].

$$F_{GAP} \in \mathbb{R}^{128} \quad (17)$$

A compact, high-dimensional feature vector (f) is extracted through that layer and transferred for further processing as described in Eq. (18) [49].

$$f = F_{DenseNet}(x) \in \mathbb{R}^d \quad (18)$$

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i2.1380

Eq. (19) describes the automated extraction of spatial, spectral, and local features using DenseNet121 and spectrograms. A total of 128 features were extracted per segment using the DenseNet121 model.

$$F_{spatial-spectral,local} = DenseNet(x_{spectrogram}) \quad (19)$$

In this work, DenseNet121 is employed with a transfer-learning strategy. The network is initialized with ImageNet pre-trained weights, and the convolutional backbone layers are frozen during training to preserve generic spatial feature representations. Only the newly added convolutional, Global Average Pooling, and fully connected layers are fine-tuned on EEG spectrogram data as implemented in our experimental setup. This design choice improves generalization, reduces overfitting on limited seizure samples, and ensures compatibility with the input spectrogram dimensions $224 \times 224 \times 3$.

D. Handcrafted Feature Extraction

The study simultaneously extracts manual features from EEG data in both the time and frequency domains and computes spatial features.

1. Time Domain (Temporal) and Statistical Features

The time-domain features quantify the signals' shapes, variability, and temporal distributions. The following temporal and statistical features are extracted from the EEG data: Mean, Standard Deviation (SD), Skewness, Kurtosis, Permutation Entropy (PE), Variance, Zero Crossing Rate (ZCR), and Root Mean Square (RMS).

2. Frequency Domain (Spectral) Features

Frequency-domain features, such as Band Power, Spectral Entropy, and Power Spectral Density (PSD) for the Delta, Theta, Alpha, Beta, and Gamma frequency bands, are computed and extracted. For a pre-processed EEG segment $x(t)$, the Power Spectral Density (PSD), $P(f)$ is estimated using Welch's method. The band power for a frequency band $[f_1, f_2]$ is computed as in Eq. (20) [51].

$$P_{band} = \int_{f2}^{f1} p(f)\, df \quad (20)$$

The mean Spectral Amplitude is computed as in Eq. (21) [51].

$$\mu_{amp} = \frac{1}{N} \sum_{k=1}^{N} \sqrt{P(f_k)} \quad (21)$$

Where N denotes the number of frequency bins. The Spectral Entropy quantifies the complexity or disorder of the EEG signal's frequency content. It is derived by normalizing the PSD and applying Shannon entropy as described in Eq. (22) [51].

$$H_{Spec} = -\frac{1}{log_2 N} \sum_{k=1}^{N} \tilde{P}(f_k)\, log_2 \tilde{P}(f_k) \quad (22)$$

These features facilitate analysis of the frequency distributions of both seizure and non-seizure EEG signals. Specifically, analyzing waveform-based spectral features enables detection of variations in brain activity using amplitude and frequency. Spectral features in the frequency domain serve as a powerful representation, capturing key differences in the brain's functional and behavioural characteristics.

3. Spatial Features

Cross-correlation: This metric quantifies the similarity between two EEG signals from different channels. It computes the correlation for each pair and stores the mean cross-correlation for that pair. The final output is the average cross-correlation for each segment across all pairs of channels.

Coherence: For each pair of channels, it calculates the cross-spectral density by multiplying the FFT of the first channel with the Fast Fourier Transform (FFT) of the other channel, using the conjugate of the first channel's FFT. Finally, it computes coherence by normalizing the cross-spectral density by the product of the two channels' PSDs.

Eq. (23) formulates the handcrafted feature extraction. It extracts 243 handcrafted features per segment.

$$F_{Handcrafted} = Extract(x_{Handcrafted}) \quad (23)$$

E. Feature Fusion

The handcrafted and automated features extracted from the Transformer and DenseNet121 models are fused as described in Eq. (24), yielding a combined feature vector of size (None, 2931).

$$F_{fused} = F_{temporal,Global} \,||\, F_{spatial-spectral,local} \,||\, F_{Handcrafted} \quad (24)$$

The details are as follows. The automated features through Transformer (1 to 2560), DenseNet121(2561 to 2668), and Handcrafted features (Frequency (Delta: 2689 to 2691, Theta: 2692 to 2694, Alpha: 2695 to 2697, Beta: 2698 to 2700, Gamma: 2701 to 2703), Spatial (Correlation: (2704 to 2725) Coherence: (2726 to 2747), Time (Mean: 2748 to 2770, SD:2771 to 2793, Skewness: 2794 to 2816, Kurtosis: 2817 to 2839, PM: 2840 to 2862, Variance: 2863 to 2885, ZCR:2886 to 2908, RMS: 2909 to 2931)

F. Classification

The resulting fused feature representation $F_{\text{fused}} \in \mathbb{R}^{2931}$ is provided as input to a fully connected Artificial Neural Network (ANN) for final classification. The ANN consists of three hidden layers with 128, 64, and 32 neurons, respectively, each employing the ReLU activation function, followed by dropout regularization to mitigate overfitting. The output layer contains a single neuron with sigmoid activation to perform binary classification (seizure vs. normal). The forward propagation of the ANN is defined in Eq. (25) [52].

$$h_1 = ReLU\,(W_1 F_{\text{fused}} + b_1),\ h_2 = ReLU\,(W_2 h_1 + b_2),$$
$$h_3 = ReLU\,(W_3 h_2 + b_3),\ \hat{y} = \sigma\,(W_0 h_3 + b_0) \quad (25)$$

where W_i and b_i denote the trainable weights and biases of the i^{th}layer, and $\sigma(\cdot)$represents the sigmoid

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

activation function. The network is optimized using the Adam optimizer by minimizing the binary cross-entropy loss function as described in Eq. (26) [52].

$$\mathcal{L} = -[y \log(\hat{y}) + (1-y)\log(1-\hat{y})] \qquad (26)$$

where $y \in \{0,1\}$denotes the ground-truth class label and $\hat{y}$ is the predicted probability of seizure occurrence.

The detailed architecture of the proposed model, comprising its core components, is shown in Fig. 3.

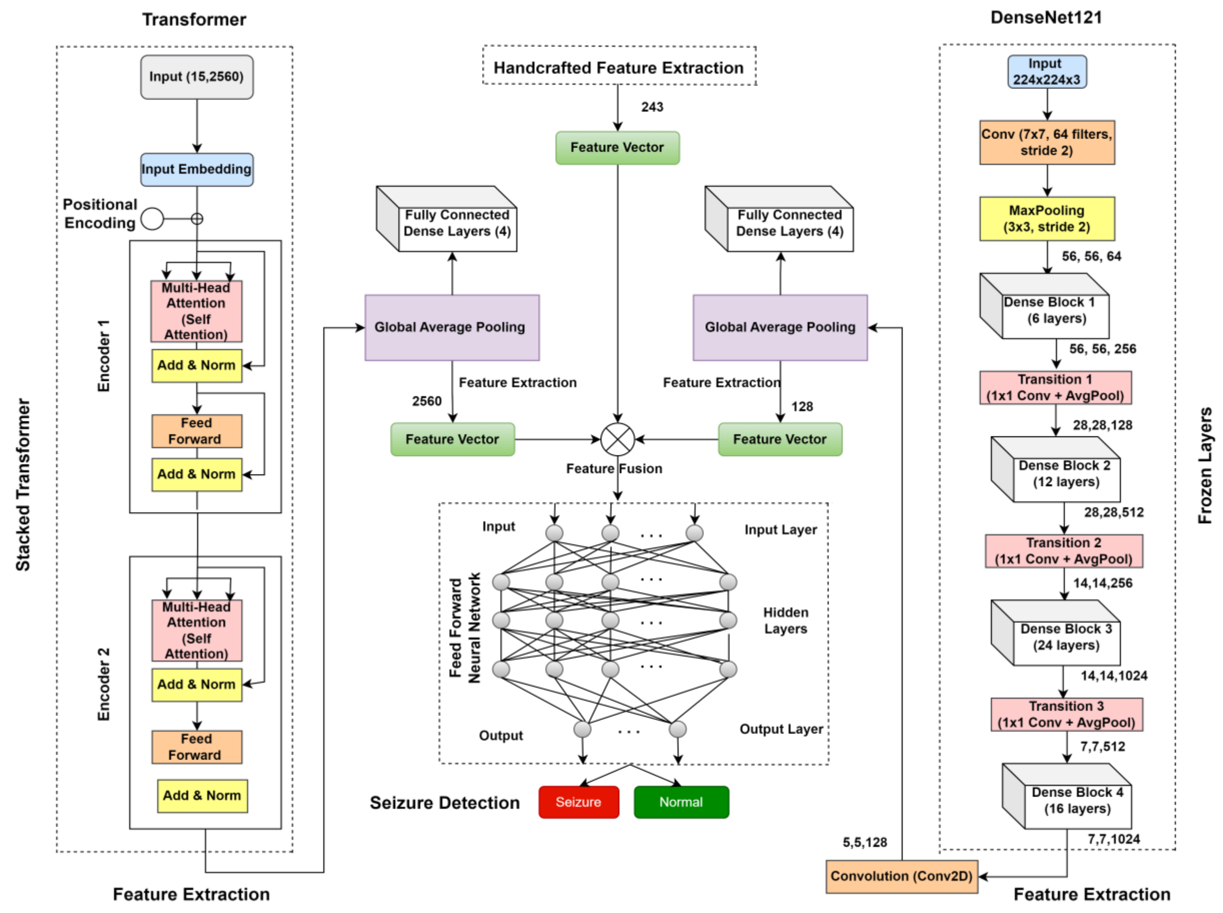


Fig.3. Architecture of Hybrid Transformer-DenseNet121-ANN Seizure Detection Model

G. Multimodal Explainable AI module

The proposed model integrates various XAI techniques, such as Attention Weights, Saliency Maps, and SHAP values, with the Hybrid Transformer-DenseNet121-ANN model to interpret the model's decision-making process. The visual explanation is presented using heatmaps of attention weights computed by the Transformer's self-attention mechanism and saliency maps for DenseNet-121, highlighting the input regions critical for seizure detection. The quantitative explanation is incorporated using attention weights for important features. The proposed model produces structured textual explanations by analyzing Transformer attention weights. The attention scores are aggregated and mapped to predefined EEG frequency bands (delta, theta, alpha, beta, and gamma), and the relative contributions of these bands are expressed as percentages. It reports the dominant frequency band affecting model predictions in a compact, rule-based textual format, thereby providing a human-readable, interpretable explanation in the frequency domain without using post hoc language generation.

SHAP assigns an importance value to each feature based on its contribution to the model's output, grounded in cooperative game theory. The SHAP value ϕ_i for the i^{th} feature is computed using the Shapley value formulation, as described in Eq. (27) [53].

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \qquad (27)$$

where F denotes the set of all features, S is a subset of features excluding the feature i, and $f(\cdot)$ represents the trained model's prediction function. This formulation

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i2.1380

ensures properties such as local accuracy, consistency, and missingness. Eq. (28) describes the multimodal explainable model that utilizes the above XAI techniques to explain the prediction made by the classifier for the input signal.

$$\hat{y} \rightarrow MXAI(AttentionWeight, Saliency, SHAP) \quad (28)$$

We evaluated the faithfulness of the generated explanations using an accuracy-drop perturbation test on the features. Faithfulness determines how well the explanation aligns with the model's prediction. The attribution values are computed for the features, and the features are then ranked based on the attribution scores generated by the XAI technique. The top 40% of features by attribution were removed, and the modified feature vectors were then passed through the trained model to determine how their removal affects the model's predictive performance. In this work, we define the faithfulness score as the reduction in model accuracy after removing the most important features, as shown in Eq. (29), as follows [54].

$$Faithfulness = Original\ Accuracy - Perturbed\ Accuracy \quad (29)$$

A high score indicates that removing highly attributed features causes greater performance degradation. This indicates that the explanation is more representative of the model's true decision-making behavior. This method is consistent with perturbation-based faithfulness evaluation in the explainable AI literature [54], where performance drops have been used as an indicator of how well aligned explanations are with model reliance.

Algorithm 2. Pseudocode for Proposed HTD-MXAI Model for Epileptic Seizure Detection

```
Input  :  Raw EEG Signals
Output:   Seizure and Normal signals, Visual,
          Quantitative, and Textual Explanation
              //Data Augmentation//
(1)   Acquire EEG recordings
(2)   Separate seizure and non-seizure recordings
(3)   Apply GAN to generate synthetic seizure samples using Eq. (4) and Eq. (5)
(4)   Concatenate the synthetically generated data with the original EEG data
              //Preprocessing//
(5)   for concatenated input EEG data in step 4 do
(6)       Apply Independent Component Analysis (ICA)
(7)       Apply a Bandpass filter
(8)       Apply Segmentation and overlapping
(9)       Apply normalization using $S' = \frac{S-\mu}{\sigma}$
(10)      Apply Short-Time Fourier Transform (STFT) using Eq. (6)
              //Transfer Learning//
(11)  Load the pretrained DenseNet121 model
(12)  Train the model on concatenated EEG data
              //Automated Feature Extraction//
(13)  for the preprocessed EEG segment in step 9 do
(14)      Apply the proposed Transformer model
(15)      Reduce feature dimensions using Global Average Pooling (GAP)
(16)      Store the feature vector
(17)  endfor
(18)  for 2D spectrograms in step 10 do
(19)      Apply the DenseNet121 model
(20)      Reduce feature dimensions using Global Average Pooling (GAP)
(21)      Store the feature vector
(22)  endfor
              //Handcrafted Feature Extraction//
(23)  for the normalized EEG segment in step 9 do
(24)      Compute time-domain features,
          Mean $\mu = \frac{1}{N}\sum_{i=1}^{N} s_i$,
          SD $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i-\mu)^2}$
          Skewness $Sk = \frac{1}{N}\sum_{i=1}^{N}(\frac{s_i-\mu}{\sigma})^3$,
          Kurtosis $K = \frac{1}{N}\sum_{i=1}^{N}(\frac{s_i-\mu}{\sigma})^4$ ,
          PE, Variance, ZCR, and RMS
(25)      Compute frequency-domain features using Eq. (20) to Eq. (22)
(26)      Compute Spatial features
(27)      Store a handcrafted feature vector
(28)  endfor
              //Feature Fusion//
(29)  Fuse the handcrafted features and the automated features extracted through
      the Transformer and DenseNet121 in steps 16,21, and 27
              //Seizure Detection//
(30)  Input the fused features to a feed-forward Artificial Neural Network (ANN)
(31)  Apply a feed-forward Artificial Neural Network (ANN) on test samples
(32)  Classify seizure and normal EEG segments
```

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeemi.v8i2.1380

```
                //Integrating Multimodal XAI //
                    //Attention Weights//
(33)   for all transformer encoder blocks
(34)       for all heads do
(35)           Extract Attention Weights from
               the attention layers
(36)       endfor
(37)       Plot attention maps to visualize the
           time points
(38)   endfor
                    //Saliency Maps //
(39)   Apply Saliency Maps to compute
       gradients of DlenseNet121 and plot a
       heatmap
       //SHAP //
(40)   for each input feature in step 29
(41)       Apply SHAP
(42)       Visualize feature importance using
           SHAP summary plots and bar plots
(43)   endfor
       //Textual Explanation//
(44)   Generate a textual explanation based
       on the prediction outcome by the model
(45)   endfor
```

Another measure used to evaluate the performance of XAI techniques is the completeness score. It measures whether the sum of all feature attributions approximates the model output. SHAP satisfies the completeness (local accuracy) axiom, where the prediction for an input can be expressed as a sum of the output for a baseline instance and feature attributions, as shown in Eq. (30) [55]. ϕ_i denotes SHAP attribution of feature 'i'.

$$f(x) = f(x_{baseline}) + \sum_{i=1}^{n} \phi_i \tag{30}$$

Since the original definition of SHAP does not yield a scalar value for completeness fidelity, we defined a Completeness Score using Eq. (31) on the basis of the completeness axiom, referring to Eq. (30), on how well the SHAP attributions reconstruct the model output. A score of 1 indicates perfect fidelity, while lower values indicate deviation.

$$Completeness = 1 - |\, f(x) - (f(x_{baseline}) + \sum_{i=1}^{n} \phi_i)| \tag{31}$$

Table 2. Implementation Parameters of the Proposed Model

Parameters	Transformer	DenseNet121	ANN
Number of Transformer Encoders	2	N/A	N/A
Number of attention heads	3	N/A	N/A
Dropout Rate	0.2	0.2	0.3
Learning Rate	0.001	0.001	0.001
Loss Function	Binary Cross-Entropy	Binary Cross-Entropy	Binary Cross-Entropy
Batch Size	32	32	32

The pseudocode of the proposed model is presented in Algorithm 2. It describes the workflow and step-by-step procedure of the proposed model. It is included to improve the understanding, clarity, and reproducibility of the proposed work, thus allowing the researchers to implement and validate the proposed work.

H. Experimental Settings and Performance Metrics

The proposed model is evaluated using data from 21 patients in the CHB-MIT dataset, comprising 160 seizures. Cases 12 and 13 have recurrent variations in channel configuration, whereas case 24 has insufficient seizure data during the EEG recordings. Thus, these cases were excluded during the experimentation. The dataset is divided into 60% for training, 20% for validation, and 20% for testing. In addition, the performance of the HTD-MXAI Seizure Detection

Table 3. Comparative Performance of MobileNetV2, EfficientNetB0, and DenseNet121

Classifier	Epileptic Seizure Detection Performance (%)						
	#Parameters	Accuracy	Precision	F1-score	Sensitivity	Specificity	AUC
MobileNetV2	3.5 M	52.40	52.40	68.77	100	0.0	50.45
EfficientNet-B0	5.3 M	52.40	52.40	68.77	100	0.0	52.17
DenseNet121	8.1 M	95.43	93.35	95.18	97.09	93.99	97.89

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

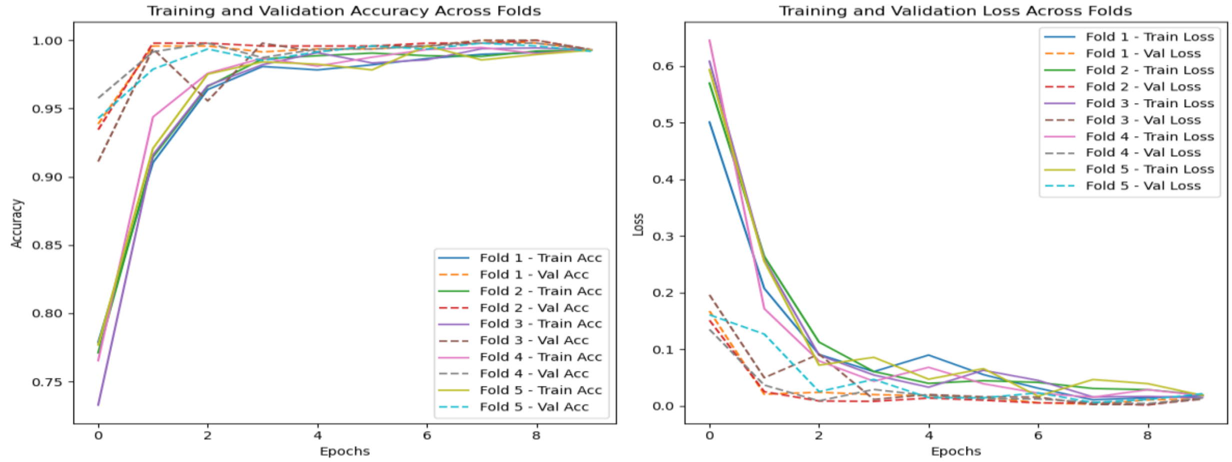


Fig. 4. **Training and Validation Accuracy and Loss Curves across Folds**

Model is further validated via 5-fold cross-validation on the CHB-MIT dataset. A 5-fold cross-validation ensures generalization to limited seizure data through a robust and unbiased evaluation of the seizure detection model. The details of the training protocol and hyperparameter settings used to implement the Transformer, DenseNet121, and ANN models are presented in Table 2. To evaluate the performance of the Hybrid epileptic seizure detection model, standard performance metrics such as Accuracy, Precision, F1-Score, Sensitivity, Specificity, and Area Under the ROC Curve (AUC) are used. To assess the performance of XAI techniques, metrics such as Faithfulness and Completeness are employed in this study. These metrics provide a quantitative evaluation of the SHAP-based feature attributions with respect to model prediction.

III. Result

This section presents the results of an extensive evaluation of our proposed model on the CHB-MIT dataset. The performance evaluation was conducted considering different components and configurations to identify and report the most effective version. We presented the results from the experimental evaluations, ablation studies, and validation in this section.

A. Performance Comparison of Pretrained Models

Initially, we compared the performance of CNN-based pre-trained models on the CHB-MIT dataset, including MobileNetV2, EfficientNet-B0, and DenseNet121, for extracting deep features from EEG data and classifying seizure and normal EEG signals. We have chosen MobileNetV2 and EfficientNet-B0 models, considering the computational cost. Because these models have fewer than 10 million parameters, they are considered lightweight and substantially reduce computational cost. As these models are CNN-based, pre-trained on large-scale datasets, they are fine-tuned on EEG data and applied to test data to extract features and classify the input. The input to these models is a 2D spectrogram. MobileNetV2 and EfficientNet-B0 have

Table 4. **Impact of Combined Feature Extraction on Seizure Detection (in %)**

Feature Extraction	Accuracy	Precision	F1-score	Sensitivity	Specificity	AUC
Transformer	83.48	82.35	81.55	80.77	85.71	84.78
DenseNet121	95.43	93.35	95.18	97.09	93.99	97.89
Hybrid-Automated (Transformer+ DenseNet121)	97.63	100	97.38	94.90	100	99.81
Handcrafted	94.58	93.84	94.49	95.15	94.05	98.44
Combined Handcrafted+ Automated	99.14	99.62	99.05	98.49	99.68	99.81

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

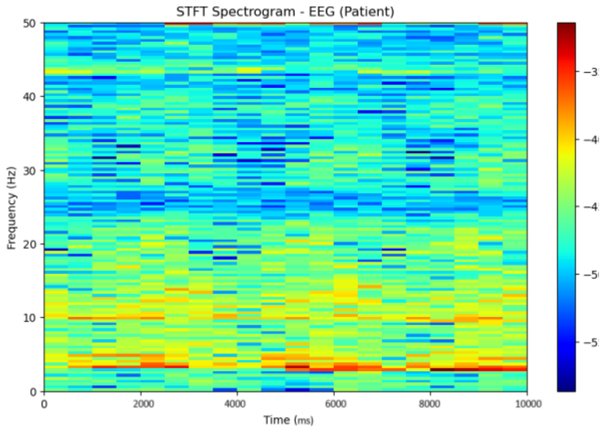


Fig. 5. Input Spectrogram

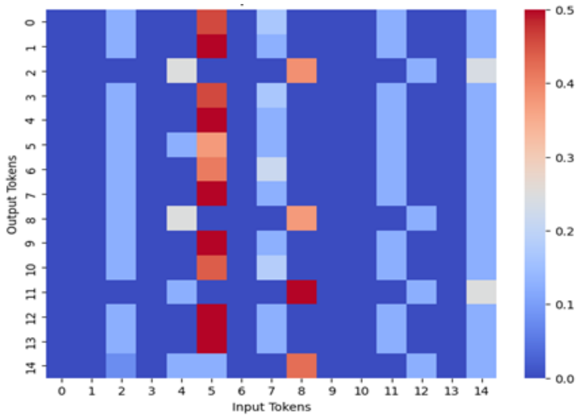


Fig. 6. Attention Heatmap

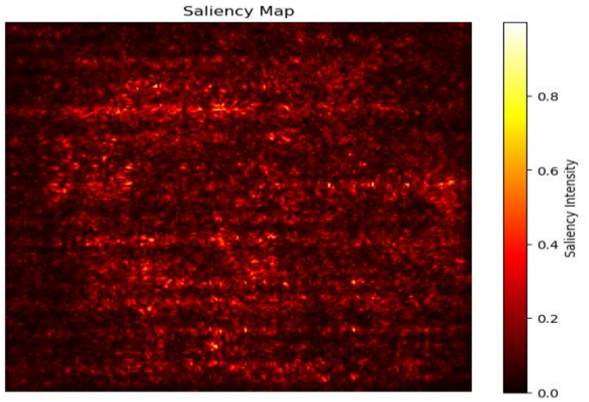


Fig. 7. Saliency Map

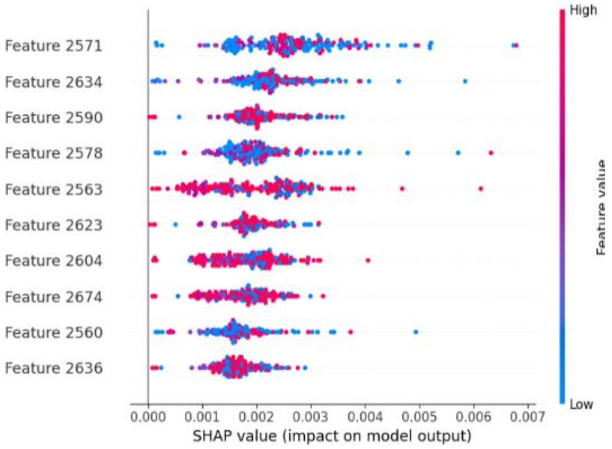


Fig. 8. SHAP Summary Plot

shown poor performance on the CHB-MIT dataset. Table 3 provides a comparative performance analysis of MobileNetV2, EfficientNet-B0, and DenseNet-121 models for Epileptic Seizure Detection on the CHB-MIT dataset. The DenseNet121 outperforms the other two models, achieving an Accuracy of 95.43%, a Precision of 93.35%, an F1-score of 95.18%, a Sensitivity of 97.09%, a Specificity of 93.99%, and an AUC of 97.89%.

B. Ablation Study and Model Performance Validation

An ablation study was conducted to analyze the individual and combined contributions of the Transformer and DenseNet121 models in extracting automated features for seizure detection. Additionally, an ablation study was conducted to evaluate classification performance with handcrafted features alone and with handcrafted and automated features combined. Table 4 illustrates the impact of combining automated and handcrafted features on seizure detection. The results demonstrate that DenseNet121 outperforms the Transformer model. However, the Hybrid approach achieves the highest performance, with an accuracy of 97.63%, a precision of 100%, an F1-score of 97.38%, a sensitivity of 94.90%, a specificity of 100%, and an AUC of 99.81% after fusing features extracted from the Transformer and DenseNet-121. The classification performance achieved using handcrafted features alone is lower than that obtained with automated features. The combined approach of automated and handcrafted features yields optimal classification performance with an accuracy of of 99.14%, a precision of 99.62%, an F1-score of 99.05%, a sensitivity of 98.49%, a specificity of 99.68%, and an AUC of 99.81%.

The HTD-MXAI model was trained for 10 epochs using a stratified 5-fold cross-validation. It was observed that both training and validation losses stabilized within the initial epochs, leading to early convergence. Fig. 4 presents the training and validation accuracy and loss trends across 5-folds, which are examined to assess the model's learning

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i2.1380

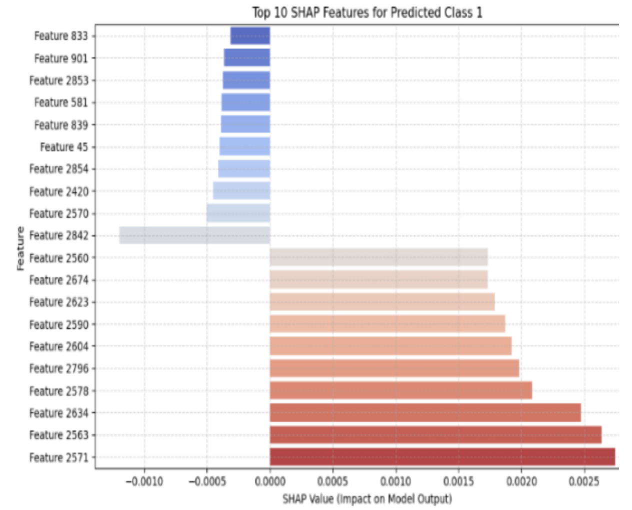


Fig. 9. **Top 10 SHAP Features for Predicted Seizure Class**

behavior. In our experiments, the training and validation accuracies remained consistently high across all folds, and the small gap between them indicates that the model did not merely memorize the training data. Instead, it appeared to learn the characteristics of the dataset effectively. Another reassuring observation was that both loss curves decreased steadily and approached very low values after the initial epochs, suggesting that the optimization process converged efficiently. Taken together, these trends suggest that the proposed model not only fits the data well but also generalizes to unseen samples, which is crucial for its practical applicability.

C. Performance Evaluation of XAI Techniques

Further, this subsection reports the outcome of XAI techniques integrated with the proposed epileptic seizure detection model. The XAI techniques are evaluated on the number of input test samples, including seizure and normal segments. However, we report findings on a seizure-specific input sample to emphasize the proposed model's ability to correctly identify the seizure segment and to demonstrate the effectiveness of XAI techniques in interpreting the model's decision. The STFT spectrogram of the input test sample is shown in Fig. 5. The spectrogram indicates seizure activity between 5000 and 7000 milliseconds, as indicated by a high-power concentration in the lower-frequency bands, delta (0.5 to 4 Hz) and theta (4 to 8 Hz). It could represent the onset and propagation of an epileptic seizure. The model correctly classified it as a Seizure segment. The following explanations are generated with the XAI techniques. Fig. 6 shows the attention heatmap generated for the output of a Transformer block. Attention weights provide temporal explainability by highlighting EEG segments that most strongly contribute to the classification decision. The attention weights predominantly emphasize temporally localized EEG segments corresponding to seizure-related rhythmic activity and sustained discharges. Such temporal concentration aligns with established clinical observations that epileptic seizures exhibit distinct time-evolving EEG patterns rather than uniform activity across the entire recording [56]. Saliency maps are used to highlight regions of importance for seizure detection. Fig. 7 shows the saliency map generated for the input EEG spectrogram. The saliency maps highlight time-frequency regions corresponding to increased low-frequency (delta, theta) and broadband power during seizure intervals, which are well-known EEG signatures of epileptic activity. These regions align closely with the dominant energy components of the input spectrograms, indicating that the model focuses on clinically meaningful seizure-related patterns rather than artifacts. Similar spectral characteristics during epileptic events have been consistently reported in clinical EEG studies [56]. The SHAP is applied to explain the impact of automated, handcrafted, and combined features on the model's prediction. Fig. 8 presents the SHAP summary plot for automated features extracted through Transformer and DenseNet121. It indicates that Feature 2571 is the most important because it has the widest range of SHAP values.

Fig. 9 visualizes the contribution of combined features to the model's outcome. The SHAP values are sorted to identify the top positive (supporting) and negative (opposing) features. It helps interpret results by identifying which features pushed the prediction toward class 0 (Normal) or 1 (Seizure). This indicates that automated features dominate handcrafted features. Among the top 10 supporting features for classifying the input segment as seizure, 8 features are extracted through DenseNet121 (Feature 2571, 2563, 2634, 2578, 2604, 2590, 2623, 2674), 1 feature is extracted through Transformer (Feature 2560), and 1 is a handcrafted feature (Feature 2796 referring to skewness). After

Table 5. **XAI Performance Evaluation**

XAI Technique (SHAP)	Original Accuracy	Perturbed Accuracy	Faithfulness	Completeness
	99.14%	58.20%	40.94%	1.00

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

performing SHAP analysis on handcrafted features and grouping them, the results demonstrate that the Time-domain features dominate the Spectral and Spatial features. The “Skewness” is quantified as the most important feature, along with “Kurtosis,” which also has the same importance. The neurophysiological basis for the pivotal role of higher-order time-domain statistics, in particular skewness and kurtosis, lies in the occurrence of sharp transients and non-Gaussian amplitude distributions in epileptic seizures. The prevalence of

Table 6. Subject-wise Performance Comparison of Individual and Hybrid models on the CHB-MIT Dataset

Patient ID	Classifier	Performance (%)					
		Accuracy	Precision	F1-Score	Sensitivity	Specificity	AUC
CHB01	Transformer	98.97	100	98.92	97.87	100	99.96
	DenseNet121	97.94	97.87	97.87	97.87	98	98.29
	Hybrid	98.97	100	98.92	97.87	100	100
CHB02	Transformer	95.08	87.5	90.32	93.33	95.65	97.75
	DenseNet121	100	100	100	100	100	100
	Hybrid	93.44	92.30	85.71	80	97.83	99.42
CHB03	Transformer	95.69	97.43	95	92.68	98.07	99.62
	DenseNet121	100	100	100	100	100	100
	Hybrid	100	100	100	100	100	100
CHB04	Transformer	94.82	100	80	66.67	100	97.95
	DenseNet121	96.55	100	87.5	77.78	100	99.77
	Hybrid	100	100	100	100	100	100
CHB05	Transformer	98.14	100	98.03	96.15	100	100
	DenseNet121	100	100	100	100	100	100
	Hybrid	100	100	100	100	100	100
CHB06	Transformer	98.96	100	98.92	97.87	100	100
	DenseNet121	96.90	100	96.70	93.61	100	99.57
	Hybrid	100	100	100	100	100	100
CHB07	Transformer	100	100	100	100	100	100
	DenseNet121	100	100	100	100	100	100
	Hybrid	100	100	100	100	100	100
CHB08	Transformer	93.79	93.40	94.97	96.59	89.47	97.15
	DenseNet121	100	100	100	100	100	100
	Hybrid	93.10	92.39	94.44	96.59	87.71	98.21
CHB10	Transformer	97.94	97.87	97.87	97.87	98	99.70
	DenseNet121	100	100	100	100	100	100
	Hybrid	98.96	97.92	98.95	100	98	99.74
CHB23	Transformer	98.59	94.74	97.29	100	98.11	100
	DenseNet121	100	100	100	100	100	100
	Hybrid	100	100	100	100	100	100
CHB24	Transformer	98.83	100	98.41	96.87	100	99.59
	DenseNet121	96.51	91.42	95.52	100	94.44	99.82
	Hybrid	100	100	100	100	100	100

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

time-domain features over spectral-spatial features suggests that short-term temporal irregularities are more important than long-term frequency patterns for differentiating events of different sizes. This observation corroborates clinical EEG interpretation, in which seizure onset is more strongly determined by acute temporal features than by sustained spectral changes, supporting the conclusion that the developed feature set is both clinically interpretable and relevant.

In addition to visual and quantitative explainability mechanisms, the proposed model produces brief textual explanations based on the attention factors of the model. In particular, attention weights computed by the Transformer are pooled over feature dimensions and translated to hand-crafted EEG frequency bands (i.e., delta, theta, alpha, beta, gamma). The relative contribution of individual bands is converted into percentages of the total attention, and the most dominant band is interpreted in a human comprehensible form (such as a strong influence on the theta band). These textual explanations are lightweight and automatically generated, aimed at giving clinicians attention visualizations and feature attributions based on SHAP. The quantitative explanation is formulated as scores computed from the transformer's attention weights and SHAP values. These values quantify the features that are important for seizure detection.

The SHAP technique is applied globally to the ANN classifier to assess how accurately its explanations represent the model's decision process. The evaluation examines whether the features identified by SHAP actually influence the model's decision. It was observed that the classification accuracy dropped significantly, from 99.14% to 58.20%, after removing the top 40% of features with the highest attribution scores, as identified by the SHAP XAI technique. The Perturbed Accuracy of 58.20% indicates that the model relies heavily on the removed features for seizure detection. Therefore, in our case, SHAP is faithful in analyzing feature importance and depicting the model's behavior. Table 5 illustrates that SHAP achieves a Faithfulness score of 40.94% and a Completeness score of 1.00.

D. Subject-Specific Performance Evaluation

The model's performance was evaluated on a single dataset across two settings: aggregated data across all subjects and subject-specific data. It is essential to assess the model's performance on individual subjects, as seizure morphology, frequency patterns, and EEG characteristics vary substantially across individual subjects. While the proposed model was trained and evaluated on EEG recordings from 21 patients in the CHB-MIT dataset, Table 6 reports subject-wise performance for a subset of patients. We evaluated the individual models, Transformer, DenseNet121, and the Hybrid model, combining features from both models across different subjects. The results demonstrate that the proposed model can adapt to subject-specific variations in seizure patterns. The remaining subjects contribute to the overall performance metrics and exhibited comparable classification trends. Although the proposed hybrid model consistently performs competitively across subjects, its effectiveness is influenced by differences in individual EEG characteristics and data distributions. Further, the findings from the experimental evaluation are interpreted in the following section.

IV. Discussion

The proposed HTD-MXAI model aims to improve the performance of Deep Learning-based epileptic seizure detection while providing interpretability for the model's detection. The experimental evaluation reveals that the hybrid approach using Transformer and DenseNet1 21 for automated feature extraction, along with handcrafted features, improves the model's performance in detecting seizures. Initial experimentation guides the selection of suitable DL models to achieve optimal performance. The results reveal that the components of HTD-MXAI ensure

optimal, robust performance in detecting epileptic seizures. The significance of integrating these components is discussed here, highlighting their complementary roles in improving model performance

Table 7. Comparative Performance of Transfer Learning-based Epileptic Seizure Detection Work on CHB-MIT Dataset

Authors	Year	Approach	Performance (%)		
			Accuracy	Sensitivity	Specificity
A. A. Ein Shoka, et al., [14]	2023	Alexnet, Darknet19, GoogLenet, ResNet50, SqueezeNet	86.11	88.89	-
S. Pattnaik, et al., [15]	2024	ResNet50	95.23	99.54	90.28
Proposed	2025	Hybrid DenseNet121-Transformer-ANN	99.14	98.49	99.68

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i2.1380

Table 8. **Comparative Performance of Transformer-based Approaches for Epileptic Seizure Detection on CHB-MIT Dataset**

Author	Year	Approach	Performance (%)		
			Accuracy	Sensitivity	Specificity
N. Ke., et al., [21]	2022	Convolutional Transformer	97.56	96.02	97.94
S. Rukhsar, et al., [22]	2023	Lightweight Convolution Transformer	96.31	96.82	-
Y. Ru, et al., [23]	2024	CNN + Transformer Self-attention	92.89	96.17	92.99
Proposed	2025	Hybrid -Transformer-DenseNet121-ANN	99.14	98.49	99.68

and interpretability. The performance of the proposed HTD-MXAI is compared with several existing Epileptic Seizure Detection research studies [2, 14, 15, 21, 22, 23, 24, 25, 26, 29, 37, 38, 41], including CNN-based pre-trained models, transformer models, hybrid models, and XAI-based models**.** All benchmark comparisons are reported using aggregated performance metrics across the evaluated subjects to enable a fair, high-level comparison with previously published CHB-MIT-based studies.

Initially, the study employs a transfer-learning approach using CNN-based pre-trained models, such as MobileNetV2, EfficientNet-B0, and DenseNet121, to extract features from EEG data. These models are trained on a large ImageNet dataset. Using a transfer-learning approach, these models are fine-tuned on EEG data to extract features. Instead of training a model from scratch, the learned parameters on ImageNet are transferred to these models. It takes less time for training and can be used with scarce labelled data. This results in reduced computational cost and improved generalization on limited data. As demonstrated in Table 3 of the results section, DenseNet121 outperforms MobileNetV2 and EfficientNet-B0. It shows that MobileNetV2 and EfficientNet-B0 completely ignore the normal class. The 100% sensitivity indicates that these models cannot distinguish between non-seizure patterns and predict all samples as seizures, resulting in poor real-world performance on the CHB-MIT dataset for seizure detection. In contrast, DenseNet121 demonstrated balanced performance and effectively classified seizure samples. Hence, we selected DenseNet121 over MobileNetV2 and EfficientNet-B0 to capture deep spatial features from EEG data.

Table 7 presents a comparative analysis of the proposed HTD-MXAI model with existing studies that employ a transfer-learning-based approach for Epileptic Seizure Detection on the CHB-MIT dataset. The study reported by A. A. Ein Shoka et al. [14] evaluated multiple CNN-based pretrained models, such as AlexNet, Darknet19, GoogLeNet, ResNet50, and SqueezeNet. GoogLeNet outperformed other models, achieving moderate accuracy and sensitivity, thus indicating their limited ability to fully capture the complex seizure patterns. The recentstudy by S. Pattnaik et al. [15] employs transfer learning with a pre-trained ResNet-50 model for seizure classification, using 2D scalograms as input. The study reports improved Sensitivity; however, specificity remains relatively low, suggesting potential false-positive detections. In contrast, the proposed hybrid Transformer-DenseNet121-ANN approach achieves the highest accuracy of 99.14% and specificity of 99.68%, demonstrating a more balanced performance. It suggests that combining dense spatial features with Transformer-based temporal modeling is more effective than standalone pretrained models in capturing both local and long-range dependencies via transfer learning.

The performance of the proposed HTD-MXAI model is compared with recent Transformer-based Epileptic Seizure Detection approaches on the CHB-MIT dataset, as depicted in Table 8. The studies by N. Ke et al. [21] and S. Rukhsar et al. [22] demonstrate that the convolutional transformers have a strong ability to capture the temporal dependencies among the EEG data, but with lower accuracy and sensitivity. Y. Ru et al. [23] combined CNNs with self-attention to improve the temporal modeling, but it resulted in reducing overall accuracy. In contrast, the effectiveness of the proposed hybrid Transformer-DenseNet121-ANN model in the epileptic seizure detection task is demonstrated by achieving the highest accuracy, Sensitivity, and Specificity of 99.14%, 98.49%, and 99.68%, respectively. It indicates that joint modeling of rich spatial representations and long-range temporal dependencies improves seizure detection reliability.

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

Table 9. Comparative Performance of Hybrid Approaches for Epileptic Seizure Detection on CHB-MIT Dataset

Author	Year	Approach	Performance (%)		
			Accuracy	Sensitivity	Specificity
M. K. Alharthi, et al., [2]	2022	CNN-Bi-LSTM-AM	96.87	96.85	-
T. Zhou, et al., [24]	2023	CNN-LSTM	98.79	97	-
J. Xu, et al., [25]	2024	GCN-BiGRU	97.35	98.85	95.83
X. Dong, et al., [26]	2024	TCN-Bi-LSTM	97.09	94.13	97.13
Proposed	2025	Hybrid Transformer-DenseNet121-ANN	99.14	98.49	99.68

Table 9 compares the performance of the proposed HTD-MXAI model with existing hybrid approaches applied for epileptic seizure detection on the CHB-MIT dataset. The prior studies combined CNNs with RNNs or utilized Graph-based learning. M. K. Alharthi et al. [2] applied CNN-BiLSTM with an attention mechanism to model the temporal dynamics of EEG and reported promising accuracy and sensitivity. T. Zhou et al. [24] employed a CNN with an LSTM to leverage sequential modeling and reported competitive accuracy and sensitivity. The proposed model shows improved accuracy and sensitivity, which may be attributed to the inclusion of Transformer-based global attention. J. Xu et al. [25] presented a GCN-BiGRU model to capture both spatial and temporal relationships among EEG data. While effective, its reported accuracy and specificity are lower than those of the proposed method. The difference observed in performance may be influenced by the architectural choice, where the study relies on a static graph structure and recurrent temporal modeling. In contrast, the proposed model applied Transformer-based attention to effectively capture long-range dependencies. X. Dong et al. [26] combined temporal convolutions with Bi-LSTM to capture multiscale temporal patterns, resulting in balanced performance. The proposed model surpasses this approach by jointly modeling spatial, temporal, and long-range dependencies. The proposed approach achieves the highest accuracy, sensitivity, and specificity compared with existing hybrid approaches, suggesting that integrating deep feature extraction and Transformer attention improves seizure detection.

As demonstrated in the Results section, the integrated multimodal XAI approach effectively supports the interpretation of the model's decision-making process by providing visual, textual, and quantitative explanations. The study aims to provide a quantitative evaluation of the XAI techniques to better assess their performance and address limitations in existing studies. As illustrated in Table 5, the faithfulness score of 40.94% indicates that removing the feature identified by SHAP as important to the model significantly reduces seizure classification performance, corroborating that the explanations align with the model's actual decision-making behavior. In the context of seizure detection using EEG, such a loss in performance is clinically significant, as it indicates that the model is learning in a physiologically meaningful spectral-temporal manner rather than merely from spurious patterns. The completeness score of 1.00 provides additional evidence for clinical trust, because the total effect of all the SHAP attributions reconstructs the input-output relation of the model, in compliance with the completeness (local accuracy) axiom. Cumulatively, these findings indicate that the XAI framework elucidates sound and transparent rationales suitable for clinical decision support applications.

The robustness of the proposed multimodal explainability framework was empirically verified by qualitatively comparing explanation outputs when the EEG was slightly perturbed, including small changes in amplitude and temporal delay. During these perturbations, we observed that the attention weights were still focusing on similar temporal locations, saliency maps were still highlighting similar discriminative time-frequency patterns, and SHAP values maintained identical feature importance ranks. Such qualitative consistency across diverse explanation modalities suggests that the explanations do not strongly depend on small input perturbations, thereby reinforcing their robustness and supporting their potential use for reliable clinical decision-making.

The visual explanations produced by saliency maps and attention weights were qualitatively examined to assess their consistency with known neurophysiological characteristics of epileptic seizures. During seizure

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

segments, saliency maps tend to emphasize high-energy time-frequency regions in the spectrograms, particularly within frequency bands commonly reported in clinical EEG literature during ictal activity, while attention weights focus on temporally localized segments surrounding seizure onset. Although this assessment is qualitative, the observed alignment with established EEG patterns suggests that the learned explanations are clinically reasonable. SHAP further provides feature-level analysis that supports this alignment by identifying higher-order moments, such as skewness. Together, the attention weights, saliency maps, and SHAP provide complementary explanations across the temporal, spectral, and feature levels.

The performance of the proposed HTD-MXAI model is compared with the existing Explainable AI-based approaches. Table 10 presents a comparison of existing Explainable AI-based Approaches with the proposed HTD-MXAI model for Epileptic Seizure Detection on the CHB-MIT Dataset. Earlier studies, such as those by Mansour et al. [29] and Ding et al.[37], employ feature-relevance analysis and SHAP-based explanations, but report comparatively lower classification performance. Other recent works, including Sánchez-Hernández et al. [38], primarily focus on qualitative interpretability while achieving modest detection accuracy. In contrast, Mazurek et al. [41] incorporate attention mechanisms and GNN-based explanations, but do not explicitly assess the reliability of the generated explanations. The proposed HTD-MXAI model differs in that it combines strong detection performance with both qualitative and quantitative evaluation of explainability, including faithfulness and completeness measures. This comparative analysis suggests that the proposed approach offers a more balanced trade-off between predictive accuracy and interpretability.

Table 10. **Comparative Analysis of XAI-based Epileptic Seizure Detection on CHB-MIT Dataset**

Author/ Year	Performance (Classification Model)					XAI Techniques	Performance Metrics for XAI	
	Acc (%)	Sen (%)	Spec (%)	Precision (%)	F1-Score (%)	Visualization	Qualitative	Quantitative
M. Mansour, et al., 2020, [29]	97.04	97.65	95.58	95.40	-	Feature Relevance Score	Correlation between the Feature relevance	×
Y. Ding et al.,2024, [37]	95.31	92.42	95.32	-	-	SHAP	Optimal Channel combination determination	×
S. E. Sánchez-Hernández, et al.,2024, [38]	84	-	-	-	-	SHAP, LIME	×	Spearman's rank correlation coefficient
S. Mazurek, et al.,2024, [41]	91.32	-	-	-	89.94	Attention Mechanism, GNN Explainer	Interpretation of EI and FI Scores	Edge and Feature Importance Score
Proposed	**99.14**	**98.49**	**99.68**	**99.62**	**99.05**	**Attention Weights, Saliency Maps, SHAP**	**Feature Importance**	**Faithfulness: 40.94% Completeness 1.00**

Note: (‘ - ' indicates that the values are not reported in the reference study.)

The superior performance of the proposed HTD-MXAI model can be attributed to the complementary strengths of its architectural components rather than a single design choice. The Transformer encoder can handle long-range temporal dependencies and global EEG dynamics, and DenseNet121 extracts localized spatial-spectral patterns from time-frequency representations. By combining automatically learned deep features with handcrafted time-frequency and spatial-domain features, the system's discrimination

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

performance is improved by retaining clinically meaningful signal properties that may not be fully captured by deep models alone. Furthermore, the intra-modal multimodality input enables the model to jointly capture temporal, spectral, and spatial cues that contribute to more robust and generalizable seizure representations as compared to unimodal or single-architecture approaches. The proposed model offers multimodal explainability by providing visual, textual, and quantitative explanations of the model's decision-making process.

In addition to aggregated evaluation, subject-specific performance metrics are reported to analyze inter-patient variability and robustness of the proposed model. As presented in Table 6, the proposed hybrid architecture consistently improves or stabilizes performance across most subjects, compared with the individual Transformer and DenseNet121 models. For subjects, e.g., CHB01, CHB03, CHB04, CHB05, CHB06, CHB07, CHB23, and CHB24, the hybrid model achieved perfect to near-perfect performance, suggesting the model’s ability to capture spatial-temporal representation effectively. In contrast, CHB02 and CHB08 are more challenging cases, in which the hybrid model achieved lower accuracy and F1-score than DenseNet121. This indicates that seizure manifestations may be more heterogeneous or subtle in these subjects. These observations highlight the need for subject-level evaluation to better understand the model behavior.

Despite the model demonstrating promising results, the following limitations should be acknowledged. First, its effectiveness has not yet been validated across a sufficiently diverse patient population, which is essential before considering clinical deployment. Second, the post-hoc integration of explainable AI (XAI) methods, while valuable for interpretability, may introduce additional computational overhead and increase system latency. In particular, SHAP-based feature attribution incurs a higher computational cost because the additive nature of Shapley value estimation requires repeated model evaluations to approximate each feature's contribution. Also, attention visualization and saliency map generation require extra forward and backward passes. Although post hoc XAI techniques are applied, they are not executed continuously during real-time inference. They increase analysis latency and memory usage during explanation generation, highlighting a trade-off between interpretability and computational efficiency. Future work will focus on optimizing explanation pipelines and benchmarking inference-explanation latency to assess feasibility in strict real-time clinical settings. Ethical considerations related to XAI are important, considering the clinical implications of epileptic seizure detection. Biases present in the training data, such as patient imbalance, recording conditions, or seizure subtype representation, may influence both model predictions and their explanations, potentially leading to misleading interpretations. While the proposed XAI methods improve transparency, they do not inherently eliminate such biases. Future work should therefore include bias-aware training strategies and diverse clinical datasets to ensure that explanations remain reliable, clinically meaningful, and safe for real-world deployment.

V. Conclusion

The study aimed to develop a multimodal, explainable approach for detecting epileptic seizures. The proposed model leverages advanced deep learning architectures, such as Transformers and DenseNet121, to achieve superior performance compared with existing approaches. The model achieved an overall (aggregated across subjects) accuracy of 99.14%, a sensitivity of 98.49%, and a specificity of 99.68%, outperforming the state-of-the-art models. High sensitivity and specificity are of paramount importance in medical applications because they ensure accurate identification of true clinical conditions while minimizing false diagnoses, thereby supporting safe and effective patient care. The integration of XAI techniques enhances model transparency, offering insight into decision-making through visual, textual, and quantitative explanations. The application of SHAP further enables detailed feature-importance analysis, which may be used for potential feature reduction and refinement. These interpretability mechanisms aim to improve clinical trust and facilitate broader acceptance among healthcare professionals. Future work will focus on real-time diagnosis using large-scale, diverse datasets. Additionally, the research will be extended to develop an explainable predictive model for forecasting epileptic seizures, contributing further to proactive clinical decision support.

Acknowledgment

The authors sincerely thank Dr. Veerendrasinh S. Pawar, Consultant Neurosurgeon at City Hospital, Kolhapur, Maharashtra, India, for his valuable guidance throughout the research and for validating the results. The authors also sincerely thank the Editor and the reviewers for their constructive comments and valuable suggestions. Their insightful feedback has significantly improved the manuscript's clarity, quality, and overall rigor.

Funding

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The dataset used in the study is publicly available. https://physionet.org/content/chbmit/1.0.0/

Author Contribution

Both authors contributed to the design, conceptualization, and writing of the Manuscript. Both authors read and approved the final manuscript.

Declarations

Ethical Approval

Not Applicable.

Consent for Publication Participants.

Consent for publication was given by all participants

Competing Interests

The authors declare no competing interests.

References

[1] G. Alarcon and A. Valentin, *Introduction to Epilepsy*. Cambridge, UK: Cambridge University Press, 2012.

[2] M. K. Alharthi, K. M. Moria, D. M. Alghazzawi, and H. O. Tayeb, "Epileptic disorder detection of seizures using EEG signals," *Sensors*, vol. 22, no. 17, p. 6592, 2022, doi:10.3390/s22176592.

[3] G. Amrani, A. Adadi, M. Berrada, Z. Souirti, and S. Boujraf, "EEG signal analysis using deep learning: A systematic literature review," in *Proc. 5th Int. Conf. Intell. Comput. Data Sci. (ICDS)*, 2021, pp. 1-6.

[4] M. Sazgar and M. G. Young, *Absolute Epilepsy and EEG Rotation Review: Essentials for Trainees*. Berlin, Germany: Springer, 2019.doi:10.1007/978-3-030-03511-2.

[5] A. A. Ein Shoka, M. M. Dessouky, A. El-Sayed, and E. E.-D. Hemdan, “EEG seizure detection: Concepts, techniques, challenges, and future trends,” *Multimed. Tools Appl.*, pp. 1-31, 2023. doi: 10.1007/s11042-023-15052-2.

[6] J. Yuan, X. Ran, K. Liu, C. Yao, Y. Yao, H. Wu, and Q. Liu, “Machine learning applications on neuroimaging for diagnosis and prognosis of epilepsy: A review,” *J. Neurosci. Methods*, vol. 368, p. 109441, 2022. doi: 10.1016/j.jneumeth.2021.109441.

[7] D. D. Spencer, J. L. Gerrard, and H. P. Zaveri, “The roles of surgery and technology in understanding focal epilepsy and its comorbidities,” *Lancet Neurology*, vol. 17, no. 4, pp. 373-382, 2018, doi: 10.1016/S1474-4422(18)30031-0.

[8] M. Sameer and B. Gupta, “CNN-based framework for detection of epileptic seizures,” *Multimedia Tools and Applications,* vol. 81, no. 12, pp. 17057-17070, 2022, doi:10.1007/s11042-022-12702-9

[9] X. Wang, T. Ristaniemi, and F. Cong, “One and two-dimensional convolutional neural networks for seizure detection using EEG signals,” in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO*), Amsterdam, Netherlands, Jan. 2021, pp. 1387-1391,doi:10.23919/Eusipco47968.2020.9287640

[10] S. Das, S. A. Mumu, M. A. H. Akhand, A. Salam, and M. A. S. Kamal, “Epileptic seizure detection from decomposed EEG signal through 1D and 2D feature representation and convolutional neural network,” *Information*, vol. 15, no. 5, p. 256,2024, doi:10.3390/info15050256

[11] B. Zhang, W. Wang, Y. Xiao, S. Xiao, S. Chen, S. Chen, and W. Che, “Cross-subject seizure detection in EEGs using deep transfer learning,” *Comput. Math. Methods Med.*, vol. 2020, p. 7902072, 2020, doi:10.1155/2020/7902072

[12] H. S. Nogay and H. Adeli, “Detection of epileptic seizure using pretrained deep convolutional neural network and transfer learning,” *Eur. Neurol.*, vol. 83, no. 6, pp. 602-614, Jan. 2021. [Online]. Available: https://doi.org/10.1159/000512985

[13] A. Narin, “Detection of focal and non-focal epileptic seizure using continuous wavelet transform-based scalogram images and pre-trained deep neural networks,” *IRBM*, vol. 43, no. 1, pp. 22-31, 2022. [Online]. Available: https://doi.org/10.1016/j.irbm.2020.11.002

[14] A. A. Ein Shoka, M. M. Dessouky, A. El-Sayed, and E. E.-D. Hemdan, “An efficient CNN-based epileptic seizures detection framework using encrypted EEG signals for secure telemedicine applications,” *Alexandria Eng. J.*, vol. 65, pp. 399-412, 2023. [Online]. Available: https://doi.org/10.1016/j.aej.2022.10.014

[15] S. Pattnaik, B. N. Rao, N. K. Rout, et al., “Transfer learning-based epileptic seizure classification using scalogram images of EEG signals,” *Multimed. Tools Appl.,* vol. 83, pp. 84179-84193, 2024. [Online]. Available: https://doi.org/10.1007/s11042-024-19129-4

[16] Zhao, W., Jiang, X., Zhang, B. et al. CTNet: a convolutional transformer network for EEG-based motor imagery classification. Sci Rep 14, 20237 (2024). https://doi.org/10.1038/s41598-

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

024-71118-7

[17] S. Y. Shah, H. Larijani, R. M. Gibson, and D. Liarokapis, “Random neural network-based epileptic seizure episode detection exploiting electroencephalogram signals,” *Sensors*, vol. 22, no. 7, p. 2466, 2022, doi:10.3390/s22072466

[18] M. N. A. Tawhid, S. Siuly, and T. Li, “A convolutional long short-term memory-based neural network for epileptic seizure detection from EEG,” *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1-11, 2022, doi: 10.1109/TIM.2022.3217515

[19] Y. Zhang, S. Yao, R. Yang, X. Liu, W. Qiu, L. Han, et al., “Epileptic seizure detection based on bidirectional gated recurrent unit network,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 135-145, 2022, doi: 10.1109/TNSRE.2022.3142270

[20] I. D. Mienye, T. G. Swart, and G. Obaido, “Recurrent neural networks: A comprehensive review of architectures, variants, and applications,” *Information*, vol. 15, no. 9, Art. no. 517, 2024, doi: 10.3390/info15090517.

[21] N. Ke, T. Lin, Z. Lin, X. Zhou, and T. Ji, “Convolutional transformer networks for epileptic seizure detection,” in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manag. (CIKM '22)*, 2022, pp. 4109-4113, doi: 10.1145/3511808.3557568.

[22] S. Rukhsar and A. K. Tiwari, “Lightweight convolution transformer for cross-patient seizure detection in multi-channel EEG signals,” *Comput. Methods Prog. Biomed.*, vol. 242, p. C, Dec. 2023, doi: 10.1016/j.cmpb.2023.107856.

[23] Y. Ru, G. An, Z. Wei, and H. Chen, “Epilepsy detection based on multi-head self-attention mechanism,” *PLoS One*, vol. 19, no. 6, p. e0305166, 2024, doi: 10.1371/journal.pone.0305166.

[24] T. Zhou, Y. Feng, J. Wang, Y. Tian, J. Feng and J. Li, "Real-Time Epileptic Seizure Detection Based on Deep Learning," *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Sydney, Australia, 2023, pp. 1-4, doi: 10.1109/EMBC40787.2023.10340706.

[25] J. Xu, S. Yuan, J. Shang, J. Wang, K. Yan, and Y. Yang, “Spatiotemporal network based on GCN and BiGRU for seizure detection,” *IEEE J. Biomed. Health Inform.*, 2024, doi: 10.1109/JBHI.2024.3349583.

[26] X. Dong, Y. Wen, D. Ji, S. Yuan, Z. Liu, W. Shang, and W. Zhou, “Epileptic seizure detection with an end-to-end temporal convolutional network and bidirectional long short-term memory model,” *Int. J. Neural Syst.*, vol. 34, no. 3, Art. no. 2450012, 2024, doi:10.1142/S0129065724500126.

[27] E. Tjoa and C. Guan, “A survey on explainable artificial intelligence (XAI): Toward medical XAI,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793-4813, 2021. [Online]. Available: https://doi.org/10.1109/TNNLS.2020.3027314

[28] P. Rathod and S. Naik, “Review on epilepsy detection with explainable artificial intelligence,” in *Proc. 10th Int. Conf. Emerging Trends Eng. Technol. - Signal Inf. Process. (ICETET-SIP-22)*, 2022.

[29] M. Mansour, F. Khnaisser, and H. Partamian, “An explainable model for EEG seizure detection based on connectivity features,” *arXiv Preprint*, arXiv: Learning, 2020, doi:10.48550/arXiv.2009.12566.

[30] X. Zhang, L. Yao, M. Dong, et al., “Adversarial representation learning for robust patient-independent epileptic seizure detection,” *IEEE J. Biomed. Health Inform.*, vol. 24, no. 10, pp. 2852-2859, 2020, doi: 10.1109/JBHI.2020.2971610.

[31] V. Gabeff, T. Teijeiro, M. Zapater, L. Cammoun, S. Rheims, P. Ryvlin, and D. Atienza, “Interpreting deep learning models for epileptic seizure detection on EEG signals,” *Artif. Intell. Med.*, vol. 117, p. 102084, 2021, doi: 10.1016/j.artmed.2021.102084.

[32] D. Raab, A. Theissler, and M. Spiliopoulou, “XAI4EEG: spectral and spatio-temporal explanation of deep learning-based seizure detection in EEG time series,” *Neural Comput. Appl.*, vol. 35, pp. 10051-10068, 2022, doi: 10.1007/s00521-022-07809-x.

[33] P. S. Rathod, J. M. Bhalodiya, and S. Naik, “Epilepsy detection using Bi-LSTM with explainable artificial intelligence,” in *2022 1st IEEE India Conf. (INDICON)*, 2022, pp. 1-6, doi: 10.1109/INDICON56171.2022.10039816.

[34] X. Zhao, N. Yoshida, T. Ueda, H. Sugano, and T. Tanaka, “Epileptic seizure detection by using interpretable machine learning models,” *J. Neural Eng.*, vol. 20, no. 1, p. 015002, 2023, doi: 10.1088/1741-2552/acb089.

[35] I. Al-Hussaini and C. S. Mitchell, “SeizFt: Interpretable machine learning for seizure detection using wearables,” *Bioengineering*, vol. 10, no. 8, p. 918, 2023, doi: 10.3390/bioengineering10080918.

[36] J. C. Vieira, L. A. Guedes, M. R. Santos, and I. Sanchez-Gendriz, “Using explainable artificial intelligence to obtain efficient seizure-detection models based on electroencephalography

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

signals," *Sensors*, vol. 23, no. 24, p. 9871, 2023, doi: 10.3390/s23249871.

[37] Y. Ding and W. Zhao, "Channel selection for seizure detection based on explainable AI with Shapley values," *IEEE Sens. J.*, pp. 1-1, 2024, doi: 10.1109/jsen.2024.3422388.

[38] S. E. Sánchez-Hernández, S. Torres-Ramos, I. Román-Godínez, and R. A. Salido-Ruiz, "Evaluation of the relation between ictal EEG features and XAI explanations," *Brain Sci.*, vol. 14, no. 4, p. 306, 2024, doi: 10.3390/brainsci14040306.

[39] A. Ijaz, Y. Chen, L. Lin, C. Yan, Z. Liu, I. Ullah, M. Shabaz, X. Wang, K. Huang, G. Li, G. Zhao, O. Williams, S. Chen, "An efficient feature selection and explainable classification method for EEG-based epileptic seizure detection," *J. Inf. Secur. Appl.*, 2024, doi: 10.1016/j.jisa.2023.103654.

[40] D. S. Udayantha, K. Weerasinghe, N. Wickramasinghe, A. Abeyratne, K. Wickramasinghe, J. Wanigasinghe, A. D. Silva, and C. U. Edussooriya, "Using explainable AI for EEG-based reduced montage neonatal seizure detection," in *2024 IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, 2024, pp. 463-468.

[41] S. Mazurek, R. Blanco, J. Falcó-Roget, and A. Crimi, "Explainable graph neural networks for EEG classification and seizure detection in epileptic patients," in *2024 IEEE Int. Symp. Biomed. Imaging (ISBI)*, 2024, doi: 10.1109/ISBI56570.2024.10635821.

[42] S. Liu, Y. Zhou, X. Yang, X. Wang, and J. Yin, "A robust automatic epilepsy seizure detection algorithm based on interpretable features and machine learning," *Electronics*, vol. 13, no. 14, p. 2727, 2024, doi: 10.3390/electronics13142727.

[43] P. Rathod, S. Naik, and J. M. Bhalodiya, "Epilepsy detection with CNN and explanation with layer-wise relevance propagation," in *2024 15th Int. Conf. Comput. Commun. Networking Technol. (ICCCNT)*, 2024, pp. 1-6.

[44] F. A. Khan, Z. Umar, A. Jolfaei, et al., "Explainable AI for epileptic seizure detection in Internet of Medical Things," *Digit. Commun. Netw.*, 2024, doi: 10.1016/j.dcan.2024.08.013.

[45] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should I trust you?" Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 1135-1144, doi: 10.1145/2939672.2939778.

[46] A. Shoeb, "CHB-MIT Scalp EEG Database," *physionet.org*, Accessed on 20 December 2024.

[47] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672-2680

[48] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. 31st Int. Conf. Neural Information Processing Systems (NIPS)*, Red Hook, NY, USA, 2017, pp. 6000-6010.

[49] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.

[50] Lin, M., Chen, Q., & Yan, S. (2013). Network In Network. *CoRR, abs/1312.4400* doi:10.48550/arXiv.1312.4400.

[51] S. Sanei and J. A. Chambers, *EEG Signal Processing*. Chichester, U.K.: John Wiley & Sons, 2007.

[52] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[53] L. S. Shapley, "A value for n-person games," in *Contributions to the Theory of Games, Vol. II*, A. W. Tucker and R. L. Luce, Eds. Princeton, NJ, USA: Princeton Univ. Press, 1953, pp. 307-317.

[54] I. Šimić, E. Veas, and V. Sabol, "A comprehensive analysis of perturbation methods in explainable AI feature attribution validation for neural time series classifiers,"*Scientific Reports*, vol. 15, Art. no. 26607, 2025 doi: 10.1038/s41598-025-09538-2, .

[55] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From Local Explanations to Global Understanding with Explainable AI for Trees. *Nature machine intelligence*, *2*(1),56-67. doi:10.1038/s42256-019-0138-9

[56] E. Niedermeyer and F. L. da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 5th ed. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2005.

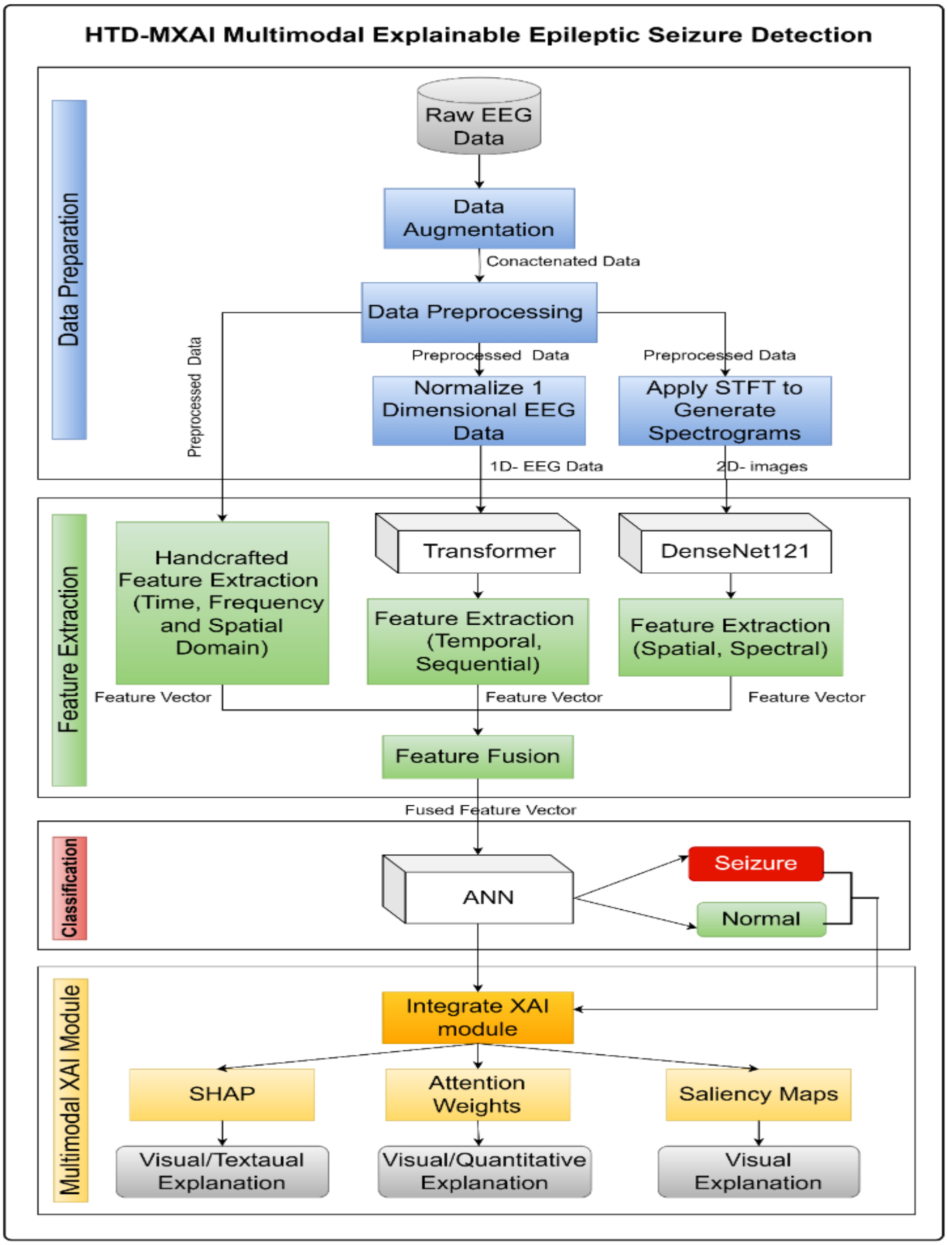
Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380

Author Biography

Ashwini Patil received her B.E. in Computer Science and Engineering from Shivaji University, Kolhapur, Maharashtra. She holds a Master's degree in Computer Engineering from the University of Mumbai and is currently pursuing her Ph.D. in Computer Engineering at the University of Mumbai. She previously served as an Assistant Professor in the Department of Computer Engineering at Thakur College of Engineering and Technology, Mumbai, where she contributed significantly to teaching and academic development. She has over 14 years of teaching experience and has guided many undergraduate students in their research projects. Her research interests include Machine Learning, Deep Learning, Explainable AI, and Data Analytics. She has published numerous papers in reputed international journals and conferences.

Megharani Patil is working as a professor in the Department of Computer Engineering at Thakur College of Engineering and Technology, Mumbai. She holds a Master's degree in Computer Science and Technology from Shivaji University, Kolhapur, Maharashtra. She received her Ph.D. from the University of Mumbai. Her research interests include user experience design and intelligent systems. She has undertaken and completed various consultancy projects. She has published many research papers in National and International journals and conferences. She has served as a reviewer of research articles for international conferences and of edited book chapters. A patent is registered in her name, and she has published 2 books.

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380



Proposed HTD-MXAI Epileptic Seizure Detection Model

Manuscript Received 03 December 2025; Revised 25 January 2026; Accepted 03 February 2026; Available online 07 February 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1380