Received: December 4, 2024. Revised: January 24, 2025.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Multimodality Deep Learning Fusion Based Epileptic Seizure Detection

Vishal Sharad Thakare[1]* **Rekha Ranawat[1]**

[1] *Sage University, India*
*Corresponding author's Email: vishal.thakare@sageuniversity.in

Abstract: Epileptic seizure is a neurological disorder which can create severe consequences when not monitored and given timely care. Various techniques have been proposed for detection of epileptic seizures using electroencephalogram (EEG) signals. But these unimodal approaches have higher false positives. Multimodal approaches combining electrocardiogram (ECG), facial cues etc. with EEG can reduce the false positives. This work proposes a deep learning fusion-based technique for detection of epileptic seizure from multimodal inputs of ECG and EEG. The solution extracts feature from multimodal inputs and applies cross modal learning in spatial and temporal context to increase the accuracy of epileptic seizure detection and reduce false positives. Through experimental analysis with EPILEPSIAE dataset and TUSZ dataset, the proposed solution is found to increase accuracy by at least 1% and reduce false positives by at least 1% compared to recent multimodality solution combining ECG and EEG modality with CNN features and late decision fusion for seizure detection.

Keywords: Epileptic seizure, Neurological, Multimodal, Electroencephalogram, Electrocardiogram, Deep learning, Cross modal learning.

1. Introduction

Epilepsy is a neurological disorder which has the higher risk of death and can create life crippling situations [1]. Though most of epileptic cases respond well to pharmaceutical drugs, 30-40% have drug resilient epilepsy. Epilepsy is often marked by irregular electrical activity in the brain, leading to focal seizures. These seizures are confined to a specific area of the body and may present as unusual sensations, brief lapses in awareness, altered behavior, or confusion, often without visible convulsions. Depending on the abnormal electrical activity in brain and how it spreads, the effects can vary from person to person [2]. Electroencephalography (EEG) is the most popular method for epileptic seizure diagnosis. EEG is the electrical signal measurements from electrodes attached to scalp area. Analysis of these electrical measurements can provide various cues about seizure onset and class of seizures [3]. Manual analysis of EEG is tedious and error prone as the signal is complex, high dimensional and noisy. Towards solving this problem, various machine learning (ML) techniques using both conventional [4] and deep learning [5] schemes have been proposed for automatic diagnosis of epileptic seizures from EEG signals.

Conventional techniques extract various handcraft features in frequency and time domain from EEG signals. The features are then classified to seizure using various machine learning classifiers like support vector machine (SVM), K-nearest neighbor (KNN), Artificial neural networks (ANN), random forest etc. Deep learning techniques avoid handcraft features and learn intricate features through convolutions and pooling. The features are learnt either in 1-dimension or 2-dimension signal representation using various deep learning architectures to classify seizures. The existing ML based techniques have two important issues: (a) for the pre-seizure samples that are further in advance of the onset (one hour), the classification often leads to larger false negatives (low sensitivity) (b) for the non-seizure samples the false positive tends to be larger as it gets closer to the pre-seizure period (low specificity). Multi modal technique integrating various biological time series like EEG, intracranial electroencephalogram (iEEG), or ECG can solve these problems [6]. Multi-modality seizure detection can improve the classifier's robustness by minimizing variance and maximizing overall performance. But there are very few works on multi modal integration especially with consideration for spatial and temporal context. This work addresses this gap and proposes a multimodality deep learning fusion integrating EEG and ECG signals with consideration for spatial and temporal context in both signal dimensions for increasing the accuracy of epileptic seizure prediction and reducing the false positives. The proposed solution has following novel contributions

(i) A novel signal segmentation algorithm based on HRV features of ECG.

(ii) A deep learning network with cross modality learning to extract features from each of EEG and ECG modality and provide enhanced fused feature with minimal adversarial loss.

(iii) The spatial context fused features in different time segments are sequenced to capture temporal context and classified using long short-term memory (LSTM) classifier to predict probability of seizure and non-seizure class. With consideration of both spatial and temporal context, the accuracy of prediction increases and false positives reduces.

The rest of paper contents are structured as follows. Section 2 presents the existing approaches for epileptic seizure prediction. Section 3 presents the proposed multimodality deep learning fusion technique for seizure prediction. The performance comparison results and discussion on results in presented in Section 4. Section 5 presents the conclusion and scope for future research.

2. Literature survey

Wang et al [7] proposed a multi modal approach to classify seizures. EEG signals were processed in 1D or 2D mode with repeated convolutions & pooling along with LSTM to extract spatial and temporal features. The features are then classified using hybrid neural network to seizure classes. Multimodality was applied only in context of feature extraction but the same EEG signal was used which can create bias resulting in higher false positives. Harikumar et al [8] extracted Singular Value Decomposition (SVD) features from EEG signals and classified it to seizures using extreme learning machine. The temporal effect over longer duration is lost in SVD feature extraction and approach is based only on spatial features. Kumar et al. [9] extracted features from EEG signals using combined variable mode decomposition and Hilbert transform. The features are then classified to seizure using stacked neural networks. Temporal context was not considered in feature extraction and signal was processed as whole for feature extraction. Murariu et al. [10] decomposed the EEG signal to intrinsic mode functions (IMF) using empirical mode decomposition (EMD) and extracted Power spectral density features from IMF. The features are classified by KNN and Naïve Bayes classifier. IMF's segmentation using EMD cannot properly segment short duration seizures and hence they cannot be accurately classified. Ficici et al. [11] split the EEG signals to fixed duration epochs. Discrete wavelet transform (DWT) is applied over each epoch to get sub bands and energy features are extracted from each sub band. The energy features for each epoch are classified to seizure class using ensemble classifier. Temporal correlation between the epochs were not considered for seizure classification. Ghazali et al. [12] applied DWT on EEG signals to get sub bands. From these sub bands features related to time domain are extracted. The features are classified using feed forward neural network to seizure class. Temporal feature extraction was not considered. Jana et al. [13] split the EEG signals to two second duration segments and transformed them to a spectrogram matrix. This spectrogram matrix is then classified to seizure class using 1D convolutional neural network (CNN). Without consideration of temporal resolution features the accuracy was less than 80% in this method. Hassan et al [14] proposed a hybrid approach combining deep learning features with traditional machine learning classifiers for seizure detection. CNN features extracted from EEG signals are flattened to 1D feature vector. The best set of features are selected using mutual information entropy. The features are then classified using various traditional classifiers. Segmentation of EEG signal and temporal correlation between features of different segments were not considered in this work. Lih et al. [15] proposed a deep learning transformer architecture called EpilepsyNet to classify seizures. EEG signals were first segmented in 23.6 second duration and spectral entropy features are extracted from each segment. The features are then classified using Bi-LSTM. Though temporal correlation was considered, the segments were on fixed duration and limited handcrafted features are extracted from each segment. Saeizadeh et al. [16] proposed a decision fusion approach combining EEG and ECG modalities. Both modality signals are segmented into 4s duration epochs. For each epoch, convolutional features are extracted and classified using softmax. Both results are then ensembled using logistic regression to

Received: December 4, 2024. Revised: January 24, 2025.

provide seizure class as output. Both modalities worked independently and only their decision was fused. Majzoub et al. [17] et al trained deep learning Alexnet with multichannel EEG to classify seizures. Use of multichannel input avoided information loss but without temporal correlation between signals across duration increases false positives. Christos et al. [18] developed a multi modal approach to detect seizure in home environment combining EEG with accelerometer and gyroscope. The accelerometer and gyroscope readings were used to filter noised EEG signal caused by patient movement. Time/frequency domain features extracted from EEG signal are classified by SVM to seizure. Nielsen et al. [19] proposed a multi modal approach to detect seizure combining EEG, ECG and accelerometry signals. Handcrafted features extracted from each modality were combined as one feature vector and classified to seizure type using SVM. Each modality feature is used individually without any cross-modal learning. Also, temporal correlation was not considered in this work. Vandecasteele et al [20] proposed a multi modal seizure detection approach combining behind-the-ear EEG and ECG signals. Handcraft features are extracted from EEG and ECG signals and classified separately using SVN and RF classifier. The results of each are then late fused to provide the seizure class. The classification accuracy improved due to ECG integration. Qaraqe et al. [21] proposed a multimodal seizure detection technique combining EEG with ECG signals. From ECG, Heart rate variability (HRV) features are extracted and classified to seizure using SVM. EEG signal is split to EEG spectral band. Common spatial pattern features are extracted from each spectral band and classified using SVM classifier. The approach used both feature and decision fusion strategies for seizure detection. Sabor et al [22] combined EEG and ECG modality to detect epileptic seizure onset. The signals are split to 5 second duration segments.

From ECG, HRV features and from EEG, frequency domain features are extracted. Features are enhanced using CNN. The enhanced features are classified by SVM. Each modality was processed separately without any cross-modal learning. Sigsgaard et al. [23] proposed a multi modal approach combining EEG and ECG signals. RF classifier detects seizure from the time/frequency domain features extracted from EEG and ECG signals separately for each modality. The decision of each modality is then late fused to get final decision. Yang et al. [24] proposed a multimodal seizure detection technique combining iEEG and sEEG. Transfer learning is applied to enhance the convolutional features extracted from each modality and decision fusion is done on results of each modality.

Problem definition

Table 1 summarizes the solutions detailed so far. From the Table 1, it can be seen that integration of other modalities to EEG reduces the false alarms and helps to filter noisy signal processing. In most of solutions, the integration was done in early or late fusion mode but signals were processed individually without any cross-modality learning. Also in existing solutions, the temporal correlation over various epochs of signal and segmentation of epochs based on spike characteristics were not considered. Integration of cross modality learning in spatial and temporal context over multi-modality signal is a gap in the existing works. Addressing this gap can minimize the false positives and maximizes the accuracy of seizure detection. Based on this observation, this work proposes a multimodality deep learning fusion technique integrating cross modality learning in spatial and temporal context.

3. Multimodality deep learning fusion

The architecture of the proposed multimodality deep learning fusion technique is given in Fig. 1. The solution has four stages: (i) segmentation, (ii) scaleogram generation, (iii) cross modality learning, and (iv) detection. In the segmentation stage, a common segmentation boundary is established for the joint EEG and ECG signals based on the HRV analysis. 2D scaleogram image is generated on the segmented EEG and ECG signals in the scaleogram generation stage. In the feature enhancement stage, convolutional features are extracted from scaleogram image and enhanced with cross modal learning. In the detection stage, the temporal correlation between the cross modality enhanced features is used to predict the probability of seizures using LSTM. The functional components of the solution are detailed in below subsections. The notations used in subsequent equations are given in Table 3

3.1 Segmentation

Moridani et al. [25] observed that HRV features of RR interval, mean heart rate, low frequency and high frequency components of Poincare plots exhibited significant variations during onset and duration of seizure. But these variations alone cannot be taken as indication of seizure. Based on this observation, this work proposes a novel segmentation algorithm to partition the EEG and ECG signals using HRV feature variations. Different from segmentation

Table 1. Summary of survey

Work	Modality	Technique	Research Gap
Wang et al. [7]	EEG	EEG features are extracted in two modes of 1D and 2D. These features were classified by hybrid neural network to seizures	• Multimodality input was not considered • Classification did not consider temporal correlation • Signals were processed as whole without segmentation
Harikumar et al. [8]	EEG	SVD features are extracted from EEG and classified by Extreme learning machine to seizures	
Kumar et al. [9]	EEG	Features are extracted from EEG using variable mode decomposition and Hilbert transform. Features are classified to seizures with stacked neural networks	
Murariu et al. [10]	EEG	EEG segmented and power spectral features are extracted. Classification using Naïve Bayes classifier	• Multimodality input was not considered • Though segmentation was done on time duration feature correlation between temporal segments were not considered
Ficici et al. [11]	EEG	EEG split to fixed duration segments. Energy features are extracted from each segment and classified as whole by ensemble classifier	
Ghazali et al. [12]	EEG	EEG split to sub band using DWT. Time domain features extracted from each sub band and classified as whole by feed forward neural network	
Jana et al. [13]	EEG	EEG split to 2s duration segments. Spectrogram matrix is created with FFT features of each segment mapped to row in matrix. Matrix is classified by 1D CNN.	
Hassan et al. [14]	EEG	CNN features extracted from EEG signal and flattened to 1D vector. Classification is done by traditional classifier	
Lih et al. [15]	EEG	EEG signal split to 1 second segments. Deep learning transformer was trained to classify EEG segments to seizures	
Majzoub et al. [17]	Multi-channel EEG	AlexNet was trained to classify multichannel EEG features to seizure	
Saeizadeh et al. [16]	ECG + EEG	ECG and EEG signal split to 4s duration segments. CNN features extracted from each of ECG and EEG are classified separately and late decision fused to predict seizure class.	• Though multimodality input was considered, each input was processed separately without any cross-reference learning • Though segmentation was done on time duration feature correlation between temporal segments were not considered
Christos et al. [18]	EEG with accelerometer and gyroscope	Noisy EEG portions filtered using accelerometer and gyroscope readings. EEG is processed as whole	• Segmentation of signal and temporal

Received: December 4, 2024. Revised: January 24, 2025.

		by extracting time/frequency domain features and classified to seizures using SVM	correlation between segments was not considered • Though multimodality input was considered, each input was processed separately without any cross-reference learning
Nielsen et al. [19]	EEG, ECG and accelerometry signals	An aggregated feature vector combining handicraft features of each modality classified using SVM to seizure class.	
Vandecasteele et al. [20]	behind-the-ear EEG and ECG	Decision fusion of classification results of each modality of EEG and ECG to final decision on seizure	
Qaraqe et al. [21]	EEG + ECG	Strategies of early and late fusion to detect seizure	
Sabor et al. [22]	ECG + EEG	The signals are split to 5 second duration segments. Frequency domain features are extracted from EEG and HRV features are extracted from ECG. Features are enhanced using CNN. The enhanced features are classified by SVM to seizure class	
Sigsgaard et al. [23]	ECG + EEG	Features of frequency/time domain are extracted from each modality and classified by RF. Decision is then fused.	
Yang et al. [24]	iEEG + sEEG	convolutional features extracted from each modality and decision fusion is done on results of each modality	

Table 2. Feature difference across solutions

Solutions	**Multimodality**	**Segmentation**	**Temporal correlation**	**Cross modal learning**
[7],[8],[9]	×	×	×	×
[10],[11],[12],[13],[14], [15],[17]	×	✓	×	×
[16]	✓	✓	×	×
[18],[19],[20],[21],[22],[23],[24]	✓	×	×	×
Proposed	✓	✓	✓	✓

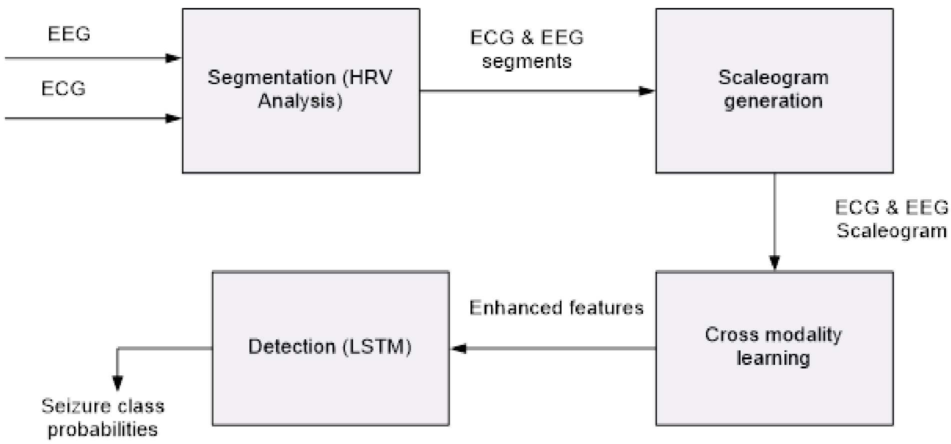

Figure. 1 Proposed Multimodality architecture

 DOI: 10.22266/ijies2025.0331.49

Received: December 4, 2024. Revised: January 24, 2025.

Table 3. Equation notation list

Equation no	Variable	Description
1	w	Frequency
2	α	Scaling factor for low pass filter with value in range of 0 to 1.
2	β	Scaling factor for low pass filter with value in range of 0 to 1.
2	m	Number of samples
3	$\theta(w)$	Daubechies wavelet filter response
5	$O_{11} \dots . O_{1n}$	Softmax classifier outputs for EEG scaleogram
5	$O_{21}, \dots O_{2n,}$	Softmax classifier outputs for ECG scaleogram
5	E_1	Output entropy for EEG scaleogram classification
5	E_2	Output entropy for ECG scaleogram classification
7	p	Prediction label
7	k	Number of output class
8	$\emptyset_t$	Tangent activation function
8	W_c	Weights for input vector
8	U_c	Weights for hidden input vector
12	L	Loss function to be minimized

based on uniform interval, this work adopts segmentation based on characteristic regions boundaries. This reduces the time for detection of seizure and also reduces weightage of non-seizure intervals over LSTM cell states during seizure detection stage.

The ECG signal is processed to detect R peaks using Pan and Tompkins algorithm [26]. A segmentation start is marked at onset of R peak. The RR interval to next immediate R peak is measured. If the RR interval is less than 550-700 ms (milli sec), the process is repeated. If the RR interval is greater than 700 ms, segmentation end is marked and the ECG & EEG signal duration from segment start to segment end is marked as one segment. This process is repeated till the entire signal is segmented. The threshold of 550-700 is fixed based on observation in [25] Fig. 3.

3.2 Scaleogram generation

The ECG and EEG segments are converted to scaleogram image as shown in Fig. 4 by processing with tunable Q-Factor wavelet transform (TQWT). Scaleogram is the plot of energy distribution of a time series signal. The choice of TQWT for scaleogram generation is due to more intricate energy distribution plot for oscillatory signals like EEG and EEG. TQWT is a fully discrete wavelet transform with use of filter banks (high pass or low pass) in sequence. It is represented in terms of high pass filter ($H_1^j(w)$)and low pass filter ($H_0^j(w)$) in Eq. (1) and Eq. (2).

$$H_0^j(w) = \begin{cases} \prod_{m=0}^{j-1} H_0\left(\frac{w}{\alpha^m}\right), \; |w| \le \alpha^j \pi \\ 0, \alpha^j \pi < |w| \le \pi \end{cases} \tag{1}$$

$$H_1^j(w) = \begin{cases} H_0\left(\frac{w}{\alpha^m}\right), \\ H_1\left(\frac{w}{\alpha^{j-1}}\right) \prod_{m=0}^{j-2} (1-\beta)\alpha^{j-1}\pi \le \\ |w| \le \alpha^{j-1}\pi \\ 0, for\ other\ w \in [-\pi, \pi] \end{cases} \tag{2}$$

Where

$$H_0(w) = \theta\left(\frac{w+(\beta-1)\pi}{\alpha+\beta-1}\right) \tag{3}$$

$$H_1(w) = \theta\left(\frac{\alpha\pi - w}{\alpha+\beta-1}\right) \tag{4}$$

In above equations, $\theta(w)$ is the filter response for Daubechies wavelet. The number of decomposition level is given as J. Through experimentation, with various values J is set as 3 in this work. α, β represents the scaling factor of low pass and high pass filter respectively.

3.3 Cross modality learning

The ECG and EEG scaleogram images belonging to same segmentation slot is processed for feature extraction and enhancement using cross modality learning. Since the scaleogram image is an energy plot and there is no difference between ECG or EEG, cross modality fusion becomes easier at scaleogram level. The cross modality enhanced feature extraction is realized using the architecture given in Fig 2. The architecture has repeated convolutions to extract more intricate features. Convolution features at each level are dimension reduced using fully connected (FC) layer and the reduced feature is classified by softmax classifier to seizure class.

Received: December 4, 2024. Revised: January 24, 2025.

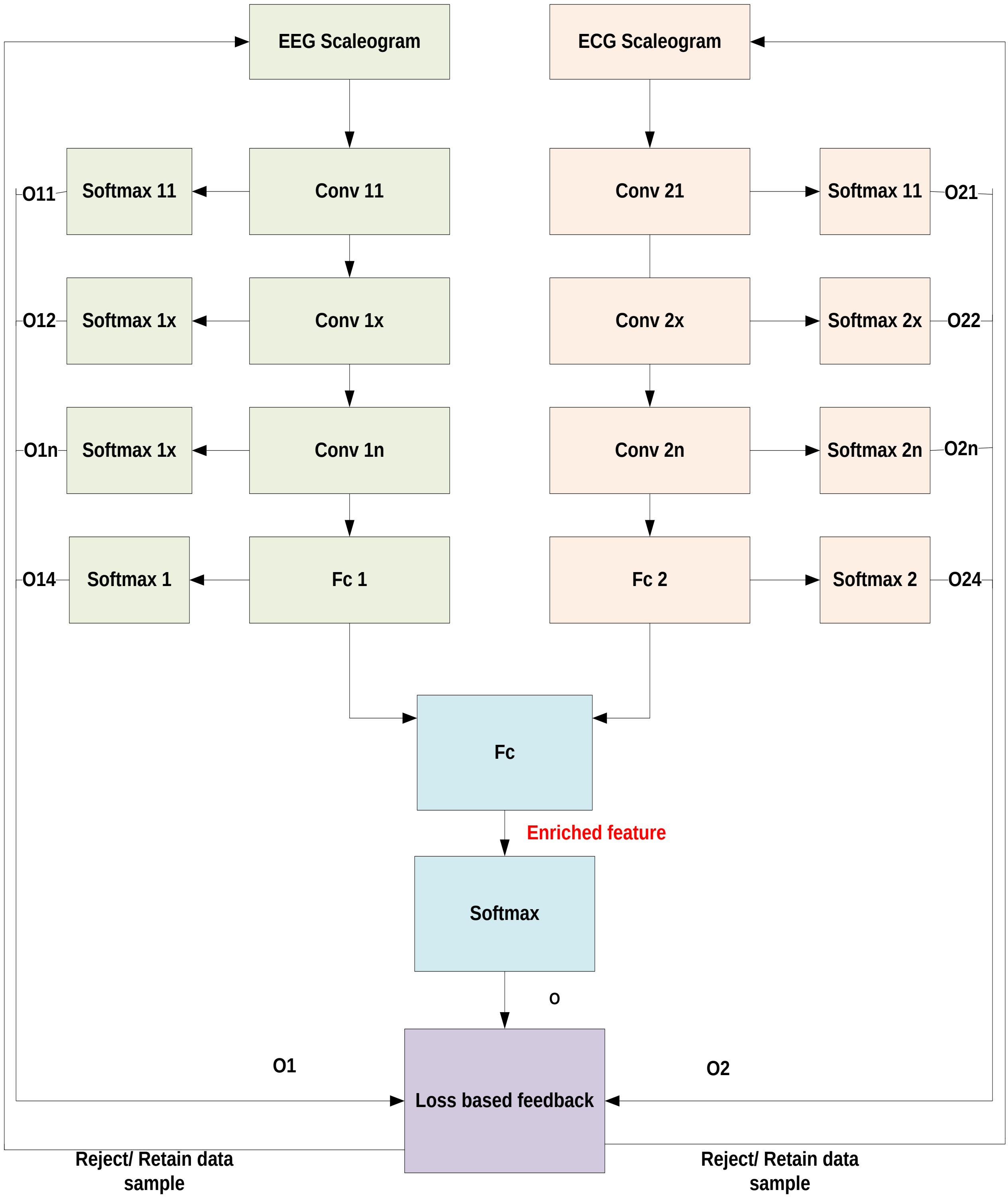

Figure. 2 Cross model learning

The feature without dimension reduction is passed to convolution layer at next level. A consistency test is done by cross checking results of each softmax $O_{11} \ldots . O_{1n}, O_{21}, \ldots O_{2n,}, O_1, O_2, O)$. Majority rule is followed and if more than 90% of the outputs are consistent, the input feature is selected as valid. The valid features are merged as enhanced feature and passed to final stage softmax classifier. The enhanced feature is selected or dropped with criterion of loss minimization. Loss (L) is calculated using Eq. (5).

$$L = \frac{O_{1n}}{O_{1n}+O_{2n}} E_1 + \frac{O_{2n}}{O_{1n}+O_{2n}} E_2 + E \tag{5}$$

Where

Received: December 4, 2024. Revised: January 24, 2025.

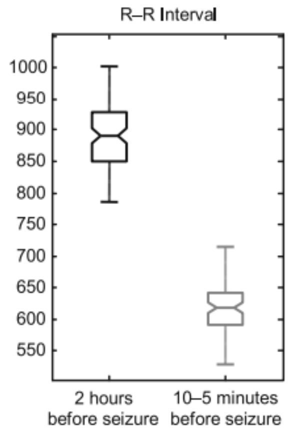

Figure. 3 RR interval observation [25]

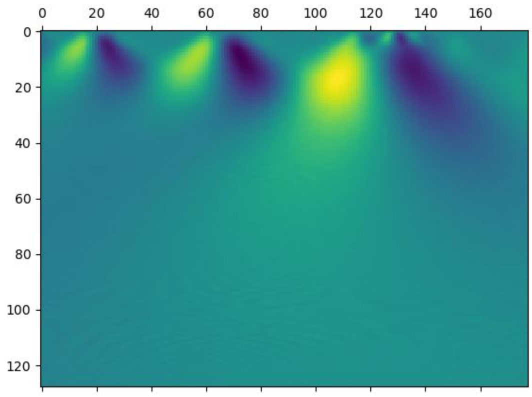

Figure. 4 Scaleogram

$$E_x = \sum_{j=1}^{n-1} O_{xj}.e_{xj} \tag{6}$$

In the above equations, E represents the final output entropy and e_{xj} represents the entropy till the final output, calculated using Eq. (7).

$$e_{xj} = -\sum_{k=1}^{m} p_i^k \log p_i^k \tag{7}$$

In the above Eq. (7), k represents number of output classes. In this work k is 2 for seizure and no seizure class. p represents the prediction label.

The sample set with reduces the loss are added to training set and the sample set increasing the loss are dropped. The cross-model network is then retrained with this updated training set. When the ECG and EEG scaleogram image is passed as input to the cross-model network, the features from the final FC layer are collected as enhanced features. This enhanced feature captures the more intricate details necessary for seizure classification at spatial context level. Learning at temporal context level is done in the detection stage.

3.4 Detection

The enhanced features are sequenced to a length N and this sequence is used to train a multivariate LSTM to predict seizure probabilities. LSTM is the refined version of most used recurrent neural networks (RNN) for series-based prediction. LSTM extends RNN by adding gating mechanism to control the learning rate and forget level. Taking input feature vector and previous hidden states as input, each LSTM cell applies activation function. This cell activation output is given in Eq. (8). Cell activation function applies hyperbolic tangent activation on sum of weighted input ($W_c x_t$), weighted hidden state ($U_c h_{t-1}$) and bias (b_c)

LSTM adds gating mechanism to recurrent neural network. This gating mechanism allows LSTM to retain or forget a level of information. This allows LSTM to control information to be passed to next cell.

$$c_t = \emptyset_t(W_c x_t + U_c h_{t-1} + b_c) \tag{8}$$

Gates control the level of information to be preserved or forgot at the LSTM nodes. Input gates control the preserving factor and forget gates control the forget factor. There is a final gate in LSTM cell to calculate hidden state information for next cell. The resulting gate vector from each of the gates is given below using Eqs. (9) to (11).

$$f_t = \emptyset_s(W_f x_t + U_f h_{t-1} + b_f) \tag{9}$$

$$i_t = \emptyset_s(W_i x_t + U_i h_{t-1} + b_i) \tag{10}$$

$$o_t = \emptyset_s(W_o x_t + U_o h_{t-1} + b_o) \tag{11}$$

f_t is the forgot gate vector. i_t is the input gate vector.o_t is the output gate vector.

Taking the sequence of enhanced features ($Z = (Z_1, Z_2, \ldots Z_T)$), the LSTM is trained to predict the probabilities of two classes: seizure and not seizure. Z_i is the enhanced feature vector and T is the sequence length. The probability of output class is found by adding a softmax classifier at end of last LSTM cell. The softmax classifier operates in regression mode. It is trained to minimize the loss (L).

$$L = -[\sum_{i=1}^{m} \sum_{k=0}^{1} 1\{y^{(i)} = k\} \log P(y^{(i)} = k|z^{(i)}; \theta)] \tag{12}$$

 DOI: 10.22266/ijies2025.0331.49

Where

$$P\left(y^{(i)} = k \middle| z^{(i)}; \theta\right) = \frac{\exp(\theta^{(k)} z^{(i)})}{\sum_{j=1}^{K} \exp\left(\theta^{(k)} z^{(i)}\right)} \tag{13}$$

Where $\theta^{(1)}, \theta^{(2)}, \ldots \theta^{(k)}$ are the parameters of the model and $\exp(\theta^{(k)} z^{(i)})$ is the normalization of parameter with the input feature values.

4. Results

The performance comparison is done by evaluation of the proposed solution against two benchmark EEG+ECG datasets - EPILEPSIAE dataset [27] and Temple University Hospital Seizure Detection Corpus (TUSZ) dataset [28]. Though there are other benchmark datasets like University of Bonn, CHB-MIT datasets, they cannot be used in this work as they have EEG recordings alone but to demonstrate the multimodality and cross modality learning features of the proposed solution, these datasets were not suitable.

Benchmarking with EPILEPSIAE dataset

EPILEPSIAE dataset contains a wide variety of biosignals collected from 275 patients diagnosed with focal epilepsy. The data, gathered between 2009 and 2012 from three esteemed European centers, is characterized by continuous long-term recordings, averaging 165 hours per patient and an average of 9.8 seizures per patient. Among the 275 patients, 29 have non-invasive data comprising both single-channel ECG recordings from the chest and surface EEG data in the 10–20 system. In our study, we utilize all the available data from these 29 patients to ensure a comprehensive analysis of seizure prediction using non-invasive methods. using 5-fold cross validation. In each fold, the dataset is split into 80% for training and 20% for testing, with 10% of the training set further allocated for validation. The performance metrics represent the average efficacy of the model across all folds, specifically on the testing datasets.

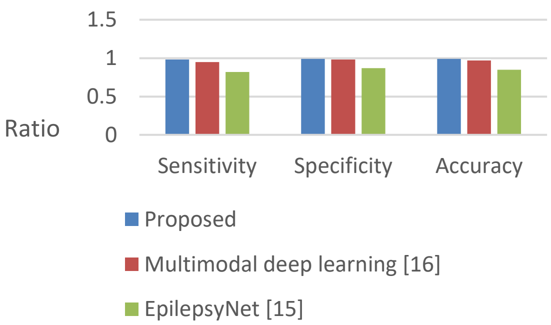

Figure. 5 Performance comparison for EPILEPSIAE dataset

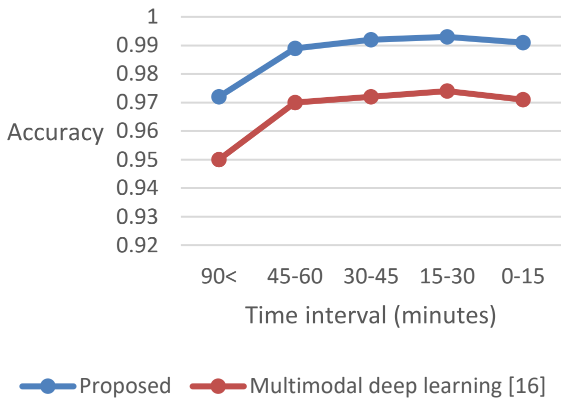

Figure. 6 Accuracy vs seizure onset time

This comprehensive approach ensures a robust assessment of the model's performance across varying data subsets. The performance of the proposed solution is compared against multimodal non-invasive deep learning solution proposed by Saeizadeh et al [16] and EpilepsyNet solution proposed by Lih et al [15]. Saeizadeh et al [16] was selected for comparison as it used both EEG and ECG for seizure prediction similar to the proposed work.

To assess the effectiveness of our model in binary classification, particularly in identifying whether a patient is within 60 minutes of seizure onset, we classify all labels under 60 minutes as pre-seizure and all others as non-seizure. The performance results for seizure prediction across the solutions are given in Fig. 5.

From the results in Fig. 5, it is seen that proposed solution has higher sensitivity, specificity and accuracy compared to multimodal deep learning solution proposed in [16] and EpilepsyNet solution proposed in [15]. The accuracy and specificity are atleast higher by 1% in proposed solution. Lower false positive is inferred from higher specificity values and lower false negatives is inferred from higher sensitivity values. Both sensitivity and sensitivity are higher in the proposed solution. Even though proposed and [16] used ECG and EEG for seizure detection, the proposed solution used cross modal learning to enhance features and temporal correlation to increase accuracy, but the solution in [16] used only late decision fusion. The accuracy of proposed solution is 14% higher and sensitivity is 13% higher compared to [15]. It is because only EEG modality was used in [15]. Thus, the dual modality of ECG and EEG improves the accuracy and sensitivity.

The accuracy trend over seizure onset is measured and plotted in Fig. 6.

Even for earlier onset of 15 minutes, the proposed solution is able to predict with 2% higher accuracy compared to multimodal deep learning solution proposed in [16]. The temporal correlation-based prediction using LSTM has increased the accuracy in proposed solution while [16] did not consider temporal context.

Benchmarking with TUSZ dataset

TUSZ dataset is a open-source EEG/ECG corpus with data of 315 patients. For each patient, 19 channel EEG recordings and one channel ECG recording is present in the dataset. There are 7 seizure types in the dataset. The performance is compared against strategy 3 (with combination of CNN + MLP+ SVM) proposed in BHI-Net [22] and DWT-Net [29] (used in [22] for comparison). The performance is compared in terms of sensitivity, specificity and False Alarm Rate (FAR). The comparison results are given in Table 4.

The proposed solution has higher sensitivity (>3%), higher specificity (>1.5%) and lower FAE (20% less) compared to strategy 3 proposed in BHI-Net. The proposed solution has higher sensitivity (>12%), higher specificity (>6.38%) and lower FAE (26% less) compared to DWT-Net [29].

The false alarm rate has reduced due to cross modal learning and temporal correlation between features of each segment in the proposed solution, but BHI-Net used only late fusion without any temporal correlation between features over segments.

The confusion matrix for the proposed solution with EPILEPSIAE dataset and TSUZ dataset is given in Fig. 7 and Fig. 8.

From the confusion matrix, it can be seen that misclassified instances were very low (0.0183 for TSUZ dataset and 0.0165 for EPILEPSIAE dataset). The misclassification has reduced due to use of multiple functional features of multimodal input, cross modal learning and temporal correlation. The contribution of each of the functional features is analyzed using Ablation study. Ablation study was conducted to find the contribution of each component to the effectiveness of proposed solution. The ablation study was conducted against following cases in Table 5.

The results of accuracy and specificity for the ablation cases are given in Fig. 9 and Fig. 10.

Table 4. Performance comparison for TUSZ dataset

Measure	Proposed	BHI-Net [22]	DWT-Net [29]
Sensitivity	71.34	68.2	59.07
Specificity	96.1	94.6	89.72
FAR (/24h)	9.5	11.9	12

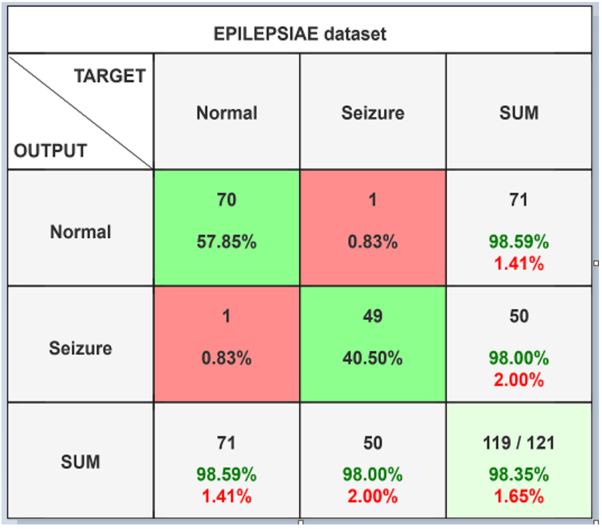

Figure. 7 Confusion matrix for EPILEPSIAE dataset

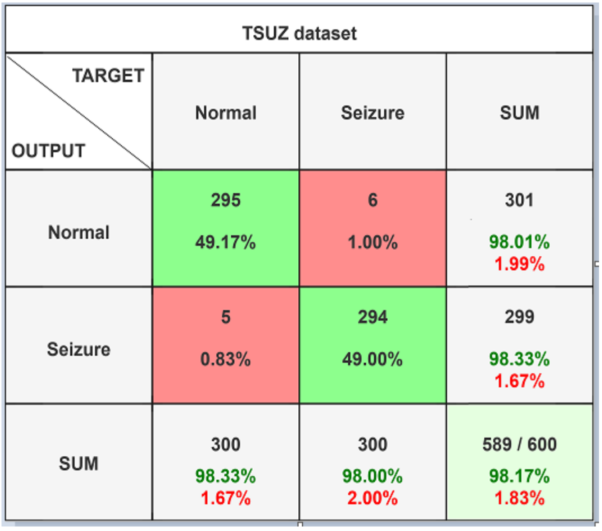

Figure. 8 Confusion matrix for TSUZ dataset

Table 5. Ablation cases

Ablation cases	Description
C1	Proposed solution without cross model learning
C2 (cross modal learning)	Proposed solution without temporal correlation but with cross model learning (realized by skipping LSTM and using softmax for classifying enhanced features)
C3	Proposed solution without temporal correlation and without cross model learning (realized by late fusion of decision on each of the spectrum features of ECG and EEG separately)

From the results of Fig. 7, the proposed solution's accuracy drops 2% without temporal correlation (C1), drops by 2% without cross modal learning (C2) and

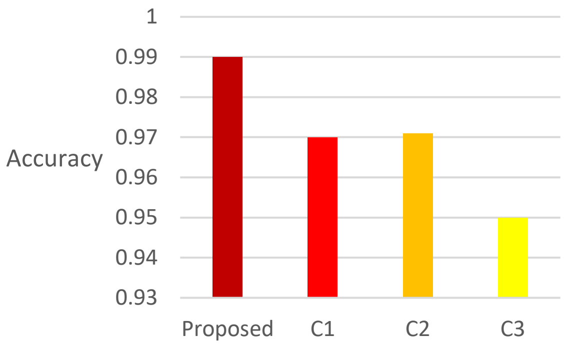

Figure. 9 Accuracy of ablation cases

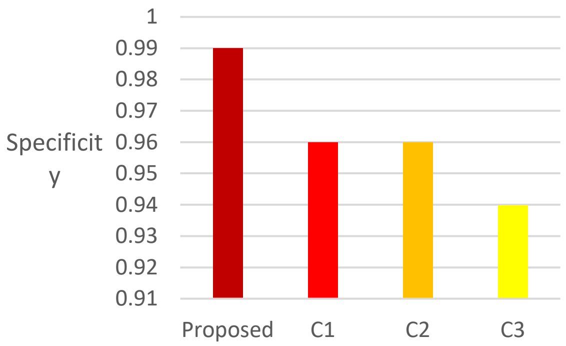

Figure. 10 Specificity of ablation cases

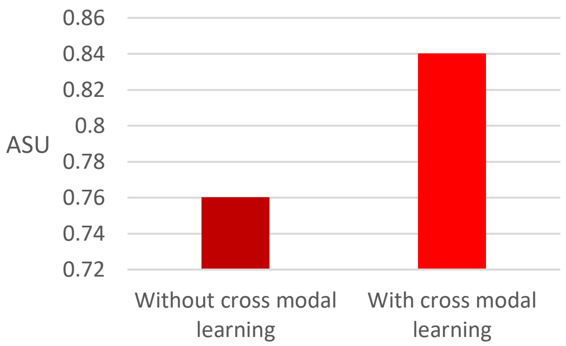

Figure. 11 Results of Average symmetric uncertainty (ASU).

drops by 4% with temporal correlation & cross modal learning (C3). The proposed solution's specificity drops by 3% without temporal correlation (C1), drops by 3% without cross modal learning (C2) and drops by 5% without both cross modal learning & temporal correlation (C3). Thus, cross modal learning and temporal correlation are the two important factors improving the performance of the proposed solution.

The effectiveness of cross modal learning is compared by measuring the average symmetric uncertainty [30] between the features and output class label (normal/seizure) for two cases of feature aggregation without cross modal learning and feature aggregation with cross modal learning. Average symmetric uncertainty (ASU) measures the correlation between the features and the output class. It's value ranges from 0 to 1 and value towards 1 demonstrates higher correlation. The results of ASU are given in Fig. 11.

The ASU is 6% higher with cross modal learning demonstrating its effectiveness.

Discussion

The study identified two important gaps of cross modality learning and temporal correlation between the features. Though many multi modal approaches using EEG and ECG were proposed, these approaches extracted features from each modality and classified seizure using two modes of early and late fusion. In early fusion, features were combined in a single feature vector with correlating the features at spatial level in cross layer manner for improving the feature effectiveness. This is important for reducing the false positives (FAR). The proposed solution solved this problem by correlating the features of each modality at spatial level by bringing them to a common spectrogram form and learning enriched features with cross layer loss measurement feedback. Though segmentation of signal and extraction of feature was considered in existing works, the features are packed to matrix and classified without considering the temporal correlation between them. Considering the temporal correlation improves the accuracy of seizure detection. The proposed solution achieved higher accuracy and reduced false positives by incorporating two important features of cross modal learning and temporal correlation between the sequence of segments.

The proposed solution was implemented using Python 3.8 with Intel®Core™i5-7500 3.4 GHz processor and 16 GB RAM machine. Seizure classification took 18.1 seconds for EPILEPSIAE dataset and 19.3 seconds for TSUZ dataset. When comparted to single modality based only of EEG, the proposed solution has higher latency. Reducing the latency is in scope of future work. Since both datasets used for experimentation did not have noises, noise filtering was not considered in this work. When the solution is applied for real time data, depending on signal acquisition a suitable bandpass filter must be used to filter noise. Designing bandpass filter was not in scope of the work.

5. Conclusion

This work proposes a multimodality deep learning fusion-based technique to detect epileptic seizure. The solution extracted deep learning features from signals by converting them to scaleogram and

Received: December 4, 2024. Revised: January 24, 2025.

improvised the features using cross model learning. The enhanced features are sequenced over temporal duration and classified using LSTM. Through experimental analysis with EPILEPSIAE dataset, the proposed solution is found to have 1% higher accuracy and 1% lower false positives compared to recent multi modal techniques. Through experimental analysis with TUSZ dataset, the proposed solution is found to reduce false alarm by 20% compared to recent multi modal techniques.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, V.S.T. and R.R.; methodology, V.S.T.; software, V.S.T; validation, V.S.T., R.R; formal analysis, V.S.T; investigation, V.S.T; resources, V.S.T; data curation, V.S.T; writing—original draft preparation, V.S.T; writing—review and editing, V.S.T; visualization, V.S.T; supervision, R.R; project administration, V.S.T.

References

[1] W. H. Organization, *Epilepsy*, 2023

[2] X.Ping, W.Hai-Jiao and L. Ling, "Risk factors for drug-resistant epilepsy: A systematic review and meta-analysis", *Medicine*, Vol. 98, No. 30, 2019.

[3] Z. Lasefr, K. Elleithy and M. Faezipour, "An Epileptic Seizure Detection Technique Using EEG Signals with Mobile Application Development", *Appl. Sci*. 2023, Vol. 13, No. 17, pp.9571, 2023.

[4] M.Farooq, A.Zulfiqar and S. Riaz, "Epileptic Seizure Detection Using Machine Learning: Taxonomy, Opportunities, and Challenges", *Diagnostics*,Vol.13,No.6,pp.1058, 2023.

[5] J. Xu, K. Yan and S. Yuan. "EEG-based Epileptic Seizure Detection using Deep Learning Techniques: A Survey", *Neurocomputing*. Vol. 610, 2024.

[6] A. Schulze-Bonhage, "Seizure prediction: Time for new, multimodal and ultra-long-term approaches", *Clin Neurophysiol*,2022.

[7] B. Wang, Y. Xu and F. Li. "Detection Method of Epileptic Seizures Using a Neural Network Model Based on Multimodal Dual-Stream Networks", *Sensors*, Vol. 24, No. 11, 2024.

[8] R. Harikumar, C. Babu and M. Shankar, "Extreme Learning Machine (ELM) based Performance Analysis and Epilepsy Identification from EEG Signals", *Journal of the Institution of Electronics and Telecommunication Engineers*, Vol. 69, No.9, pp. 6304-631, 2021.

[9] G. Kumar, S. Chander and A. Almadhor, "An intelligent epilepsy seizure detection system using adaptive mode decomposition of EEG signals", *Phys. Eng. Sci. Med.*, Vol 45, No. 1, pp.261-272, 2022.

[10] M. Murariu, F. Dorobantu and D. Tarniceriu, "A Novel Automated Empirical Mode Decomposition (EMD) Based Method and Spectral Feature Extraction for Epilepsy EEG Signals Classification", *Electronics* Vol. 12, No. 9, 2023.

[11] C. Ficici, Z. Telatar and O. Erogul, "Automated temporal lobe epilepsy and psychogenic nonepileptic seizure patient discrimination from multichannel EEG recordings using DWT based analysis", *Biomed. Signal Process. Control*, Vol. 77, 2022.

[12] S. Ghazali, M. Alizadeh and Y. Baleghi, "Modified binary salp swarm algorithm in EEG signal classification for epilepsy seizure detection", *Biomed. Signal Process. Control*, Vol. 78, 2022.

[13] G. Jana, R. Sharma and A. Agrawal, "A 1D-CNN-Spectrogram Based Approach for Seizure Detection from EEG Signal", *Procedia Computer Science*. Vol. 167, pp. 403-412, 2020.

[14] F. Hassan, S. Hussain and S. Qaisar, "Epileptic Seizure Detection Using a Hybrid 1D CNN-Machine Learning Approach from EEG Data", *Journal of Healthcare Eng*., 2022.

[15] S.Lih , V.Jahmunah and E.Palmer, "EpilepsyNet: Novel automated detection of epilepsy using transformer model with EEG signals from 121 patient population", *Comput Biol Med.*, 2023.

[16] A. Saeizadeh, D. Schonholtz and J. Neimat, "A Multi-Modal Non-Invasive Deep Learning Framework for Progressive Prediction of Seizures", In: *Proc. of the IEEE 20th International Conference on Body Sensor Networks (BSN)*, 2024.

[17] S. Majzoub, A. Fahmy and F. Sibai, "Epilepsy Detection with Multi-channel EEG Signals Utilizing AlexNet", *Circuits Syst. Signal Process.* Vol. 42, pp. 6780-6797, 2023.

[18] C. Chatzichristos, L. Swinnen and J. Macea "Multimodal detection of typical absence seizures in home environment with wearable electrodes", *Frontiers in Signal Processing*. Vol. 2, 2022.

[19] J.Nielsen, I. Zibrandtsen and P.Masulli, "Towards a wearable multi-modal seizure

Received: December 4, 2024. Revised: January 24, 2025.

detection system in epilepsy: A pilot study", *Clinical Neurophysiology*, Vol. 136, pp. 40-48, 2022.

[20] K.Vandecasteele, T. Cooman and C. Chatzichristos",The power of ECG in multimodal patient-specific seizure monitoring: Added value to an EEG-based detector using limited channels", *Epilepsia*, Vol. 62, No. 10, 2021.

[21] M. Qaraqe, M. Ismail and E. Serpedin "Epileptic seizure onset detection based on EEG and ECG data fusion", *Epilepsy Behav.*, Vol. 58, pp. 48-60, 2016.

[22] N. Sabor N, H. Mohammed, Z. Li and G. Wang, "BHI-Net: Brain-Heart Interaction-Based Deep Architectures for Epileptic Seizures and Firing Location Detection", *IEEE Trans Neural Syst Rehabil Eng.* Vol.30, pp. 1576-1588, 2022.

[23] G. Sigsgaard and Y. Gu, "Comparison of patient non-specific seizure detection using multi-modal signals", *Neuroscience Informatics*, Vol. 4, No. 1, 2024.

[24] Y. Yang, F. Li and J. Luo",Epileptic focus localization using transfer learning on multi-modal EEG", *Front Comput Neurosci*., Vol. 17, 2023.

[25] M.Moridani and H. Farhadi, "Heart rate variability as a biomarker for epilepsy seizure prediction", *Bratislava Medical Journal*, Vol. 118, No.1, pp.3-8, 2017.

[26] J.Pan and W.Tompkins. "A real-time QRS detection algorithm", *IEEE Transactions on Biomedical Engineering*, Vol. 32, No. 3, 1985.

[27] M. Ihle, H. Feldwisch and C. Teixeira, "Epilepsiae-a european epilepsy database", *Computer Methods and Programs in Biomedicine*, Vol. 106, No. 3, pp. 127-138, 2012.

[28] I. Obeid and J. Picone, "The temple university hospital EEG data corpus", *Frontiers Neurosci*., Vol. 10, pp. 1-5, 2016.

[29] Z. Zhang,Y.Ren and N.saboor."DWT-Net: Seizure detection system with structured EEG montage and multiple feature extractor in convolution neural network", *J. Sensors*, Vol. 2020, pp. 1-13,2020.

[30] L. Zhang and X. Chen, "Feature Selection Methods Based on Symmetric Uncertainty Coefficients and Independent Classification Information", *IEEE Access*, Vol. 9, pp. 13845-13856, 2021.