Received: March 16, 2023. Revised: June 9, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Cross Model Attention based Deep Learning for Multi Modal Epilepsy Detection

Venkat Reddy Kumbam[1]* **A. Viji Amutha Mary[1]**

[1] *Department of Computer Science and Engineering,*
Sathyabama Institute of Science and Technology, India
* Corresponding author's Email: venkatreddyk231@gmail.com

Abstract: Epilepsy is a neurological disorder. It is a problem of abnormality in brain resulting in seizures, unusual behaviours, loss of awareness and sensations. Recently many deep learning-based methods to detect epilepsy from multi lead electroencephalogram (EEG) signals have been proposed. Most of these approaches fused the features extracted in one dimensional or two-dimensional mode from EEG through aggregation without any cross-reference learning. As the result, their performance gain in terms of accuracy is limited and also false positives are higher. Learning the cross-correlation information across multiple leads will solve this problem and there has been no works exploiting cross correlation learning. Existing approaches on usage of multiple modalities are mostly feature fusion or decision fusion approaches without cross correlation learning. This work proposes a cross model attention-based learning to learn enriched features from multiple modalities of EEG with aim to increase the accuracy and reduce the false positives. The performance of the proposed solution was tested against University of Bonn EEG dataset. Due to cross correlation learning from multiple modalities, the accuracy of epilepsy detection in proposed approach has increased to 98% which is atleast 2% higher compared to most recent approaches.

Keywords: Deep learning, Multi modal epilepsy diagnosis, Cross model attention.

1. Introduction

Epilepsy is neurological disorder affecting large number of people worldwide. It is characterized by seizures. Based on the onset of seizures and etiology Epilepsy can be categorized to: Localization-related, Idiopathic, Symptomatic and age specific. Though the source of many of these categories are unknown, in most of these cases, epileptic diseased can live without seizures if properly diagnosed and treated with anti-epileptic drugs [1]. EEG is the most used and the golden standard for diagnosis of epileptic seizures. In a typical process of diagnosis, electrodes are placed over the patient's scale regions and electrical signals are recorded. These electric signals are then analyzed through visual inspection by physician to spot any seizures. But this method is error prone and some seizures can be missed out or diagnosed erroneously. Computer aided diagnosis (CAD) is valuable tool to assist expert in accurate diagnosis. These CAD tools extract various features from the EEG recordings and classify them to seizures using machine learning classifiers. Conventional machine learning techniques for seizure prediction extract handcrafted features from EEG signals and classify these features to seizure classes. Handcrafted features extracted using various methods like principal component analysis (PCA), Fourier transform, wavelet transform are classified using machine learning classifiers like support vector machine (SVM), Hidden Markov model and neural networks. Recently deep learning models are proposed for epileptic seizure prediction. Deep learning models are able to learn intricate features from EEG avoiding the need for handcrafted features. Deep learning features provide higher performance gains compared to conventional machine learning models. Deep learning models used EEG signal as ID or converted the EEG to image through some transformation and used it in 2D mode for feature extraction. Deep learning models like deep belief networks, restricted Boltzmann machines, auto encoders, CNN and recurrent neural network (RNN) extract features from 1D EEG signals. Various approaches have been proposed to convert EEG to 2D representation and process it with deep learning models just like images. To improve the accuracy, various multi modal based deep learning methods have been proposed where information from multiple streams are used instead of relying on features from individual streams alone for increasing the accuracy of seizure detection. The problem in these approaches is that each modality is processed separately at results are ensembled only at last stage without cross reference between the modalities. Due to this, performance gains in multimodal approaches are limited.

This work proposes a cross model attention-based decision level fusion to learn enhanced features through cross reference learning from multiple modalities. As the part of work, scalable hierarchical cross model attention learning is proposed which can be extended for any number of modalities. Features extracted from EEG are represented as high level deep learning features and these features are enhanced with proposed hierarchical cross model attention. The enhanced features are classified with multivariate LSTM to consider temporal correlation between sequences of enhanced features for higher heart disease classification accuracy. Following are the novel contribution of this work.

(i) A scalable hierarchical cross model attention to enrich the multimodality features through cross reference learning.
(ii) Temporal correlation between sequences of enriched features through multivariate LSTM.
(iii) Novel methods to convert EEG signals to deep learning high level representations.

Paper organization is as follows. Section II presents the existing deep learning solutions for epileptic seizure prediction. Section III presents the proposed cross model attention based deep learning model for seizure prediction. Section IV presents the results of proposed solution and comparison to existing works. Section V presents the conclusion and scope for future research.

2. Related Work

Gramacki et al [2] developed a deep learning approach using convolutional neural network (CNN) for detection of seizure episodes. The approach was tested for neonatal seizure in young babies less than 4 weeks old. The EEG signals were time segmented and converted to 2D matrix. This 2D matrix is processed by CNN to two classes of seizure or not. Temporal correlation of signals over a longer time window is not considered in this work. O'shea et al [3] developed a 2D fully connected architecture for classifying seizures from multichannel EEG signals. Convolutional features extracted from each channel EEG is combined as feature map and classified. Author did not consider the spatial correlation between multichannel signals and cross reference learning. Isaev et al [4] proposed attention model to provide weightage to each of multichannel EEG signals in detecting the neonatal seizure. The weightage is given based on the relevance to features extracted from channel to seizure class. Based on the weightage, feature fusion is done and fused features are used for classification. Spatial correlation across channels and temporal correlation over a longer time period is not considered in this work. Avcu et al [5] developed a modified CNN architecture for seizure detection from multichannel EEG signals. Additional drop out layers and batch normalization after every convolutional layers are added in the CNN for increasing the classification performance. Features extracted from each channel are fused before classification. But the cross reference learning between each channel features was not considered. Hossain et al [6] used CNN to extract features from single channel EEG for seizure detection. The EEG signals are processed as 1D signal and convolutional features are extracted from it for seizure detection. Temporal correlation across time window of EEG segments was not considered in this work. Covert et al [7] proposed a temporal graph CNN to extract features from EEG signals for seizure detection. Features extracted are localized and shared over both time and space. But the temporal correlation is only over a smaller time window. Emami et al [8] converted EEG signals to a 2D image by segmenting the EEG signals over fixed time window and creating a matrix from segments. The segments are processed by CNN to detect seizure. The approach was designed only for single channel EEG signals. Achilles et al [9] proposed a multimodal deep learning approach combining both EEG and video camera feed to detect seizures. Convolutional features learnt from each modality is feature fused and used for classification. But cross reference across features was not considered. Park et al [10] used CNN to extract temporal and spatial features from multichannel EEG for seizure detection. 1D convolutional layer were used to extract temporal correlation features from EEG and 2D convolutional layers were used to learn the spatial correlation between multiple channels EEG.

Received: March 16, 2023. Revised: June 9, 2023.

But the temporal correlation is only of short duration. Nejedly et al [11] developed a deep learning classifier for seizure detection from EEG. The CNN hyper parameters were optimized using genetic algorithm. EEG signals were processed in 1D mode without spatial and temporal correlation.Iešmantas et al [12] extracted power spectrum features from the EEG signal and classified it using CNN to detect seizures. The power spectrum features across time segments were not correlated. Segundo et al [13] extracted wavelet features from EEG signals using Fourier and empirical wavelet decomposition. The wavelet features were then classified using CNN to detect seizure. The approach was designed for single channel EEG and it lacked temporal correlation over the time window. Similar to above work, Ankut et al [14] extracted features from EEG using discrete wavelet transform and classified it using convolutional neural network. But temporal correlation was not considered in this work too. Turk et al [15] applied continuous wavelet transform on EEG signals to generate 2D scaleogram image. The image is then classified by CNN to three different classes of seizures. Temporal correlation of images was not considered in this work. Liu et al [16] extracted deep learning features from single channel EEG and classified to two classes of seizure or not. Though CNN was modified for improving the performance, this work did not consider temporal correlation. Tian et al [17] proposed a two stage feature extraction method where initial multi view features are constructed from EEG signal using Fast Fourier transform and wavelet packet decomposition. CNN learns the deep features from the initial multi view features. Rule based classifier is employed to classify the CNN features to seizure. Though the features perform better compared to principal component analysis, temporal correlation even over a short time period is not considered. Ansari et al [18] extracted CNN features from multi channel EEG and classified the features using random forest classifier. But no spatial correlation between multiple leads was considered in this work. Cao et al [19] used stacked CNN to extract deep features from multichannel EEG and classified the features to seizure using extreme learning machine. Authors applied weighted feature fusion to fuse multichannel features. But spatial and temporal correlation of features was not considered in this work. Daoud et al [20] used empirical mode decomposition to extract features from EEG and applied CNN for seizure classification. But no temporal correlation between the EEG segments was considered in this work. Craley et al [21] proposed a hybrid probabilistic graphical model CNN for detecting seizure from multichannel EEG. Clinically relevant features are selected by applying probabilistic graphical model. But the method lacked cross reference learning between features across the channels. Ullah et al [22] proposed a ensemble of pyramid 1D CNN for seizure detection. Compared to other CNN models, this model requires only limited dataset. Though spatial correlation across EEG signal is considered, temporal correlation is not considered. Acharya et al [23] used a 13 layer deep convolutional neural network to increase accuracy of seizure detection. Page et al [24] improved the performance of max pooling CNN using transfer learning to increase the accuracy of seizure detection. Yao et al [25] extracted features in different time scales considering temporal and spatial context using recurrent neural network. But the approach was designed only for single channel EEG. Wei et al [26] proposed two novel improvements to CNN to increase the accuracy of seizure detection in terms of data sequences and data augmentation. Fukumori et al [27] used recurrent neural network to detect epileptic spikes in EEG signals. Though the work considered temporal correlation using recurrent neural network, it is designed only for single channel EEG. Lin et al [28] used deep ConvNet to detect seizures from EEG. Author increased the performance using data augmentation, but without temporal and spatial correlation, the accuracy is low. Hussein et al [29] used long short term memory (LSTM) to learn high level representation features from EEG. These features are then classified using a fully connected network. Though use of LSTM brings temporal context into features, the sequence duration is low and the approach is designed only for single channel EEG. Geng et al [30] combined Stockwell transform (S-transform) and bidirectional LSTM to detect seizures from EEG signal. S-transform is applied on EEG signal and the time frequency blocks are grouped. These grouped features are classified by Bi-LSTM. But the approach lacks temporal and spatial correlation of signals. Llias et al [33] proposed a multimodal deep neural network for epilepsy detection. Short Fourier transform was applied to single channel EEG signals and an image is created from three channels output. The image is then classified to epilepsy using pre-trained EfficientNet-B7 model. The approach did not consider the temporal correlation in each channel and designed to work only for short duration EEG signals.

The summary of the survey is presented in Table 1.

Table 1. Survey summary

Works	Gap
Gramacki et al [2]	Temporal correlation of signals over a longer time window was not considered in this work
O'shea et al [3]	Author did not consider the spatial correlation between multichannel signals and cross reference learning
Isaev et al [4]	Spatial correlation across channels and temporal correlation over a longer time period is not considered in this work
Avcu et al [5]	cross reference learning between each channel features was not considered
Hossain et al [6]	Temporal correlation across time window of EEG segments was not considered in this work
Covert et al [7]	Temporal correlation is only over a smaller time window
Emami et al [8]	The approach was designed only for single channel EEG signals
Achilles et al [9]	cross reference across features was not considered
Park et al [10]	Temporal correlation is only of short duration
Nejedly et al [11]	EEG signals were processed in 1D mode without spatial and temporal correlation
Iešmantas et al [12]	The power spectrum features across time segments were not correlated
Segundo et al [13]	It lacked temporal correlation over the time window
Ankut et al [14]	Temporal correlation was not considered in this work
Turk et al [15]	Temporal correlation of images was not considered
Liu et al [16]	Temporal correlation was not considered
Tian et al [17]	Temporal correlation even over a short time period is not considered
Ansari et al [18]	No spatial correlation between multiple leads was considered in this work
Cao et al [19]	Spatial and temporal correlation of features was not considered in this work.
Daoud et al [20]	No temporal correlation between the EEG segments was considered in this work
Craley et al [21]	Method lacked cross reference learning between features across the channels.
Ullah et al [22]	Though spatial correlation across EEG signal is considered, temporal correlation is not considered
Fukumori et al [27]	It is designed only for single channel EEG
Lin et al [28]	Without temporal and spatial correlation, the accuracy is low
Hussein et al [29]	The sequence duration is low and the approach is designed only for single channel EEG
Geng et al [30]	Approach lacks temporal and spatial correlation of signals
Llias et al [31]	Approach did not consider long temporal correlation in each channel EEG signals.

From the survey, most of the solutions were based on single modality of EEG features either in 1D mode or 2D mode by converting EEG to image. Very few works processed multichannel EEG, but features extracted from each channel are just fused without any cross reference learning based feature enrichment. Though spatial correlation is considered, temporal correlation over longer time window is considered in any of the works. But considering temporal correlation over longer time window and cross reference based feature enrichment can increase the accuracy of detection and reduce false positives. The solution proposed in this work is based on this observation.

3. Multi modal cross model attention networks

The architecture of proposed multi modal cross model attention networks for epilepsy detection is given in Fig. 1. The EEG features are extracted from multiple channels and in two modes of 1D and 2D. In the 1D mode, the EEG segment is processed by 1D CNN to extract features. In the 2D mode, the segment is converted to 2D spectrogram by applying continuous wavelet transform. The features from each of the mode from multiple channels are passed to feature enrichment instead of usual aggregation based feature fusion in existing works. The enriched features are passed as input to multivariate LSTM to classify seizures considering temporal correlation over longer duration. The proposed solution has following important stages: feature extraction, feature enhancement and classification. Each of the stages is detailed in below subsections.

A. Feature extraction

The features are extracted from multi channel EEG in two modes: 1D and 2D.

The EEG signals are segmented into fixed duration samples. In the 1D mode, each segment is processed with Tunable Q wavelet transform(TQWT). TQWT is fully discrete wavelet transform noted for feature extraction from oscillatory data like EEG signals. It is a sequence of filter bank. The low pass output from each filter bank is passed as input to successive high pass filter in the next filter bank. Each filter bank has a low pass filter $H_0^j(w)$ and high pass filter $H_1^j(w)$. The filters are defined as

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

$$H_1^j(w) = \begin{cases} H_1\left(\frac{w}{\alpha^{j-1}}\right) \times \\ \prod_{m=0}^{j-2} H_0\left(\frac{w}{\alpha^m}\right), (1-\beta)\alpha^{j-1}\pi \leq |w| \leq \alpha^{j-1}\pi \end{cases} \quad (2)$$

Where

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

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

$\theta(w)$ is the Daubechies filter frequency response.

J is the number of decomposition levels and it is set as 3 for this work

α is the low pass scaling factor and it is set as 0.6 for this work

β is the high pass scaling factor and it is set as 1 for this work

The coefficients are arranged in row major form and passed to a 1D CNN. The 1D CNN can extract the effective and representative features of 1D time-series sequence data through performing 1D convolution operations using multiple filters. The convolutional filters and feature maps of the 1D CNN are all one-dimensional, thus it can match the one-dimensional characteristic of the TQWT processed EEG signal. Bydeepening the number of convolutional layers, the CNN can gradually extract higher-level features which are robust and discriminative for the epileptic seizure recognition tasks. The configuration of 1D CNN used in this work for feature extraction is given in Table 2.

The output of the 1D CNN is a 1×1024 vector.

In the 2D mode, the EEG segment is converted to a 2D scaleogram image using wavelet transform. which helps analysis of signals with dynamic

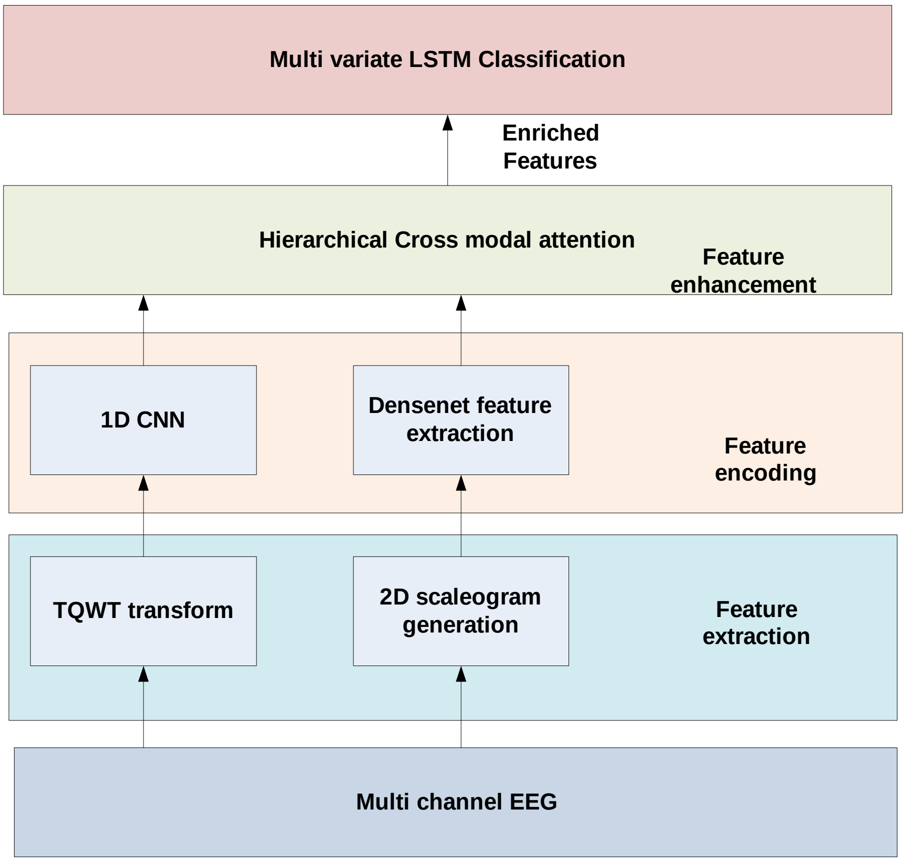

Figure. 1 Multi modality cross model attention architecture

Received: March 16, 2023. Revised: June 9, 2023.

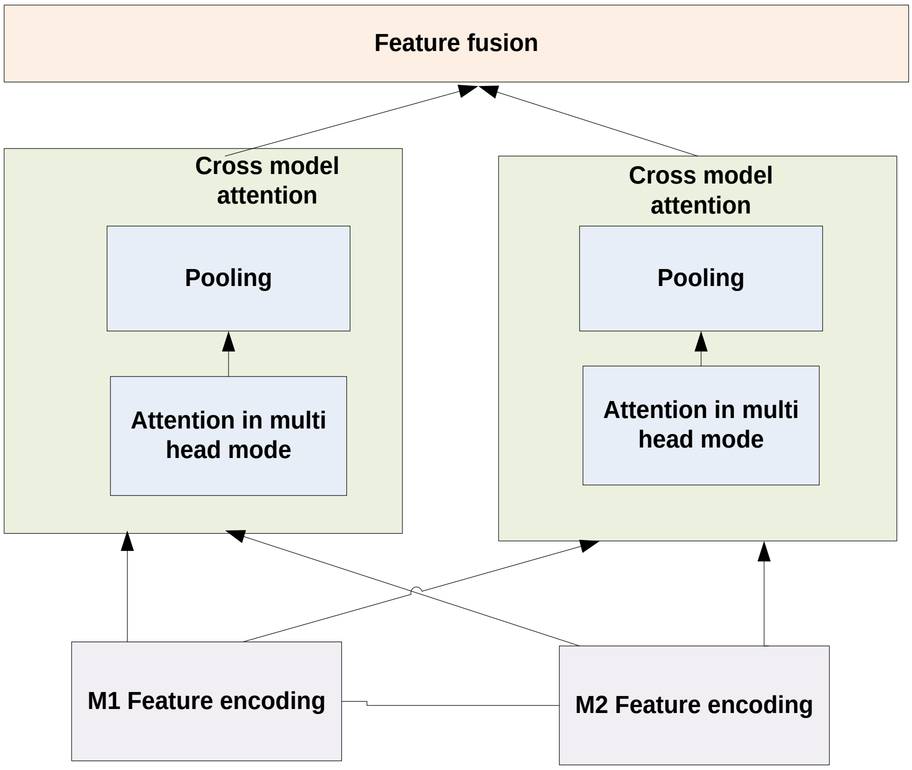

Figure. 2 Cross reference learning

Table 2. 1D CNN configuration

Layer	Parameter
L1:1D convolution	Input size:(640,2) Filter:32 Stride:1 Activation: ReLU
L2:1D convolution	Input size:(640,2) Filter:32 Stride:1 Activation: ReLU
L3:1D convolution	Input size:(640,2) Filter:32 Stride:1 Activation: ReLU
L4:Average pooling 1D	Input size:(616,2)
L5:1D convolution	Input size:(308,32) Filter:32 Stride:6 Activation: ReLU
L6:Flatten	Input size: 1,9696

frequency spectrum. Wavelet transforms are of two categories: continuous and discrete. Continuous wavelet transform is given as

$$X_w(a,b) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} x(t)\varphi\left(\frac{t-b}{a}\right)dt \tag{5}$$

$\varphi(t)$ is the mother wavelet with scale factor of a and translation factor of b. The application of continuous wavelet transform on a EEG signal results in 2D scaleogram which provides the detailed information about the state space of the system. This scaleogram can be used to understand the dynamical behavior of the system and to distinguish different types of signals. In this work continuous wavelet transform with Gaussian as mother wavelet is applied onto the raw EEG signal to generate the 2D scaleogram.

This 2D scaleogram is passed as input to Densenet deep learning model to extract features. Densenet model is used in this work due to capability to learn more intricate features from the images. The configuration of the Densenet model used for feature extraction is given in Fig. 3.

Densenet is developed as a solution to the reduced accuracy in high level neural networks, due to vanishing gradient. In these neural networks, the existence of longer path between the input layer andthe output layer make the information to vanish before reaching its destination. Densenet is an adaptation of Resnet by modifying the additive method with concatenation of previous layer outputs. The goal of Densenet is to increase the depth of network at same time without increasing training time. This is achieved by use of shorter connections between layers. Densenet increases the depth by cross connections and this maximizes information flow in the network. Even in cross connection, feed forward is respected by each layer collecting all

 DOI: 10.22266/ijies2023.1031.07

Received: March 16, 2023. Revised: June 9, 2023.

Figure. 3 Densenet configuration

Layers	Output size	DenseNet-121
Convolution	112×112	
Pooling	56×56	
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	$1 \times 1 \times 128$ conv
	28×28	
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	$1 \times 1 \times 256$ conv
	14×14	
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$
Transition Layer (3)	14×14	$1 \times 1 \times 512$ conv
	7×7	
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$

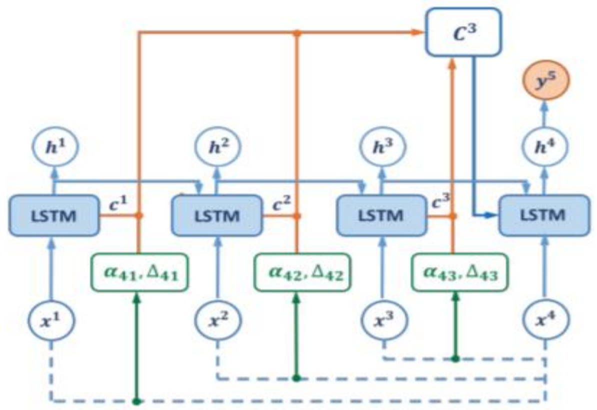

Figure. 4 LSTM Architecture

inputs before passing to next layer.

The 2D scaleogram is converted to feature dimension of size 1×1024 and passed to next stage of feature enhancement.

B. Feature enhancement

The features extracted from 2D scaleogram of different EEG leads are the input for feature enhancement process. Different from earlier works of just fusing the features by aggregating them, this work proposes cross reference learning using cross modal attention. Feature extracted from 1D & 2D mode from each of the EEG channel is a modality. Say there are two modalities $\{m_1, m_2\}$. The cross modal attention for modal m_1 takes output of its feature encoding layer as query vector and the output of m_2 feature encoding layers as key and value vectors. It then applies multi head scaled dot product attention. It helps each modality to learn cross reference information from other modality. Finally features from the cross modality attention of m_1 and m_2 are pooled and passed to prediction using softmax classifier. This cross modal attention proposed in [31] is enhanced using a novel hierarchical cross reference approach for multiple modalities as shown in Fig. 2. In this the cross modalities feature learnt from two modalities $\{m_1, m_2\}$ is passed to next cross modal attention taking output of $\{m_{2+i}\}$ feature encoding layer for learning cross reference information between $\{m_1, m_2\}$ and $\{m_{2+i}\}$. This process is repeated till all *N* modalities are covered.

C. Seizure classification

In most of existing approaches, the features are classified to beats without considering the temporal correlation between features over a time interval. This work solves this problem by using a multivariate LSTM for classification considering temporal variation over a time period.

The architecture of the multivariate LSTM used to predict the emotion classes from the series of feature vectors is shown in Fig. 4.

As shown in Figure, each LSTM node takes the current input vector x and the previous hidden state as input. With this input, it calculates the cell activation as weighted sum of inputs($\mathrm{W_c x_t}$) along with the bias(b_c). The cell activation got as result is then processed with a hyperbolic tangent activation function($\emptyset_\mathrm{t}$) as below

$$c_t = \emptyset_t(W_c x_t + U_c h_{t-1} + b_c) \tag{6}$$

In the above equation, $\mathrm{h_{t-1}}$ is the cell activation result of previous LSTM node in the sequence. The values $\mathrm{W_c}$ and $\mathrm{U_c}$ are the weights for input and the hidden state vector. The level of activation to be retained or forgot is done by controlling the gates.

The hidden state information is calculated at the final state. The gates control how much of activation must be retained and how much must be forgot. Input gate control how must activation to retain and forget gate decided how much cell activation must be forgot. The final gate is incorporated to calculate the hidden state. The final gate takes two information, forgot vector ($\mathrm{f_t}$) and input vector (($\mathrm{i_t}$) as input to provide the output vector ($\mathrm{o_t}$).

$$f_t = \emptyset_s(W_f x_t + U_f h_{t-1} + b_f) \tag{7}$$

$$i_t = \emptyset_s(W_i x_t + U_i h_{t-1} + b_i) \tag{8}$$

$$o_t = \emptyset_s(W_o x_t + U_o h_{t-1} + b_o) \tag{9}$$

Table 3. Equations notations

Variable	Description
$H_0^j(w)$	Low pass filter
$H_1^j(w)$	High pass filter
α	Low pass scaling factor
β	High pass scaling factor
$X_w(a,b)$	Continuous wavelet filter with scale factor a and translation factor b
c_t	Cell activation
$\emptyset_t$	Hyperbolic tangent activation function
W_c	Input weights for neurons
U_c	Hidden state vector
L	Loss function

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

It takes the $Z = (Z_1, Z_2, \ldots Z_T)$, where T sea extent observation are used to predict the rainfall at time T+1 and each Z_i is the input embedding of the transformed original sequence $X = (X_1, X_2, \ldots X_T)$. The final LSTM layer output is passed to a Softmax classifier in regression setting . The output of the softmax classifier is one of two classes: disease or normal. The loss function for training the softmax regression classifier is given as

$$L = -[\sum_{i=1}^{m} \sum_{k=0}^{1} 1\{y^{(i)} = k\} \log P(y^{(i)} = k|z^{(i)}; \theta)] \quad (10)$$

Where

$$P(y^{(i)} = k|z^{(i)}; \theta) = \frac{exp(\theta^{(k)} z^{(i)})}{\sum_{j=1}^{K} exp\ (\theta^{(k)} z^{(j)})} \quad (11)$$

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

The notations used in the equations are summarized in Table 3.

4. Results

The performance of the proposed solution is tested against University of Bonn EEG dataset[32]. The dataset has 4097 EEG signals sampled at 173.61 GHZ with EEG signals labelled into classes of healthy, interictal, ictal, epileptic. An additional experimental dataset was created by reducing the labels to two classes of healthy and epileptic. By this way, performance is conducted for epilepsy detection and specific epilepsy categorization was conducted separately. The dataset is pre-processed by applying bandpass filter in range of 0.53Hz to 40Hz to remove noises. This dataset is used as it is open and suits the need for long term analysis and multichannel analysis.

The performance of the proposed solution is compared against three multimodal deep neural networks proposed by Llias et al [33]. Three networks of ResnetXt-50,VGG16 and EfficientNetB7 proposed in Llias et al [33] were used for comparison.

The performance of seizure detection is measured in terms of accuracy, precision, recall and F1-score.

The results of seizure detection are given in Table 4.

Table 4. Performance comparison

Metrics	Proposed	ResnetXt-50	VGG16	EfficientNetB7
Accuracy	98.84	93.50	97.25	97.50
Precision	98.97	93.98	97.55	98.55
Recall	94.70	93.50	97	96.50
F1-score	95.11	93.49	97.23	90.56

Table 5. Comparison of cross reference learning

Channels	With cross model attention	Without cross modal attention (fusion)
5	97.12	93.92
10	97.27	94.11
15	97.89	94.56
20	98.11	95.78
23	98.34	96.11
Average	**97.74**	**94.89**

The accuracy in proposed solution is atleast 1% higher compared to three deep learning networks proposed by Llias et al [33]. Cross reference learning to enrich features and temporal correlation with multivariate LSTM has increased the accuracy in proposed solution. Though multimodality is considered in LLias et al, use of short term Fourier transform on each EEG channel and constructing image only from Fourier coefficients reduced the temporal correlation. Also cross correlation between features were not considered and the Fourier coefficients of each channel were fused into image. But the proposed solution performed better due to cross model attention between multi-channel EEG to

Table 6. Comparison of temporal correaltion

Channels	With LSTM temporal correlation	Without LSTM
5	97.42	94.92
10	97.87	95.11
15	98.10	95.56
20	98.15	95.81
23	98.34	95.92
Average	**97.97**	**95.46**

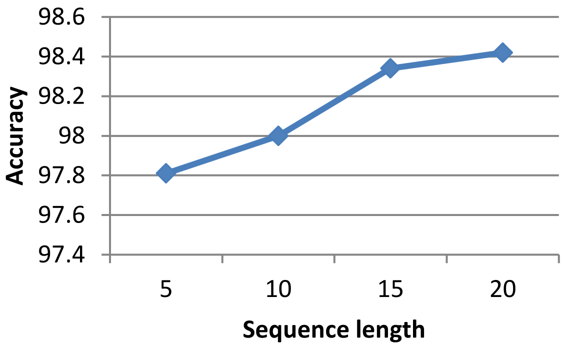

Figure. 5 Accuracy vs sequence length

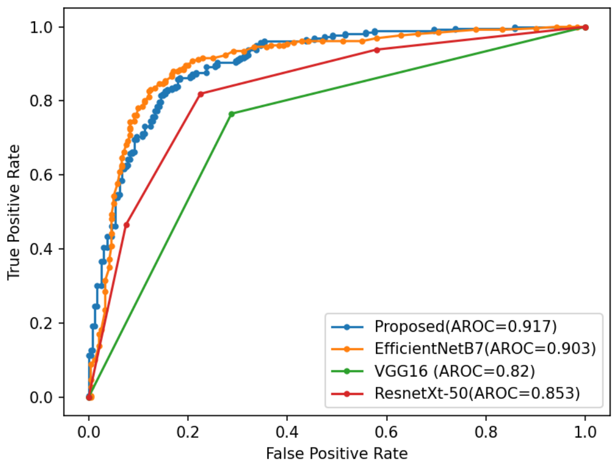

Figure. 6 ROC plot

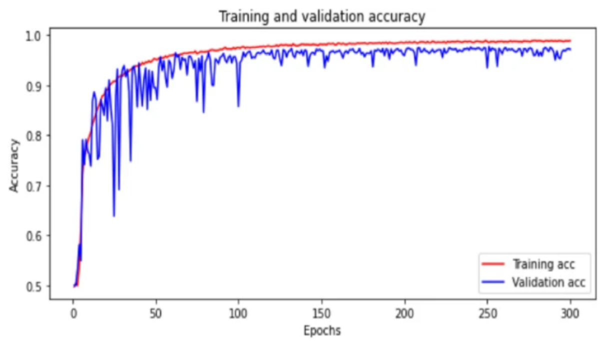

Figure. 7 Accuracy vs epochs

Table 7. Comparison against different seizure types

Seizure types	Proposed	ResnetXt-50	VGG16	EfficientNetB7
interictalseizure	98.44	95.10	94.32	95.12
ictalseizure	98.32	94.89	94.10	95.45
epileptic seizure	98.75	95.30	94.52	95.10

enrich the features.

The results of accuracy of proposed solution due to cross reference learning when compared to aggregation based feature fusion are given in Table 5.

The accuracy increases with processing of features from more channels but use of cross reference learning in proposed solution has increased the average accuracy by 2.85%.

The result of using temporal correlation over a period of time is given in Table 6.

Use of temporal correlation has increased the accuracy by atleast 2.51% compared to without temporal correlation.

The sequence length for LSTM is varied in the proposed solution and the result is given in Fig. 5.

The ROC plot of proposed solution is comparison to existing works is given in Fig. 6.

The ROC is higher in proposed solution compared to existing works. This shows the better sensitivity in proposed solution compared to existing works.

The accuracy plot of proposed solution over various epochs is given in Fig. 7.

The proposed solution achieve peak accuracy of 98% at 150 epochs.

The accuracy of the proposed solution for three different seizure types is measured and result is given in Table 7.

The accuracy in proposed solution is consistently higher compared to all three networks across three epileptic classes. Cross correlation between each channel EEG in proposed solution has increased the accuracy both for epileptic detection and epileptic categorization.

Discussion

Ilias et al [33] multimodality deep learning network is the most recent solution on Epileptic seizure detection. In this solution, short time Fourier transform coefficients were extracted from three channels EEG and they are fused to createa image. The image is then classified to seizure using deep learning networks. CNN was invoked only for extracting more intricate features from fused image. Solution considered only spatial correlation between the multichannel EEG signals and temporal correlation over longer period was considered during Fourier coefficient extraction. But the proposed solution extracted CNN features in the initial stage itself by processing the EEG one dimensional signal and enriching it successive stages. By this way more intricate features were extracted from the EEG signals and this has increased the accuracy in proposed solution. The proposed solution was not compared against uni modality approaches and other traditional machine learning techniques discussed in literature as they were found to underperform compared to multimodality approaches. A thorough discussion on this was presented in the works of Ilias et al [33].

5. Conclusion

A cross reference learning based approach for seizure detection from multi channel EEG is proposed in this work. Two modalities of deep learning features extracted from EEG are used in a cross model learning framework to generate enhanced features. The enhanced features are classified to seizure class using multivariate LSTM. The proposed solution was able to achieve a peak accuracy 98% which is atleast 1% higher compared to most recent existing work.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Venkat Reddy kumbam is the main author conceptualized and developed this article under the guidance of the second author.

References

[1] World Health Organization. Epilepsy, 2021, https://www.who.int/en/news-room/fact-sheets/detail/epilepsy.

[2] A. Gramacki and J. Gramacki, “A deep learning framework for epileptic seizure detection based on neonatal EEG signals”, *Scientific Reports*, Vol. 12, No. 13010, 2022.

[3] A. O. Shea, G. Lightbody, G. Boylan, and A. Temko, “Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture”, *Neural Netw*, Vol. 123, pp. 12–25, 2020.

[4] Y. Isaev, D. Tchapyjnikov, and M. Cotten, “Attention-based network for weak labels in neonatal seizure detection”, In: *Proc. of 5th Machine Learning for Healthcare Conference*, Vol. 126, pp. 479–507, 2020.

[5] T. Avcu, Z. Zhang, and S. Chan, “Seizure detection using least eeg channels by deep convolutional neural network”, In: *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1120–112, 2019.

[6] S. Hossain, U. Amin, M. Alsulaiman, and G. Muhammad, “Applying Deep Learning for Epilepsy Seizure Detection and Brain Mapping Visualization”, *ACM Trans. Multim. Comput. Commun. Appl*, Vol. 15, No. 1, pp. 1–17, 2019.

[7] C. Covert, B. Krishnan, I. Najm, J. Zhan, and M. Shore, “Temporal graph convolutional networks for automatic seizure detection”, In: *Proc. of the Machine Learning for Healthcare Conference*, pp. 160–180, 2021.

[8] A. Emami, N. Kunii, T. Matsuo, and T. Shinozaki, “Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images”, *NeuroImage Clin.*, Vol. 22, No. 101684, 2019.

[9] F. Achilles, F. Tombari, and V. Belagiannis, “Convolutional neural networks for real-time epileptic seizure detection”, *Comput. Methods Biomech. Biomed. Eng. Imaging Vis*, Vol. 6, pp. 264–269, 2016.

[10] C. Park, G. Choi, and J. Kim, “Epileptic seizure detection for multi-channel EEG with deep convolutional neural network”, In: *Proc. International Conference on Electronics, Information, and Communication (ICEIC)*, pp. 1–5, 2018.

[11] P. Nejedly, V. Kremen, and V. Sladky, “Deep-learning for seizure forecasting in canines with epilepsy”, *J. Neural Eng.*, Vol. 16, No. 036031, 2019.

[12] T. Iešmantas and R. Alzbutas, “Convolutional neural network for detection and classification of seizures in clinical data”, *Med. Biol. Eng. Comput*, Vol. 58, pp. 1919–1932, 2020.

[13] R. S. Segundo, M. G. Martín, and F. D. H. Enríquez, “Classification of epileptic EEG recordings using signal transforms and

convolutional neural networks", *Comput. Biol. Med.*, Vol. 109, pp. 148–158, 2019.
[14] R. Akut, "Wavelet based deep learning approach for epilepsy detection", *Health Inf. Sci. Syst.*, Vol. 7, pp.1–9, 2019.
[15] O. Türk and S. Özerdem, "Epilepsy Detection by Using Scalogram Based Convolutional Neural Network from EEG", *Signals. BrainSci.*, Vol. 9, pp. 115, 2019.
[16] J. Liu and B. Woodson, "Deep learning classification for epilepsy detection using a single channel electroencephalography (EEG)", In: *Proc. of the 3rd International Conference on Deep Learning Technologies*, pp. 23–26, 2019.
[17] X. Tian, Z. Deng, and W. Ying, "Deep Multi-View Feature Learning for EEG-Based Epileptic Seizure Detection", *IEEE Trans. Neural Syst. Rehabil. Eng*, Vol. 27, No. 10, pp. 1962–1972, 2019.
[18] H. Ansari, J. Cherian, A. Caicedo, and G. Naulaers, "Neonatal Seizure Detection Using Deep Convolutional Neural Networks", *Int. J. Neural Syst.*, Vol. 29, No. 4, 2019.
[19] J. Cao, J. Zhu, and W. Hu, "Epileptic Signal Classification with Deep EEG Features by Stacked CNNs", *IEEE Trans. Cogn. Dev. Syst.*, Vol. 12, No. 4, pp. 709–722, 2020.
[20] G. Daoud, M. Abdelhameed, and M. Bayoumi, "Automatic epileptic seizure detection based on empirical mode decomposition and deep neural network", In: *Proc. of International Colloquium on Signal Processing & Its Applications (CSPA)*, pp. 182–186, 2018.
[21] J. Craley, E. Johnson, and A. Venkataraman, "Integrating convolutional neural networks and probabilistic graphical modeling for epileptic seizure detection in multichannel EEG", In: *Proc. of International Conference on Information Processing in Medical Imaging. Springer*, pp. 291–303, 2019.
[22] I. Ullah, M. Hussain, and E. Qazi, "An automated system for epilepsy detection using EEG brain signals based on deep learning approach", *Expert Syst. Appl*, Vol. 107, pp. 61–71, 2018.
[23] R, Acharya, Y. Hagiwara, and H. Tan, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals", *Comput. Biol. Med*, Vol. 100, pp. 270–278, 2018.
[24] A. Page, C. Shea, and T. Mohsenin, "Wearable seizure detection using convolutional neural networks with transfer learning", In: *Proc. International Symposium on Circuits and Systems (ISCAS)*, pp. 1086–1089, 2016.
[25] X. Yao, Q. Cheng, and Q. Zhang, "A novel independent RNN approach to classification of seizures against non-seizures", *arXiv*, Vol. 09326, 2019.
[26] Z. Wei, J. Zou, and J. Zhang, "Automatic epileptic EEG detection using convolutional neural network with improvements in time-domain", *Biomed. Signal Process. Control*, Vol. 53, p. 101551, 2019.
[27] K. Fukumori, T. Nguyen, N. Yoshida, and T. Tanaka, "Fully Data-driven Convolutional Filters with Deep Learning Models for Epileptic Spike Detection", In. *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2772–2776, 2019.
[28] L. Lin, S. Ouyang, and R. Wu, "Alternative Diagnosis of Epilepsy in Children without Epileptiform Discharges Using Deep Convolutional Neural Networks", *Int. J. Neural Syst.*, Vol. 30, p. 1850060, 2019.
[29] R. Hussein, H. Palangi, and K. Ward, "Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals", *Clin. Neurophysiol*, Vol. 130, pp. 25–37, 2019.
[30] M. Geng, W. Zhou, and G. Liu, "Epileptic Seizure Detection Based on Stockwell Transform and Bidirectional Long Short-Term Memory", *IEEE Trans. Neural Syst. Rehabil. Eng.*, Vol. 28, pp. 573–580, 2020.
[31] A. Vaswani, N. Shazeer, and N. Parmar, "Attention is all you need", In: *Proc. of Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.
[32] R. Andrzejak, K. Lehnertz, and F. Mormann, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state", *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, Vol. 64, p. 061907, 2002.
[33] L. Ilias, D. Askounis, and J. Psarras, "Multimodal detection of epilepsy with deep neural networks", *Expert Systems with Applications*, Vol. 213, p. 119010, 2022.