RESEARCH PAPER

SEPARABLE RESIDUAL ATTENTION U-NET (SRAUNET) AND STACKED MULTI-SCALE CORRELATION DENOISING AUTOENCODER (SMCDAE) FOR EPILEPTIC SEIZURE DETECTION

Ramya. K[1,*], Dr. M. Kokilamani[2]

[1]Research Scholar, Department of Computer Science, Kamalam College of Arts and Science, Anthiyur, Udumalpet, Affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India.
Email: ramya.msc0709@gmail.com

[2]Assistant Professor, Department of Computer Science, Kamalam College of Arts and Science, Anthiyur, Udumalpet, Affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India.
Email: Kokilamanikcas1@gmail.com

***Corresponding author: Ramya. K, Research Scholar, Department of Computer Science, Kamalam College of Arts and Science, Anthiyur, Udumalpet, Tamil Nadu, India.**
Email: ramya.msc0709@gmail.com

ABSTRACT: EEG - electroencephalogram is the most common way to find seizures since it monitors the brain's electrical activity. Interictal EEG in individuals with epilepsy may exhibit asymmetric spikes or sharp waves, signifying the presence of epileptic activity. Recent methods are clearly time-consuming and tiresome, making proper diagnosis crucial to mitigate the risk of subsequent seizures and seizure-related problems. Because of the residual block and SCBA - Separable Convolutional Block Attention, the residual attention block is added to the U-Net architecture. This paper utilizes the IEWT - Improved Empirical Wavelet Transform on EEG recordings for signal pre-processing. The method works by splitting the spectral domain into flexible bounds, which makes it possible to accurately find the important areas in the Fourier spectrum that together make up the signal's underlying trend component. In a few specific frequency bands, the Separable Residual Attention U-Net - SRAUNet model figure out the properties of EEG signals. The residual attention block was added to the U-Net architecture because of the residual block and SCBA. Stacked Multi-scale Correlation Denoising Autoencoder (SMCDAE), the training process is executed by an encoder and a decoder is introduced for seizure detection. The layers are stacked sequentially, second layers' input is the output of the first layer and so on. This step-by-step learning process helps the system capture strong, noise removed patterns which are closely related to seizure activity. The experimental assessment utilizes the CHB-MIT scalp EEG dataset, a renowned collection produced through collaboration between Boston Children's Hospital and the Massachusetts Institute of Technology, this proposed method has obtained highest recall/sensitivity, F-measure, accuracy and precision than other methods.

INDEX TERMS: Epileptic seizure detection, scalp EEG, Improved Empirical Wavelet Transform (IEWT), Separable Residual Attention U-Net (SRAUNet) model, and Stacked Multi-scale Correlation Denoising Autoencoder (SMCDAE).
How to cite this article: Ramya K, Kokilamani M. Separable Residual Attention U-Net (SRAUNet) and Stacked Multi-scale Correlation Denoising Autoencoder (SMCDAE) for Epileptic Seizure Detection. Int J Drug Deliv Technol. 2026;16(63s):536-584. DOI: 10.25258/ijddt.16.63s.59
Source of support: Nil.
Conflict of interest: None

1. INTRODUCTION

Epileptic seizures happen when the brain reacts in an unusual way again and over again because of the strange electrical signals that neurons send out [1,2]. They cause mood swings, changes in sensation, movement, or mental function [3,4], delays in attention, whole-body tremors, and loss of consciousness. To provide proper treatment and care for epilepsy patients, it is important to accurately identify the type of epileptic seizure they experience. [5]. Clinically, seizures occur due to loss of consciousness, motor convulsions, or subtle alterations in behavior, frequently leading to falls, injuries, or fatalities [6]. There are three varieties of elliptic seizures: focal, generalized, and unknown. These categories depend on when and how the seizure originates in the patient's brain. Focal seizures start in a single area or a small group of cells on one side of the brain. They are further classified into two types: simple partial seizures and complex partial seizures. In contrast, generalized seizures begin in groups of cells on both sides of the brain at the same time. They are divided into myoclonic, clonic, tonic, and tonic–clonic seizures. Seizures that have no recognized cause are called "unknown seizures" [7].

Electroencephalography (EEG) is the most widely used method for diagnosing epilepsy. It works by detecting voltage changes produced by ionic currents in neurons and measuring the difference in electrical potential between electrodes placed on the scalp. It also gives information about the brain's time and space [8]. To find out if someone has a problem with their EEG, a doctor must look directly at the person and observe carefully. Moreover, experts with different diagnostic experience can express conflicting views regarding the diagnostic outcomes [9]. There are now a number of algorithms that can use EEG signals to find seizures in people with epilepsy [8,9]. The unprocessed EEG output frequently contains noise and other information.

Computer-aided detection systems provide automated tools to help identify important EEG features. Consequently, a computer-assisted approach is essential for the diagnosis of epilepsy [10]. A lot of these research use Wavelet Transform (WT) to break down signals and get rid of noise. The goal of using these methods is to lower the number of dimensions in a signal by getting signal representations from diverse frequency bands. This makes it easier for Deep Learning (DL) models to analyse less input. In wavelet transform (WT) denoising, a basis function is used to generate wavelet coefficients that indicate how well the basis matches the original signal. Noise is then reduced by applying a threshold to these coefficients and removing those that fall below it and to rebuild the denoised signal, the rest of the coefficients were used.

Yedurkar and Metkar [12] used a combination of Discrete Wavelet Transform (DWT) and adaptive filtering to suppress low-frequency physiological noise while preserving the clinically important components of the EEG signal. However, a major limitation of wavelet-based denoising is that it requires choosing a fixed basis function, which cannot be adaptively adjusted during processing. Empirical Mode Decomposition (EMD) addresses this issue by decomposing a complex signal into multiple Intrinsic Mode Functions (IMFs), each representing different frequency components. A denoised signal is acquired by setting a threshold to get rid of the bad IMFs and then rebuilding the ones that are left. Empirical Wavelet Transform decomposes a signal into different oscillatory modes, because EWT based signal denoising is an adaptive technique, so it removes noise from the detail coefficients (often representing higher frequencies), and then reconstructs the denoised signal [13]. EWT separates EEG signals into different subbands. These subbands and their coefficients are used to get different features that help classify EEG data associated to seizures [13].

The advancement of deep learning methodologies has introduced novel opportunities for enhancing the identification and prediction of epilepsy [14,15]. Compared to traditional methods, these newer algorithms can process and analyze large volumes of EEG data much more efficiently, enabling neurologists to more accurately diagnose epilepsy and predict potential seizure occurrences. These technologies can help patients take steps to avoid seizures by giving them timely warnings. This helps lessen the physical and mental effects of seizures. Models that use both WT and DL have always been very accurate. The main goal is to get rid of EEG signal’s motion distortion with the help of empirical wavelet transform (EWT) method.
To acquire a good result for EEG epileptic seizures, a feature extraction and network design frequently focus on RNN and CNN (Recurrent Neural Network and Convolutional Neural Networks respectively). RAUNet - Residual Attention U-Net is a DL method which is used for feature extraction and the strengths of U-Net, attention mechanisms and residual connections were combined by other tasks. This DL approach use an encoder-decoder U-Net structure, for capturing the features at different scales, further the residual blocks are combined to increase the ability to learn complex features and enable deeper networks, leading to improved feature representation for tasks.

In this paper, processing of EEG signals concentrates on epilepsy detection and prediction. EEG dataset, preprocessing method, feature extraction, and deep learning as major steps of proposed model. Improved Empirical Wavelet Transform (IEWT) is utilized as the initial pre-processing technique for EEG recordings. IEWT enhances the traditional EWT by introducing a more adaptive and refined mechanism for spectral boundary detection. Feature extraction is performed using a Separable Residual Attention U-Net (SRAUNet) model. The residual blocks facilitate deeper feature propagation and mitigate gradient-vanishing issues, while the attention mechanism adaptively emphasizes the most informative EEG regions and suppresses redundant features. Seizure Detection, Stacked Multi-scale Correlation Denoising Autoencoder (SMCDAE) is constructed by stacking multiple denoising autoencoder layers, where each layer is trained to reconstruct noise-corrupted input while learning robust latent representations which helps to identify seizure activity more accurately.

2. LITERATURE REVIEW

LCT - Lightweight Convolution Transformer is a novel deep learning architecture which is proposed by Rukhsar and Tiwari [16]. The Transformer model is capable of simultaneously learning both spatial and temporal relationships from multi-channel EEG signals. This makes it possible to find seizures even with shorter segment lengths. The proposed study addresses the limitations of the Vision Transformer (ViT)—notably its absence of translation equivariance and inadequate localization—by implementing convolution-based tokenization. Additionally, sequence pooling is utilized to derive comprehensive feature representations from the Transformer encoder, replacing the conventional learnable class token. Extensive experiments demonstrate that the proposed cross-patient learning framework can accurately detect seizures from raw EEG signals. Performance evaluations on the CHB-MIT dataset show that incorporating inductive biases and attention-based pooling enhances overall accuracy while reducing the required number of Transformer encoder layers, thereby significantly lowering the computational complexity of the system. For multi-channel automated seizure detection, this paper presents an optimized architecture that enhances efficiency and performance.

Multiple feature augmentation procedures were given by Pandey et al., [17], to develop a hybrid feature space that can capture how seizures don't follow a straight line. The predictor learns better in this complex feature space, which improves the prediction of seizures. The new hybrid FB-SARO - Forensic-based-Search-and-Rescue Optimization, enhances the predictor that makes the seizure prediction better. A hybrid feature space is built for performance enhancement and then the possible non-linear patterns for seizure classification is identified. The classifier's learning also helps recognizing the best SPH - seizure prediction horizon. The SPH helps make predictions early while still being accurate and keeping the False Prediction Rate (FPR) as low as possible. It also helps sound the alert so that patients have enough time to get ready for medical treatment. Using publicly available datasets, this suggested method is examined and compared its result with the best methods that are already out there.

Zhao et al. [18] introduced HAN - Hybrid Attention Network that can automatically find seizures. GAT - graph attention network gets information about space from the front end, while the Transformer component captures temporal information at the back end, while the attention mechanism in the HAN architecture effectively models the spatiotemporal relationships within EEG signals. Additionally, a focal loss function is incorporated into HAN to address the class imbalance present in the EEG-based seizure detection dataset. The experiments are conducted using the publicly available CHB-MIT database, including both patient-specific and patient-independent evaluation scenarios. The results demonstrate that HAN performs effectively across both experimental settings.

Duan et al. [19] introduced an automated epileptic seizure detection system based on deep metric learning, providing a novel approach to the few-shot learning problem by significantly reducing the reliance on large-scale training datasets. "The method uses two one-dimensional convolutional embedding modules to get deep feature representations from single-channel and multi-channel EEG signals, respectively. Next, we show a deep metric learning model in detail and explain how to train it step by step. The Bonn University dataset, which is a widely used benchmark, is available to the public and the bigger and more clinically relevant CHB-MIT dataset were both employed in experiments. The proposed model is good at finding seizures; it gets roughly 86.68% of them right and 93.71% of them wrong on the CHB-MIT dataset.

Li et al. put out a new channel-embedding spectral-temporal squeeze-and-excitation network called CE-stSENet to help find seizures in EEGs [20], the model includes a loss that maximizes information based on the maximum mean discrepancy. CE-stSENet is the first framework to combine multi-level spectral analysis with multi-scale temporal feature extraction simultaneously. Then, a type of the squeeze-and-excitation block is employed to get hierarchical multi-domain representations in one approach. Lastly, the classification net uses information from preceding subnetworks to tell if an EEG is showing signs of epilepsy. To tackle the problem of infrequent seizures, which limits data distribution and leads to significant overfitting in seizure detection, CE-stSENet uses a maximum mean discrepancy–based information-maximizing loss to reduce overfitting. The proposed framework's ability to identify epileptic EEGs is demonstrated by superior experimental outcomes on three EEG datasets compared to the most effective existing methods. This shows that it can find seizures on its own very well.

Geng et al. [21] suggested a successful automatic seizure detection technique utilizing the Stockwell transform (S-transform) and a bidirectional long short-term memory (BiLSTM) network for intracranial EEG recordings. First, the S-transform is used on the raw EEG segments to make a matrix. After that, the BiLSTM network gets the matrix in time-frequency blocks. It picks out the features and does the classification. Then, postprocessing is applied to improve the detection. This includes employing a moving average filter, making a threshold decision, fusing multiple channels, and the collar technique. The proposed methodology is evaluated utilizing 689 hours of intracranial EEG data from 20 patients. The event-based evaluation has a sensitivity of 96.30% and a false detection rate of 0.24 per hour." These encouraging findings indicate that the seizure detection approach possesses significant potential for clinical application.

Hu et al. [22] presented an innovative seizure detection methodology employing a deep bidirectional long short-term memory (Bi-LSTM) network. They used local mean decomposition (LMD) and statistical feature extraction on the EEG signals to keep the signal from getting stuck and make the calculations easier. Two separate LSTM networks, were put together by making the deep architecture, that send information in opposite directions: one sends it from the front to the rear and the other sends it from the back to the front. This lets the deep model figure out the present output by using knowledge from both the past and the future. A long-term scalp EEG database test demonstrated an average sensitivity of 93.61% and an average specificity of 91.85%. Comparative analyses with existing methodologies—utilizing either convolutional neural networks or conventional machine learning techniques — further validated the enhanced efficacy of the proposed strategy in seizure identification.

Yao et al. [23] proposed a method that integrates an attention mechanism with a Bi-LSTM to capture both spatial and temporal discriminative features, along with the unpredictable characteristics of seizures. The attention mechanism helps in identifying the spatial features based on how different parts of the brain affect seizures. It separates EEG signals from different parts of the brain and gives each channel's data the right amount of attention. To explain the attention weights, examples of EEG data segments were provided. To get temporal characteristics, Bi-LSTM is used which explains the difference between two things in both the forward and backward directions. Cross-validation and cross-patient experiments were conducted on the noisy CHB-MIT dataset to assess the proposed strategy. In average 87.30% of sensitivity, 88.30% of specificity, and 88.29% of precision obtained in cross-validation studies surpass the performance of existing cutting-edge methods.

Khan et al. [24] suggested a trainable hybrid approach that integrates a shallow autoencoder (AE) with a conventional classifier for the detection of epileptic seizures. The AE's encoded representation is employed as a feature vector to tell the difference between EEG signal segments that are epileptic and those that aren't. The method works well on single-channel data and doesn't need a lot of computing power, thus it's good for body sensor networks and wearable devices that just employ one or a few EEG channels for user comfort. This makes it possible to keep an eye on and diagnose people with epilepsy at home for a longer period of time. The shallow autoencoder is trained to minimize the reconstruction error of the original signal in order to get the encoded EEG signal segments. After testing a lot of different classifiers, two versions of the hybrid method were suggested: (a) one that works better than current methods when using the k-Nearest Neighbor (kNN) classifier, and (b) another that is better for hardware while still getting the best classification performance among reported methods in this category when using a Support Vector Machine (SVM) classifier. Researchers tested the approach on EEG data sets from CHB-MIT and the University of Bonn.

Tawhid et al. created a useful framework using a deep spatiotemporal neural network called a convolutional long short-term memory (ConvLSTM) [25], for finding epilepsy in EEG waves. The proposed model initially selects standard 19-channel EEG data and subsequently resamples it at 256 Hz. After then, the signals are broken up into 3-second chunks. Then, the ConvLSTM model uses the segmented data to tell the difference between normal people and people with epilepsy. To avoid the experiment's biases, LOOCV - leave-one-out cross-validation and five-fold cross-validation are applied. The experimental output show that the proposed model does better than the best results so far for the datasets analyzed. This means that it is a good choice for an automated system for diagnosing epilepsy.

A novel hybrid architecture for epileptic seizure prediction was by Quadri et al., [26], which uses a deep learning approach by stacking the CNN and Bi-LSTM layers. The proposed approach employs a series of 1-D convolution layers, each with several filters with lengths varying exponentially. The deep Bi-LSTM layers are subsequently integrated to the design to create a densely connected feed-forward structure. The model effectively prioritizes spatiotemporal information, thus extracting key insights for identification of interictal and preictal features. Dataset used here is the Boston Children's Hospital–MIT and train the model using fivefold cross validation. Our model is 3.44% more accurate than the best method available, and it also predicts times better.

PROPOSED METHODOLOGY

The noise, and artifacts of the EEG signals were removed by Improved Empirical Wavelet Transform (IEWT). It is based EEG signal in the frequency domain and dividing it into different subbands based on important points in the spectrum. After the signals are processed, a Separable Residual Attention U-Net (SRAUNet) model is used to extract features which include residual blocks and attention mechanisms. Seizure detection, Correlation Stacked Denoising Autoencoder (CSDAE) is introduced by stacking several denoising autoencoder layers; each layer removes the noises of inputs, and then propagates its output to the following layer for effective seizure detection. Need to examine the performance metrics like F-measure, classification recall/sensitivity and accuracy other problems with automatically detecting seizures using the CHB-MIT database when compared to the other methods.

Figure 1. explains the proposed method.

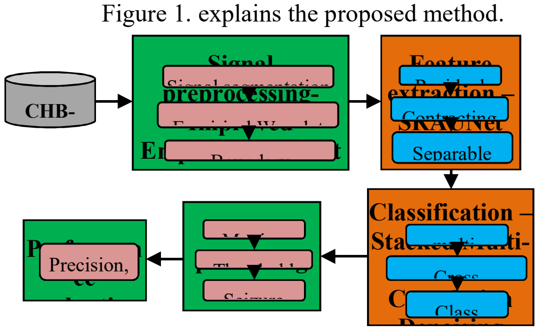

FIGURE 1. OVERALL FLOW OF PROPOSED MODEL

3.1. Improved Empirical Wavelet Transform (IEWT) based signal denoising

The Empirical Wavelet Transform breaks the vibration acceleration signal down into a number of empirical modes. There are three primary steps: (1) adaptively partitioning the spectrum, (2) using the boundaries that were found to make a suitable empirical wavelet filter bank, and filtering the signal and (3) utilize the Hilbert transform to demodulate and evaluate the empirical modes [27].

Step 1: Segregate the boundaries in the frequency domain

The EEG signal's frequency domain is $[0, \pi]$ and is split into N frequency bands, each with its own bandwidth. Then, a method for estimating spectral trends based on a key function is employed to find the edge of each frequency band. Gilles' method, on the other hand, uses the midpoints between consecutive maxima to establish borders.

Step 2: Develop an empirical wavelet filter bank

The Meyer wavelet is selected by Gilles as the foundation function for building an empirical wavelet. The boundary of the frequency band is located during the transition phase, where, a group of trigonometric functions that are all orthogonal to each other is created, and a constant is created in the frequency band.

Step 3: Perform an empirical wavelet transform

The EWT is built on a mechanism similar to the traditional wavelet transform. Here, F(.) denotes the Fourier transform, and $F^{-1}(.)$ represents the inverse Fourier transform. Spectrum segmentation and boundary optimization are two problems with EWT that are spoken about and explained. Fourier transform function of the spectrum is identified by introducing the IEWT approach, which gets the trend part of the spectrum. For reconstructing the EEG signal, the boundary is set as he minimum points of the trend component. To change the trend components, various points in key functions is chosen. Acceptable limits for EWT is given by the significant trend components.

3.1.1. Trend Estimation Method

The suggested method allows for the reconstruction of the spectrum's trend through trend estimate using the key function. The Fourier transform provides the theoretical basis for this procedure. When compared to the usual empirical wavelet transforms, this algorithm is superior, because it takes less time to compute.

Step 1: Attain the spectrum of the measured signal with the help of the FFT algorithm by equation (1),

$$□(□) = □□□(□(□)) \quad (1)$$

Step 2: Equation 2 Computes the Fourier transform function of □(□))

$$□(□) = □□□\ □(□) \quad (2)$$

Step 3: Equation 3 explains the components of key functions are closely related to the spectrum by equation (3),

$$□_{□}(□) = □□□□\left(□_{□}(□)\right) \quad (3)$$

$□_{□}(□)$ shows the left side of boundary B in K(f) and $□_{□}(□)$ shows a part of the trend. When the number of reconstructed points is less than the initial peak value of the key function $□_1$, the trend component becomes smooth, with very little change, and looks nearly like a straight line. It is hard to tell it apart from the spectral fluctuation pattern. There is a lot of information in the first section, but it doesn't separate the frequency components in the spectrum. To rebuild the trend component, thirty points are used, with the lowest values setting the bounds. The impact and modulation components show modulation information in the spectrum as a series of sidebands. So, the trend estimate method that uses the key function works better with EEG signals. The partition will be rough and won't be able to distinguish the difference between information with similar frequencies, if too many EEG signal components are chosen from the key function for the inverse transformation.

However, it will still be possible to keep the side-band frequencies and the modulation information. The suggested strategy can keep some parts and get rid of others that aren't needed. When a lot of points are picked for inverse transformation on simulated EEG signals, various details can be seen, including the components that have very similar frequency ranges. This will split the spectrum into various sections. The threshold noise-reduction method is presented to improve these problems [27].

3.1.2. Threshold Denoising method

Most of the time, for wavelet threshold denoising, this method makes use of the threshold denoising. Wavelet decomposition helps to identify the wavelet coefficients, they are denoised and the original EEG signal is used to rebuild the EEG signals. Different threshold functions have an effect on both the denoising effect and the reconstruction results when threshold denoising is used. When estimating data, it is very important to set the threshold. This method doesn't have to show the whole trend; thus, the noise-induced spectrum fluctuation needs to be decreased during threshold denoising. Equation (4) [27] shows the hard threshold function, which is one of the most used threshold functions.

$$\hat{□} = \begin{cases} □|□| \geq □ \\ □|□| < □ \end{cases} \quad (4)$$

And equation 5 gives the soft threshold function,

$$\hat{□} = \begin{cases} □□□(□).(|□| - □)|□| \geq □ \\ □|□| < □ \end{cases} \quad (5)$$

Where, the EEG signal that needs to be processed is denoted by s and the EEG signal after denoising is represented by $\widehat{□}$, with a predetermined threshold of λ [27].

$$□ = □\sqrt{2□□(□)} \quad (6)$$

The length of the signal is denoted by N that needs to be processed, and the anticipated noise variance is given by $□^2$. The hard threshold does a great job of keeping the signal's local information during processing, but the fact that s is not continuous at λ and - λ causes the reconstructed signal to oscillate at that point. During soft-threshold processing, the overall continuity of ŝ is better maintained. But there is always a discrepancy between s and ŝ since their derivatives are not continuous. This can make reconstruction less accurate. This approach simply needs to get the general trend of the spectrum; thus, it can ignore the problems with continuity that come with soft-thresholding. $□ = 0.5□/□^2$ is used as the threshold to process the signal sequence.

3.2. Separable Residual Attention U-Net (SRAUnet) based feature extraction

SRAUNet uses a residual attention, and it is inspired by Separable Convolutional Block Attention (SCBA) for feature extraction. The U-Net architecture includes the residual attention block. Features of the EEG signals are extracted from different frequency bands to capture the relevant brain activity patterns. For every 1-second segment, the EEG signals from all available channels were arranged into a 256×C×1 tensor, where the number of channels is given by *C*. To reduce differences in signal amplitude across patients and recording sessions, each channel was normalized using the Zscore method. Based on the timestamp information in the CHB-MIT dataset, each segment was marked as either a seizure or non-seizure event [28].

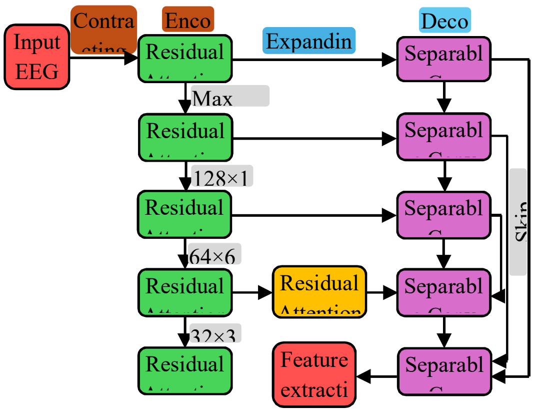

FIGURE 2. SEPARABLE RESIDUAL ATTENTION U-NET ARCHITECTURE

Figure 2 shows that the four 2×2 max pooling layers and four residual blocks are there in the contracting path. The EEG signals are in a 256×23×1 picture. Two 3×3 convolution layers are there in the residual block, each followed by ReLU activation algorithms and batch normalization. The first residual block changes the 256×23×1 input EEG signals into a 256×23×64 feature map. The max-pooling layer decrease the size of the feature map in half and doubles the number of channels. The outcome is a 16×1×1024 feature map of four residual blocks moving through the contracting path. This structure enables the network to gradually capture low-level temporal fluctuations, mid-level inter-channel features, and high-level seizure patterns. The problem occurs in the evolution region between the expanding path and the contracting path. It has a 16×1×1024 feature map and one residual block. The output from the bottleneck goes into the path that gets smaller. The growing path and the decreasing path both have four residual bricks [28]. Because the feature map gets smaller along the contracting path, it needs to be brought back to its original spatial resolution, 2x2 separable convolution layers are used to do upsampling. So, in each growing layer, the number of features is cut in half and the feature map's size is doubled. The attention module's output is added to the feature map that was up-sampled in the preceding layer. The resulting feature map is 256 × 23 × 1, which is the same size as the four residual blocks used along the expanding path. Finally, the decoded feature map is converted into a single probability value by a 1×1 convolution layer that reflects the likelihood of a seizure in the corresponding EEG segment [28].

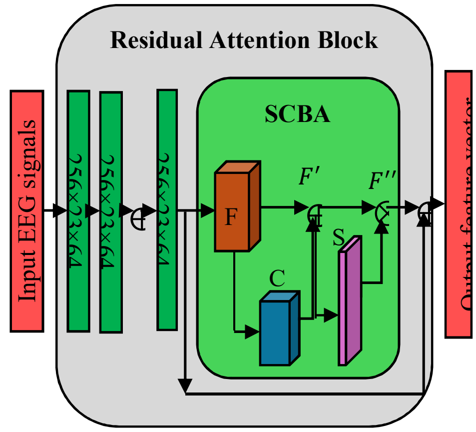

FIGURE 3. RESIDUAL ATTENTION BLOCK ARCHITECTURE

Figure 3 shows how the residual attention block is put together. This basic method analyzes the input EEG signals by adding them to the learnt function. This solves the problem of gradients disappearing as the model gets deeper. To stop performance from getting worse, a residual block is used at each step along both the contracting and expanding paths. This makes the model better at getting more features from each layer. You can say something about this process:

$$x_{i+1} = \mathrm{F}(\mathrm{x_i}) + \mathrm{x_i} \tag{7}$$

$$\mathrm{F}(\mathrm{x_i}) = \mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{K_2} * \mathrm{ReLU}\big(\mathrm{BN}(\mathrm{K_1} * \mathrm{x_i})\big)\right)\right) \tag{8}$$

where K_1, K_2 is denoted as the convolutional kernels, * is denoted as the convolutional operator. U-Net adds together the feature maps of EEG signals that come from the expanding path and contracting path one at a time via skip connections. By merging context and location data from the signal and deep layers, a sharper signal may be created, which can help find epileptic seizures more accurately. Use the attention module to focus on an EEG recording to make the detection model work better. As demonstrated in Figure 2, SCBA uses channel attention and spatial attention one after the other. You can use equations (9-10) to show how this works.

$$F' = \mathrm{CA(F)} \odot \mathrm{F} \tag{9}$$

$$F'' = \mathrm{SA}(\mathrm{F}') \odot \mathrm{F}' \tag{10}$$

The combination of MXP - max pooling and AVP - Average Pooling gives channel attention.

According to equation 11: input feature F is the input give and the obtained result is CA (channel attention).

$$\square\square(\square) = \sigma\Big(\mathrm{MLP}\big(\mathrm{AVP}(\mathrm{F})\big) + +\mathrm{MLP}\big(\mathrm{MXP}(\mathrm{F})\big)\Big) \quad (11)$$

In this case, σ stands for the sigmoid function that is employed in the batch normalization process. The spatial attention module creates the spatial attention map (SA) using the refined channel attention (CA) from the previous stage, as shown in Equation 12.

$$\square\square(\square) = \sigma\left(\mathrm{f}^{7\times7}([\mathrm{AVP}(\mathrm{F}); \mathrm{MXP}(\mathrm{F})])\right) \quad (12)$$

where $\mathrm{f}^{7\times7}$ is denoted as the convolution with 7×7 kernel, [□] [□; □] is denoted as the channel-wise concatenation. The residual attention output is described by the equation (13),

$$\square(\mathrm{x_i}) = \mathrm{SCBA}(\mathrm{F}(\mathrm{x_i}) + \mathrm{x_i}) \quad (13)$$

The attention mechanism's output is combined with the feature map from the upsampled path as shown in Equation (14).

$$\hat{\square} = \sigma(\mathrm{W_f} * \mathrm{U_0}) \quad (14)$$

where $\mathrm{U_0}$ is the final feature map from decoder, 1×1 convolution is applied with weights $\mathrm{W_f}$. This approach enhances segmentation by focusing on the recognition of epileptic seizures to identify the frequency and timing characteristics of EEG signals. By incorporating a Residual Attention Block using Residual Blocks and the SCBA into the U-Net architecture, the suggested model leverages both residual learning and attention mechanisms. In order to detect the subtle EEG patterns associated with preictal and ictal stages, the residual paths assist the network in learning deeper representations and minimizing vanishing-gradient problems. By using channel and spatial attention, SCBA simultaneously improves feature extraction, enabling the model to downweight noise and less valuable signals while concentrating on the most pertinent EEG data. By incorporating these Residual Attention Blocks into U-Net's encoder–decoder structure, the network is able to record both specific temporal variations and more general relationships between EEG channels in the CHB-MIT. This combination increases the overall reliability of seizure detection and promotes more expressive feature learning.

3.3. Stacked Multi-scale Correlation Denoising Autoencoder (SMCDAE) based seizure detection

In this study, Stacked Multi-scale Correlation denoising autoencoder (SMCDAE) is introduced for seizure detection of EEG signals, which is further divided into two parts: (1) according to the back-propagation (BP) method, the creation of a single SMCDAE and (2) the establishment of a deep network with robust non-linear mapping and advanced feature extraction capabilities, it is required to stack of a multi- SMCDAE. To represent multi-scale temporal patterns and inter-channel correlations in EEG signals, a model is enbled by this stacked architecture. Figure 4 shows how the original input vector x^i is changed into a corrupted EEG input vector x˜^i by randomly adding Gaussian noise and then using a nonlinear activation function through linear mapping. For this operation, the SMCDAE uses common weights [29], which is different from the AE.

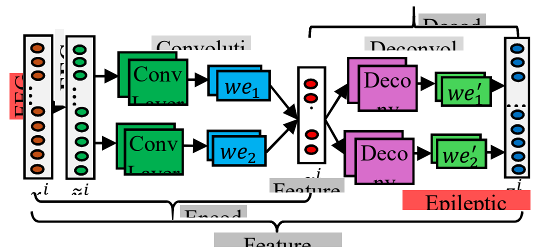

FIGURE 4. MCDAE ARCHITECTURE

Equation (15) described the possible encoding of the ith feature map for a single-channel input x_i.

$$\square^{\square} = \mathrm{f}\big(\mathrm{we_1} * \tilde{\mathrm{x}}^{\square} + \square_{l}\big) \quad (15)$$

For encoder, weighting matrix is denoted by $\mathrm{we_1}$, the encoding bias vector is given by $\square_l$, and a convolution operation is represented by * [29]. A latent feature representation is denoted by the word $\square^{\square}$, while f(·) represents a function of Leaky ReLU as articulated in equation (16).

$$\square = \begin{cases} \mathrm{x}, & \mathrm{x} \geq 0 \\ \omega\mathrm{x}, & \mathrm{x} \leq 0 \end{cases} \quad (16)$$

A coefficient is supplied ω, and feature fusion was done by making the convolution layer's size the same. In the proposed SMCDAE framework for EEG signal analysis, the decoding stage focuses on reconstructing clean EEG activity from the latent representations learned during encoding. The process begins with a feature blending step, where feature extracted from several convolutional layers—each operating at a different temporal scale is combined into new vector. By merging these multi-scale representations, the model forms a richer and more comprehensive description of the underlying neural activity. Feature-fusion mechanism further strengthens the overall representation by balancing the contribution of each feature type. Prior studies have shown that such fusion strategies significantly enhance the performance of EEG based classification. During the decoding step, the intermediate latent representation is as □□ transformed back into the time domain to reconstruct the original EEG signals [29]. To get the reconstructed EEG vector □□, a convolutional mapping is used followed by a nonlinear activation function as shown in equation [17].

$$□^{□} = g(we_2 * \alpha_i + b_2) \tag{17}$$

where $we_2$is dentoed as the decoder's convolution kernel, $b_2$is a bias term, □ indicates 1-D convolution over the EEG signal, and the activation function is given by g(.). The reconstructed output encourages the model to learn latent features that preserve important seizure-related temporal patterns while effectively reducing noise and artifacts, ultimately improving the reliability of seizure detection [29]. Collection of training EEG signals □ = □$_1$, □$_1$), (□$_2$, □$_2$), . . . , (□$_□$, □$_□$), defines the total cost function for the SMCDAE on the EEG dataset □ as follows:

$$□(□□, □) = \frac{1}{2M}\sum_{i=1}^{M}\left\|x^i - z^i\right\|^2 \tag{18}$$

In equation (18), a regularization term is included, and it is described by equation (19),

$$□(□□, □) = \frac{1}{2M}\sum_{i=1}^{M}\left\|x^i - z^i\right\|^2 + \frac{\lambda}{2}\|□□\|_2^2 \tag{19}$$

The encoder in equation (19) is the part that gets the features. It learns by minimizing the cost function's reconstruction error, with λ serving as the regularization coefficient [29]. A greedy layer-wise pre-training strategy was used to build the stacked SMCDAE model. The BP algorithm trains the first SMCDAE1.The input for the second SMCDAE2 is acquired from the output of the encoder α i 1, and trained further. The input for SMCDAE3 is from the latent feature vector α i 2. As seen in Figure 5, this process went on through several layers. From a deep stacked MGCDAE, the N stacked SMCDAEs made, which is called an SMGCDAE. Equation (20) was used to find the latent feature representation □$^{□}_{□}$.

$$\alpha_N^i = f\left(we_1^N * \alpha_{N-1}^i * \rho + b_1^N\right) \tag{20}$$

Where, b_1^N is bias vector of the N^{th} SMCDAE and we_1^N is the weight matrix, the cross correlation coefficient between the acutal EEG signal, and detected EEG signal is denoted as ρ. Cross-correlation is a way to see how similar two series are based on how far off they are from each other. For real EEG sequences of length N, ρ is described by equation (21),

$$\rho = R_{x,y}[k] = \sum_{n=0}^{N-1-k} x[n]y[n+k]\,, k = 0,1, \ldots, N-1 \tag{21}$$

In this approach, the network performs progressively deeper feature mapping, where the output of each hidden layer becomes a more abstract representation of the original EEG signal. Thus, the goal of the deep SMCDAE is to generate multiple levels of representational mappings. The initial EEG signal's feature layers are gradually abstracted until the high-level characteristics are reached. After the fusion process, these features are sent to the classifier, which finishes the last step of detection. This study used a softmax classifier to sort the signals into two groups: seizure and non-seizure. To train this deep network, a cost function was used to lower the reconstruction error. This made sure that the output class was very near to the input class.

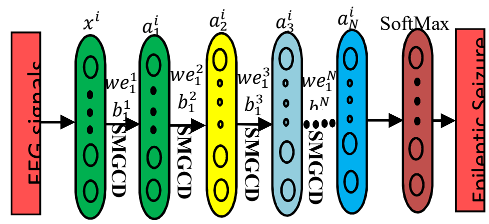

FIGURE 5. SMGCDE NETWORK ARCHITECTURE

Figure 5 illustrates the SMGCDE training method, segmented into two phases: fine-tuning and pre-training [58]. The superposition of several MGCDAEs was done without supervision and from the top down, during pre-training phase. Further, to make the whole structure work better, the supervised learning was utilized t train the softmax classifier.

4. RESULTS AND DISCUSSION

The algorithms for finding epileptic seizures runs on MatlabR2019a, provided the computer with an i7 CPU, 16GB of RAM, and an NVIDIA RTX3050 GPU. The testing data is used from all the cases in the CHB-MIT database to see how well the seizure detection methods worked.

4.1. Database

The CHB-MIT scalp EEG database, created by the Children's Hospital Boston, is publicly available at https://physionet.org/content/chbmit/1.0.0/. The dataset contains long-term EEG recordings from children with epilepsy. All signals were sampled at 256 Hz with a 16-bit resolution. Most recordings have 23 channels, while some have 24 or 26. According to the International 10–20 scheme, the electrodes were placed which makes sure that all cortical areas are covered in a standard way. To prepare the data for learning, each EEG channel was band-limited using a FIR - finite impulse response filter between 0.5–40 Hz to suppress high-frequency disturbances, EMG noise, baseline drift and power-line interference. Artifact-induced epochs were excluded to maintain signal quality. Following filtering, each recording was segmented into fixed-length windows of 256 samples (1 second) with 50% overlap, a configuration shown to improve temporal precision in seizure detection tasks. There are EEG recordings from 23 persons with epilepsy in the CHB-MIT database. These recordings were organized into 24 groups. 1.5 years apart, the same person gave us case chb21 and case chb01. Each example has between 9 and 42 continuous EEG data. Most files contain one hour of EEG recordings, while a few files—such as those from subjects chb04, chb06, chb07, chb09, chb10, and chb23—include two to four hours of data. The subjects were monitored for several days after withdrawing anti-seizure medication to capture seizure events and evaluate their suitability for surgical treatment. A total of 198 seizures have clearly annotated start and end points. The training dataset includes either an equal amount of non-seizure EEG or two to four times more non-seizure data than seizure data.

4.2 Performance evaluation

The efficacy of seizure detection systems was computed by precision, recall/sensitivity, F-measure, and accuracy serve as assessment metrics, calculable by the subsequent equations (22-25).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{22}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{23}$$

$$\text{F} - \text{measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{24}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{25}$$

True Positive (TP) and True Negative (TN) are the numbers of seizure and non-seizure segments, respectively, that the detection system correctly identifies. False Positive (FP) means the number of EEG segments that are not seizures but are wrongly categorized as seizures. False Negative (FN) means the number of seizure segments that are wrongly identified as not seizures.

TABLE 1. RESULTS COMPARISON OF SEIZURE DETECTION METHODS

Methods	Precision (%)	Recall (%)	F-Measure (%)	Accuracy (%)
Bi-LSTM	85.57	87.79	86.67	86.83
Bi-GRU	87.38	89.52	88.44	88.21
CE-stSENet	91.51	92.88	92.19	91.54
AROBILSTM	93.65	94.23	93.94	93.57
SMCDAE	94.87	95.39	95.13	96.41

Table 1 shows the performance comparison, which depicts the detection methods of seizure such as Bi-LSTM, ConvLSTM, Bi-GRU, CE-stSENet, AROBILSTM - Attention-based Residual Optimized Bidirectional Long Short-Term Memory and SMCDAE with regards to precision, accuracy, F-measure, and recall.

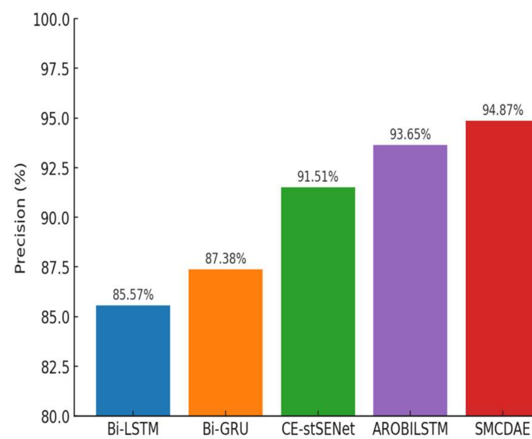

FIGURE 6. PRECISION ANALYSIS OF SEIZURE DETECTION METHODS

Figure 6 displays a detailed comparison of the accuracy of numerous cutting-edge seizure detection methods, including SMCDAE, CE-stSENet, AROBiLSTM, Bi-GRU, Bi-LSTM and ConvLSTM. As illustrated, the proposed classifier attains the highest precision value of 94.87%, indicating a superior ability to properly recognize seizure events while reducing false positives. Bi-LSTM and Bi-GRU obtain 85.57% and 87.38%, respectively. CE-stSENet and AROBILSTM perform slightly better with a precision of 91.51% and 93.65%. Proposed model has 9.30%, 7.49%, 3.36%, and 1.22% highest precision when compared to ConvLSTM, Bi-LSTM, Bi-GRU, CE-stSENet, and AROBiLSTM. Proposed architecture interprets the discriminative EEG patterns essential to precise seizure detection.

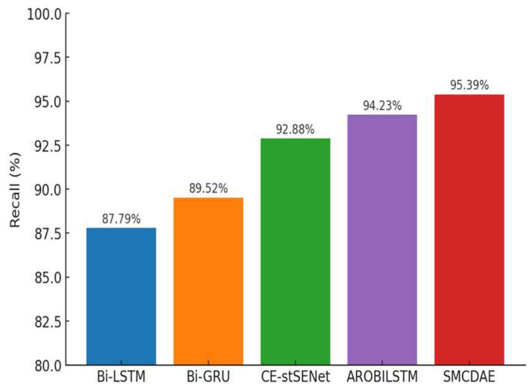

FIGURE 7. RECALL ANALYSIS OF SEIZURE DETECTION METHODS

Recall achieved by several state-of-the-art seizure detection techniques like ConvLSTM, Bi-LSTM, Bi-GRU, CE-stSENet, AROBiLSTM, and SMCDAE are illustrated in figure 7. As illustrated, the proposed classifier attains the highest recall value of 95.39%, indicating a superior ability to correctly identify seizure events while minimizing true negative, and false negatives. Bi-LSTM and Bi-GRU obtain 87.79% and 89.52%, respectively. CE-stSENet and AROBILSTM perform slightly better with a recall of 92.88% and 94.23%. Proposed model has 7.60%, 5.87%, 2.51%, and 1.16% highest recall when compared to ConvLSTM, Bi-LSTM, Bi-GRU, CE-stSENet, and AROBiLSTM. The significant performance improvement by the proposed method highlights its improved capability to extract and learn discriminative EEG representations, which is crucial for consistent and correct seizure identification.

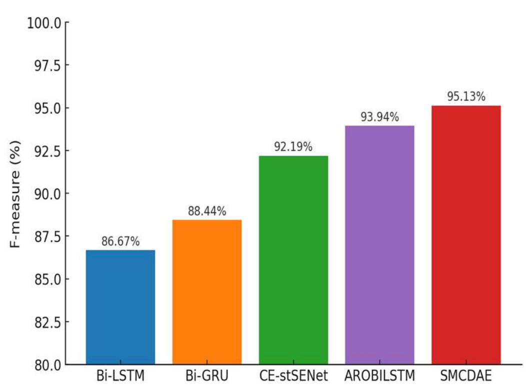

FIGURE 8. F-MEASURE ANALYSIS OF SEIZURE DETECTION METHODS

ConvLSTM, Bi-LSTM, Bi-GRU, CE-stSENet, AROBiLSTM, and SMCDAE with respect to f-measure are illustrated in figure 8. As illustrated, the proposed classifier achieves the 95.13% of f-measure, indicating a harmonic balance of recall and precision. Bi-LSTM and Bi-GRU obtain 86.67% and 88.44%, respectively. CE-stSENet and AROBILSTM perform slightly better with an f-measure of 92.19% and 93.94%. Proposed model has 8.46%, 6.69%, 2.94%, and 1.19% highest f-measure when compared to ConvLSTM, Bi-LSTM, Bi-GRU, CE-stSENet, and AROBiLSTM. The proposed model has attained highest F-measure reflecting its strong ability to accurately detect seizure events with lowest false negatives.

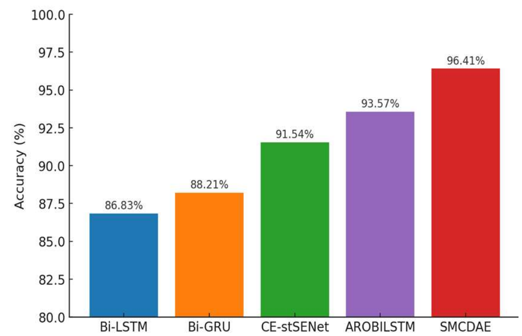

FIGURE 9. ACCURACY ANALYSIS OF SEIZURE DETECTION METHODS

Accuracy comparison of seizure detection methods like ConvLSTM, Bi-LSTM, Bi-GRU, CE-stSENet, AROBiLSTM, and SMCDAE are illustrated in figure 9. Proposed classifier attains the highest accuracy of 96.41%, Bi-LSTM and Bi-GRU obtain 86.83% and 88.21%, respectively. CE-stSENet and AROBILSTM perform slightly better with accuracy of 91.54% and 93.57%. Proposed model has 9.58%, 8.20%, 4.87%, and 2.84% highest accuracy when compared to ConvLSTM, Bi-LSTM, Bi-GRU, CE-stSENet, and AROBiLSTM.

5. CONCLUSION AND FUTURE WORK

This research demonstrates an effective approach for the early identification of epilepsy through EEG signals. Stages like dataset acquisition, feature extraction, signal denoising, detection, and analysis of performance were incorporated in this suggested method. First, the EEG recordings from the Children's Hospital Boston–Massachusetts Institute of Technology (CHB-MIT) dataset are used as the Epileptic Seizure Recognition dataset to identify epileptic seizures. Secondly, to improve the quality of the EEG recordings, an Improved Empirical Wavelet Transform (IEWT) is introduced for signal pre-processing. The Fourier transform is used by the IEWT method to find the trend component of the spectrum. Thirdly, Separable Residual Attention U-Net (SRAUnet) uses a residual attention mechanism that is based on SCBA, or Separable Convolutional Block Attention, which enforces both channel and spatial attention for feature extraction. For each step along the contracting and expanding paths, a residual block is employed. This helps the model get additional features from each layer. Fourthly, Stacked Multi-scale Correlation denoising autoencoder (SMCDAE) model is introduced to record EEG signals' multi-scale temporal patterns and correlations between channels. Using the CHB-MIT dataset, measures like recall, F-measure, accuracy, and precision were used to compare the performance of the proposed model to that of existing detection approaches. Transfer learning methods have been found very effective for reducing training time and deal with scarcity, especially in EEG-based seizure detection and other biomedical applications.

REFERENCES

1. Malekzadeh, A., Zare, A., Yaghoobi, M., Kobravi, H.R. and Alizadehsani, R., 2021. Epileptic seizures detection in EEG signals using fusion handcrafted and deep learning features. Sensors, 21(22), pp.1-28.
2. Tatum, W.O., Rubboli, G., Kaplan, P.W., Mirsatari, S.M., Radhakrishnan, K., Gloss, D., Caboclo, L.O., Drislane, F.W., Koutroumanidis, M., Schomer, D.L. and Trenite, D.K.N., 2018. Clinical utility of EEG in diagnosing and monitoring epilepsy in adults. Clinical Neurophysiology, 129(5), pp.1056-1082.
3. Farooq, M. S., Zulfiqar, A. & Riaz, S. Epileptic Seizure Detection Using Machine Learning: Taxonomy, Opportunities, and Challenges, Diagnostics, vol. 13, no. 6, pp. 1-22, 2023.
4. Ein Shoka, A.A., Dessouky, M.M., El-Sayed, A. and Hemdan, E.E.D., 2023. EEG seizure detection: concepts, techniques, challenges, and future trends. Multimedia Tools and Applications, 82(27), pp.42021-42051.
5. Aayesha M, Qureshi B, Afzaal M, Qureshi MS, Fayaz M (2021) Machine learning-based EEG signals classification model for epileptic seizure detection. Multimed Tools Appl 80(12):17849–17877.
6. Acharya, U.R., Hagiwara, Y., Deshpande, S.N., Suren, S., Koh, J.E.W., Oh, S.L., Arunkumar, N., Ciaccio, E.J. and Lim, C.M., 2019. Characterization of focal EEG signals: a review. *Future Generation Computer Systems*, *91*, pp.290-299.
7. Slimen, I.B., Boubchir, L. and Seddik, H., 2020. Epileptic seizure prediction based on EEG spikes detection of ictal-preictal states. Journal of biomedical research, 34(3), pp.162–169.
8. Mahjoub, C., Jeannès, R.L.B., Lajnef, T. and Kachouri, A., 2020. Epileptic seizure detection on EEG signals using machine learning techniques and advanced preprocessing methods. Biomedical Engineering/Biomedizinische Technik, 65(1), pp.33-50.
9. Boonyakitanont, P., Lek-Uthai, A., Chomtho, K. and Songsiri, J., 2020. A review of feature extraction and performance evaluation in epileptic seizure detection using EEG. Biomedical Signal Processing and Control, 57, pp.1-28.
10. Zhou, Y., You, J., Zhu, F., Bragin, A., Engel, J., and Li, L. (2021). Automatic electrophysiological noise reduction and epileptic seizure detection for Stereoelectroencephalography. In: 2021 43rd annual international conference of the IEEE engineering in Medicine and Biology Society (EMBC), pp. 107–112.
11. Yedurkar, D. P., and Metkar, S. P. (2020). Multiresolution approach for artifacts removal and localization of seizure onset zone in epileptic EEG signal. Biomed. Signal Process. Control 57:101794. doi: 10.1016/j.bspc.2019.101794
12. Bhagubai, M., Swinnen, L., Cleeren, E., Van Paesschen, W., De Vos, M. and Chatzichristos, C., 2024. Towards automated seizure detection with wearable eeg–grand challenge. IEEE Open Journal of Signal Processing, 5, pp.717-724.
13. Santhosh, S., Parthasarathi, V.P., Gokul, P.K., Rasswanth, G. and Devender, R., 2025. Empirical Wavelet Transform with Over Complete Dictionary for the decomposition of signals with abrupt changes. Next Research, 2(3), pp.117–130.
14. Kode, H., Elleithy, K. and Almazaydeh, L., 2024. Epileptic seizure detection in EEG signals using machine learning and deep learning techniques. *IEEE Access*, *12*, pp.80657-80668.
15. Zhang, X., Zhang, X., Huang, Q. and Chen, F., 2024. A review of epilepsy detection and prediction methods based on EEG signal processing and deep learning. Frontiers in Neuroscience, 18, pp.1-20.
16. Rukhsar, S. and Tiwari, A.K., 2023. Lightweight convolution transformer for cross-patient seizure detection in multi-channel EEG signals. Computer Methods and Programs in Biomedicine, 242, pp.1-13.
17. Pandey, A., Janghel, R. R., Mishra, P. K., & Ahirwal, M. K. (2023). An intelligent optimized deep learning model to achieve early prediction of epileptic seizures. Computers in Methods and Programs in Biomedicine, 242, 107856. https://doi.org/10.1016/j.cmpb.2023.107856.
18. Zhao, Y., He, J., Zhu, F., Xiao, T., Zhang, Y., Wang, Z., Xu, F., & Niu, Y. (2023). Hybrid attention network for epileptic EEG classification. *International Journal of Neural Systems*, 33(6), 2350031.

19. Duan, L., Wang, Z., Qiao, Y., Wang, Y., Huang, Z. and Zhang, B., 2021. An automatic method for epileptic seizure detection based on deep metric learning. IEEE Journal of Biomedical and Health Informatics, 26(5), pp.2147-2157.
20. Li, Y., Liu, Y., Cui, W.G., Guo, Y.Z., Huang, H. and Hu, Z.Y., 2020. Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(4), pp.782-794.
21. Geng, M., Zhou, W., Liu, G., Li, C. and Zhang, Y., 2020. Epileptic seizure detection based on stockwell transform and bidirectional long short-term memory. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(3), pp.573-580.
22. Hu, X., Yuan, S., Xu, F., Leng, Y., Yuan, K. and Yuan, Q., 2020. Scalp EEG classification using deep Bi-LSTM network for seizure detection. Computers in Biology and Medicine, 124, pp.1-8.
23. Yao, X., Li, X., Ye, Q., Huang, Y., Cheng, Q. and Zhang, G.Q., 2021. A robust deep learning approach for automatic classification of seizures against non-seizures. Biomedical Signal Processing and Control, 64, pp.1-18.
24. Khan, G.H., Khan, N.A., Altaf, M.A.B. and Abbasi, Q., 2023. A shallow autoencoder framework for epileptic seizure detection in EEG signals. Sensors, 23(8), pp.1-21.
25. Tawhid, M.N.A., Siuly, S. and Li, T., 2022. A convolutional long short-term memory-based neural network for epilepsy detection from EEG. IEEE Transactions on Instrumentation and Measurement, 71, pp.1-11.
26. Quadri, Z.F., Akhoon, M.S. and Loan, S.A., 2024. Epileptic Seizure Prediction using Stacked CNN-BiLSTM: A Novel Approach. IEEE Transactions on Artificial Intelligence, vol.5, no.11, pp. 5553 – 5560.
27. Bhattacharyya, A., Sharma, M., Pachori, R.B., Sircar, P. and Acharya, U.R., 2018. A novel approach for automated detection of focal EEG signals using empirical wavelet transform. Neural Computing and Applications, 29, pp.47-57.
28. Zhang, J., Lv, X., Zhang, H. and Liu, B., 2020. AResU-Net: Attention residual U-Net for brain tumor segmentation. Symmetry, 12(5), pp.1-15.
29. Wen T. and Z. Zhang, ‘‘Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals,’’ IEEE Access, vol. 6, pp. 25399–25410, 2018.