

Feature Extraction of EEG Signal Using Convolutional Neural Networks by Removing Artifacts

Padmini Chattu¹, C.V.P.R. Prasad²

¹Acharya Nagarjuna University, Nagarjuna Nagar, A.P, India

²Acharya Nagarjuna University, Nagarjuna Nagar, India

E-mail: pchattu2021@gmail.com

Keywords: feature extraction, convolution neural networks, morelette, random forest classifier

Received: June 14, 2024

Clinical depression is a neurological disease identifiable by the analysis of the electroencephalography signals (EEG). The electroencephalographic signals (EEG) are often polluted by many artifacts. Deep study models have been employed in recent years to denoise electroencephalography. The main difficulty in medical analysis is the extraction of true brain signals from the polluted EEG data. Noise reduction from recorded EEG data is very important for better brain disorder investigation. This paper proposed an effective EEG signal estimation model for the process of EEG signals. The proposed model uses the Morelette wavelet transformation model for the pre-processing of the EEG signal. With the pre-processed EEG signal model feature extraction is performed with the Convolutional Neural Network (CNN) for the EEG signal. With the pre-processed EEG signal model training and testing are estimated for the classification of the EEG signal. The EEG signal categorization was carried out utilizing characteristics derived from EEG data. Many characteristics have proven sufficiently distinctive for usage in all applications linked to the brain. The EEG may be categorized using a range of functions such as autoregression, energy spectrum density, energy entropy and linear complexity. However, various characteristics indicate varying strength of discrimination for different individuals or trials. Two characteristics are utilized in this study to enhance the performance of EEG signals. Techniques based on the neural network are used for the extraction of EEG signal. Classification methods include the Random Forest Classification. The model was tested using a random splitting method and 93.4 percent of the EEG signals were received accordingly.

Povzetek: Prispevek predstavi metodo za odstranjevanje artefaktov iz EEG-signalov z uporabo CNN in Morletove valovne transformacije.

1 Introduction

The brain-computer interface is an electronic and human brain direct communication and controlling mechanism [1]. In several areas, BCI systems offer significant application value, in particular in the realm of medical treatment [2]. Different signals for Electroencephalograms (EEGs) have been utilized in BCI systems such as P300 potentials, SSVEP and motor imaging (MI). In the diagnosis of neurologic diseases like epileptic seizures and neurophysiologists, electroencephalogram (EEG's) signals also rely largely on these EEG signals. Close to 1% of the whole global population suffers from epilepsy, a significant neurological disorder [3]. The numerous spikes are utilized to describe the EEG data during the start of epileptic attacks. Two kinds of epileptic seizures may be categorized. Focal epileptics and general epileptic seizures are included. The seizure in the cerebral hemisphere is seen as a focused seizure that shows the symptoms in the respective areas that in turn impact mental health. The later covers the brain in its whole, resulting to the awareness of bilateral motor symptoms. Epileptic seizures may affect everyone irrespective of age. EEG epilepsy patient monitoring includes two types of abnormal

activity. One is inter-ictal, aberrant EEG signals collected between episodes, which are ictal during an epileptic attack [4]. The major distinction is that the interictal activity shows transient waveforms in EEG recording in the form of varied pits, Spike trains, narrow waves or complex Spike waves, whereas the polymorphic differential range and frequency waveforms represent the ictal activity in EEG. In the early days, parametric techniques and transformations based on Fourier were employed. Frequency changes in the subband linked with EEG seizures are illustrated in μ (0.4-4 Hz), β (4-8 Hz), α (8-12 Hz), and β (4-8 Hz) (12-30 Hz).

Conventional techniques based on frequencies are usually appropriate for isolating EEG information from non-stationary and multi-component EEG signals. In comparison with traditional frequency techniques greater performance in time-frequency approaches is found. Several techniques were recommended for the reduction of Scalp EEG artifacts that record epileptic people in order to enhance diagnosis or seizure detection [5]. The main emphasis of the study is on wavelet and band modification, including seizure activity (i.e., 0.5-29 Hz). Real EEG records reveal the most often replicated items with different artifact templates. The technique

utilizes three types of false data, including completely simulated, semi-simulated, and true data to assess the identification and seizure efficiency of artifacts. Once the artifacts are eliminated, seizures may readily be distinguished from any seizures responsible for EEG. As a result, false alarms in the detection of seizures are decreased. This is a technique used to extract normal signals and epileptic seizure signals from EEG data by choosing the least number of functions for individuals with epileptic signal. Detailed EEG coefficient and approximation coefficients are produced by choosing the minimal number of WT features. The coefficient produced by the wavelet generates 40 distinct characteristics utilizing statistical techniques which induce frequency distributions and varying quantities in frequency distributions [6].

2 Related works

Researchers have utilized a series of signal processing and classification-based machine learning methods in recent years to categorize cognitive processes automatically based on intellectual arithmetic. Fatimah et al. [7] have derived from cycles of each EEG channel the standard L2, meaning the entropy of Shannon and energy parameters.

For classifying cognitive workers, before mental arithmetic (BFMAC) or rest and mental arithmetic calculations (DMAC), or active state, a supportive vector maker (SVM) classifier has been used. A rate of accuracy of 95.80 per cent was achieved utilizing the Decision Tree categorization technique. The same authors also used Fourier's decomposition technique in another research to assess the EEG sub band signals [8].

The variance, energy and entropy characteristics of the EEG data were recovered from the Fourier domain; the psychology tasks BFMAC and DMAC were separated into an SVM rating; the exactness was 98.60 percent. The model of long-term storage (LSM) was utilized to categorize mental arithmetic activity by the multi-channel EEG data with spectral and instantaneous frequency characteristics [9]. 91.67 per cent were categorized for the BFMAC vs. DMAC classification test utilizing the stacked LSTM classification.

A number of EEG data channels, including rhythms, fractals, auto-regressive model coefficients, and statistical variables, were characterized by Wang and Sourina [10]. To classify mental arithmetic tasks, we employed a support vector machine (SVM) model and principal component analysis (PCA) to reduce the vector size. The autoregressive EEG and SVM-based analyses were also used to classify the cognitive mental-arithmetic workload [11].

The methods presented in [12] comprised just statistical and spectral properties of EEG signals such the BFMAC vs. the DMAC. The method presented by [13] also provides a framework for the identification of mental arithmetic tasks using EEG signal features. In [14] the authors investigated the three categorization functions of EEG, for example the baseline vs the mental arithmetic versus the mental letter. The previous research did not identify activities for cognitive classification of working

load such as bad computations of mental arithmetic (BMAC) compared to good calculation of mental arithmetic (GMAC) using EEG signal analytics.

The methods referred to above only investigated the temporal and frequency-domain features of EEG data in the classification of mental arithmetic tasks. State-space enterothe features of a wide-spread, partial-epileptic seizure [16], emotional recognition [17] and BCI [18] are used as measurements of not linearity and randomness by means of EEG signalling [15]. The use of EEG data was not explored with respect to these non-linear entropy characteristics, including scattered entropy [19], pitfall entropy [20], and other entropy metrics. Deep learning methods like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have recently found extensive use in EEG signal processing applications [21].

When it comes to analyzing sequence data and using it for speech recognition, RNNs—deep neural network models—are the way to go. Time series data modeling [22] makes use of this kind of network to investigate the long-range dependencies. Consideration of input and data from the prior stage is necessary for evaluating the present phase. To classify mental arithmetic activities based on EEG signal properties, the LSTM-based RNN model was employed [23]. Through correlations between EEG data across different times, the RNN models can categorize mental arithmetic operations. As an example, the EEG signal BFMAC against DMAC or BMAC versus GMAC was not utilized by other RNN versions, such as the Bidirectional and Gated Recurrent units (GRU) [24], to categorize mental arithmetic processes.

2.1 Research motivation

The DLN was widely used for image processing and other applications, as we all know. Its usage in EEG is uncommon despite the efficacy of DLN [25]. This article offers a newness, resilient and efficient DLN to eliminate noise artifacts in order to address the shortcomings of the conventional techniques described above.

Like the polluted EEG. The technique suggested has the following features:

(i) This technique does not need any extra ENG (EEG with artifact Noises) recording for reference signals either offline or live, which is convenient for subjects with excellent applications for rehabilitation brain interface systems.

(ii) This technique is suited for a few EEG electrodes which are convenient for EEG recording, cost saving and suitable for use.

(iii) NAs (Noise artifacts) may be eliminated automatically using the learnt model which is quick to use and can be used online.

(iv) The generalization potential of this technique is excellent and may be extensively utilized. The technique suggested may be split into two phases (Fig. 1). The first phase is pre-processing elimination of artifacts and the second phase is the model derived from the feature.

2.2 Background convolutional neural network

A deep-learning model is a neural network of knowledge based on data, was widely shown by picture classification and object identification techniques, and includes a complete (FC) layer, a linear (Conv)

convolutional layer, a model that does not have a linear or nonlinear function. One of CNN's major advantages is the utilization of a limited spatial area for input pictures; it shares certain sharp parameters and less weights. This method is mainly more efficient than previous models. Table 1 presented the overall summary of the literature review for the EEG signal processing.

Table 1: Summary of the literature

Study	Methods	Features	Classification Tasks	Classifier	Accuracy	Key Findings
Fatimah et al. [7]	L2 norm, Shannon entropy, Energy parameters	Cycles of each EEG channel	BFMAC vs. DMAC	SVM	95.80%	Decision Tree classification achieved 95.80% accuracy; SVM was used in earlier stage for task classification.
Fatimah et al. [8]	Fourier's decomposition	Variance, Energy, Entropy	BFMAC vs. DMAC	SVM	98.60%	Fourier's decomposition of EEG subband signals improved accuracy to 98.60%.
Study [9]	Spectral and Instantaneous Frequency features	Multi-channel EEG data	BFMAC vs. DMAC	Stacked LSTM	91.67%	Stacked LSTM model used for spectral and frequency characteristics, achieving 91.67% accuracy.
Wang & Sourina [10]	PCA, SVM, Auto-regressive model coefficients	Rhythms, Fractals, Statistical variables	Mental Arithmetic vs. Rest	SVM, PCA	Not specified	PCA for feature reduction; SVM used for mental arithmetic task classification.
Study [11]	Auto-regressive EEG	Auto-regressive coefficients	Cognitive Mental Arithmetic Workload	SVM	-	Auto-regressive model coefficients used in combination with SVM for workload classification.
Study [12]	Statistical and Spectral EEG properties	BFMAC vs. DMAC	BFMAC vs. DMAC	-	-	Focused on statistical and spectral features for BFMAC vs. DMAC classification.
Study [13]	EEG signal feature extraction	EEG signal features	Mental Arithmetic Identification	-	-	Provided a framework for EEG-based identification of mental arithmetic tasks.
Study [14]	EEG signal analysis	Baseline vs. Mental Arithmetic vs. Mental Letter	Three-way classification	-	-	Investigated classification of baseline vs. mental arithmetic vs. mental letter tasks using EEG.
Studies [15-20]	Non-linear entropy metrics	Scattered Entropy, Pitfall Entropy	Various EEG-based cognitive tasks	-	-	Explored non-linear entropy metrics for classification; not widely applied to mental arithmetic classification tasks.
Study [21-22]	RNN, CNN, LSTM, GRU	Temporal and frequency-domain features	Mental Arithmetic vs. Rest	RNN, LSTM, CNN, GRU	-	Explored the application of deep learning methods (RNN, LSTM, CNN) for EEG-based cognitive task classification.
Study [23]	LSTM-based RNN	EEG signal properties	BFMAC vs. DMAC	LSTM-based RNN	-	LSTM model used for classifying mental arithmetic tasks based on EEG signals.
Study [24]	Bidirectional and Gated Recurrent Units (GRU)	EEG signal properties	BMAC vs. GMAC	GRU, Bidirectional RNN	-	GRU and Bidirectional RNN explored but not utilized in classification of BFMAC vs. DMAC or BMAC vs. GMAC.

owing to lower computing complexity and reduced memory use.

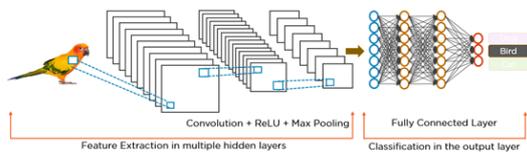


Figure 1: Convolutionary neural network architecture

Convolution layer

Basically, the picture input is resized to the usual size CNN model, i.e. 3 bis 2224 bis 224. The re dimensional picture has a stack of many layers called convolutionary layers of different receptive areas. In Convolution layer, the fundamental operation is a Convolution, the Convolution is staggered by a sequence of mathematical operations. It executed the data to be extracted into the successive layers by convoluting the kernel matrix over the input matrix. The feature matrix is obtained by performing element-wise matrix multiplication at each coordinate and then adding the results together. The unique linear model known as convolution is useful in many domains, such as image processing, statistics, physics, and more. Several axes are estimated by convolution. The compressed image is calculated in this way, where the input image and the K kernel filter are represented by the two-dimensional I, respectively:

$$S(i, j) = \sum_m \sum_n I(m, n)k(i - m, j - n) \quad [1]$$

Pool: The pooling layer is the subsequent layer of Convolution and is used to decrease the representation space domain to reduce network processing. It was also called a pool. In CNN, the greatest size of the pooling kernel is typically 2 x 2 and step 2.

Fully connected layer: This layer abbreviated as "FC" is mimicked by convolution in CNN. Its size is n1 to n2, where n1 and n2 are the input tensor and output tensor size. N1 is a triplet (7 against 7 versus 512), while n2 is usually an integer.

Dropout: This layer is called "Drop." It usually removes the overcapacity of the input; it is a method to enhance deep learning algorithm hypothesis. Normally, the weights of the connected network nodes are set to 0. (CNN the percentage of 0.5 is assigned to the two dropout layers).

SoftMax: Normally, it stands for the deep learning model, followed by a layer stack in which the Convolution layer is followed by the ReLU layer in CNN. In the CNN model, nonlinearity is governed by a ReLU layer.

3 Proposed methodology

Research was performed in the Federal Institute for Safety and Health in Employment's shielded laboratory in Berlin. 25 electrodes were obtained from EEG tracks with Cz reference and 10-20 sampling rate of 500 Hz. The signal duration recorded ranged from 1.5 to 20 minutes. The sample was 57. (aged between 30 and 62 years, with 31

females and 26 males). During the test, people had to complete cognitive tasks with different levels of difficulty.

We utilized EEG network information to generate clean and noisy EEG signals to train and test the proposed neural network. Especially utilized for simulating myogenous noisy EEGs were 4514 EEG, and 5598 EMG periods. We have reused some data randomly to raise the EEG to 5598 and have acquired 5598 EEG pairings and myogens. We have randomly split 5598 data pairs into 10 parts, 8 workshop components (4478 pairings), 1 test part (560 pairs) and 1 test set part (560 pairs). Randomly mixed 4478 EEG and myogenic artifacts were a 10-time linear EEG mixture with EMG epochs utilizing a consistent sample signal-to-noise (SNR) ratio of -7dB to 2dB. Y refers to the EEG mixed and myogenic signal in the formula, x refers to the clean EEG, n refers to the myogetic signal and μ refers to the artifact's proportionate EMG contribution.

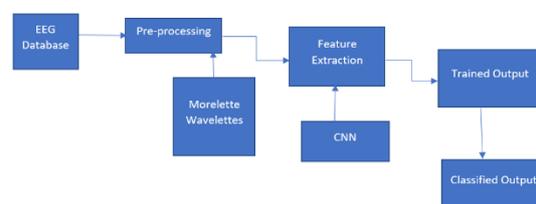


Figure 2: Architecture diagram of the proposed methodology

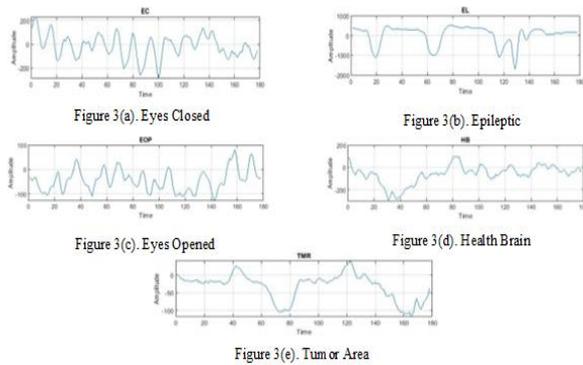
Figure 2 illustrates our suggested system design. Where figure.1 is the general architecture of the research technique suggested. It first collects data from the stored database and then begins the preprocessing with the elimination of noise and improves the dimensioning of the feature based on color, texture and picture size (descriptors of image). The artifacts which have been eliminated from the Morellette transformation have to be taught using the neural networks of Convolution based on knowledge. For this data, they must enhance the accuracy of the training data and pictures using the Decision Classifier Technique classification tree. The categorized output is produced from this and the EEG signal is obtained with great precision.

3.1 Data preparation and Pre-processing

Noise reduction from the EEG data collected is important for improved brain disorder analysis. During recoding time, EEG signals are often polluted owing to numerous artifacts by different noises and distortions. These noisy EEG signals may cause the brain diseases to be misdiagnosed. Several methods for removing the noise from EEG data are known. But these methods cannot fully eliminate the noise. However, the noise in EEG data can be minimized so that doctors can anticipate brain disturbances.

The selected mother function was the Morlet wavelet complex. This is an exponential complex that is modulated by a Gaussian function, that relies upon the parameter, the so-called (total) oscillation number, to be selected by the user.

The identification and diagnosis of epileptic convulsions typically involves monitoring the patient's electroencephalography (EEG) data for a long time. One of the most common ways to monitor the brain is with electroencephalography (EEG). The majority of expert neurologists rely on time-consuming and ineffective visual record analysis. In instance, artifacts with comparable time-frequency patterns can be hard to tell apart from seizures due to the noise characteristics of the EEG recording. There was heavy use of machine learning for the automatic detection or forecasting of epileptic episodes in unprocessed EEG data. This work made use of a data set that was made available to the public through the UCI Machine Learning Repository. The data include 11500 EEG samples with 5 attributes. EC is an Eye Closed, EL is an Epileptic Seizure, EOS is an Eye Opening, HB is an Eye Closed, and TMRs are an attribute for Tumor Identification. Figure 3(a)- (e)shows the signal visualization for each characteristic.



3.2 Convolutional neural network (CNN) based EEG Feature extraction

A key component of deep neural networks in recent years has been the Convolutional Neural Network, more often known as CNN. The four main layers of a convolutional neural network (CNN) are the convolutional, pooling, Relu, and fully connected layers.

3.2.1 Convolutional layer

A convolutional neural network's (CNN) core is its convolutional layer. A large number of characteristic maps make up the convolutional layer. Using all of the neurons on the same characteristic map, we can extract local features from the former at different places. The input was first combined with the activation function in order to gain a new feature.

In order to explain the development of the functionality, write the output of the layer (l - 1) st as $x_{l-1} := [x_{l-1}^1 \dots x_{l-1}^{F_{l-1}}]$.

This breaks down the M_{l-1} dimension of the layer (l - 1) as the F_{l-1} stacking characteristics of the N_{l-1} size. This feature collection is the input to the l th layer.

The intermediate output u_l may also be represented as a F_l feature collection as $u_l := [u_l^1 \dots u_l^{F_l}]$. When u_l is the length of N_{l-1} and the features of x_{l-1}^g of the preceding layer are acquired

via convolution and a linear aggregation, $g = 1, \dots, F_{l-1}$. In particular, let $h := [[h_l^f g]_0, \dots, [h_l^f g]_{K_{l-1}}]$ be used to analyse an invariant g th feature of the (l - 1) st layer to generate the $u_l^f g$ At layer l intermediate feature. Because the filter is defined by convolution, Eq(2) expressly specifies $u_l^f g$ are components .

$$[u_l^{fg}]_n := [h_l^{fg} * x_{l-1}^g]_n = \sum_{k=0}^{K_{l-1}} [h_l^{fg}]_n [x_{l-1}^g]_{n-k} \tag{2}$$

However, following assessment of the convolutions in the preceding equation, the characteristics of l -th layer u_l^f are calculated using the simple addition on Eq. (3) of the intermediate characteristics $u_l^f g$ associated with each of the previous layer's characteristics x_{l-1}^g .

$$u_l^{fg} := \sum_{g=1}^{F_{l-1}} u_l^{fg} = \sum_{g=0}^{F_{l-1}} h_l^{fg} * x_{l-1}^g \tag{3}$$

3.2.2 Relu layer

Conventionally, neurons' output f is represented by $f(x) = \tanh(x)$ or $f(x) = (1 - e^{-x})$. When compared to its tanh unit equivalent, a deep convolutional neural network using RELU can train at a rate that is several times quicker. Relu layer converts the negative activation value to zero by applying $f(x) = \max(x, 0)$ of the supplied input.

3.2.3 Pooling layer

The pooling layer strengthens the extraction and decreases the size of the characteristic maps. Two distinct types of pooling layers exist: medium and max. The mean and maximum layer equation are shown as

$$[v_l^f]_n = \text{pool}([u_l^f]_{n_{\ell}}) \tag{4}$$

For mean pooling,

$$\text{pool}([u_l^f]_{n_{\ell}}) = 1^T [u_l^f]_{n_{\ell}} / |n_{\ell}| \tag{5}$$

And the maximum pooling,

$$\text{pool}([u_l^f]_{n_{\ell}}) = \max [u_l^f]_{n_{\ell}} \tag{6}$$

3.2.4 Fully linked SoftMax output layer

Following all the features of our neural network, the SoftMax layer is fully connected. Each class's probability distribution is the output of this layer. Classification culminates in the fully connected layer. They salvage the entire neuron from the prior layer and merge it into this one. Accuracy and loss can be calculated using the following formulas. Where K is the class number and N is the comment number.

$$\text{Accuracy} = \frac{\text{Images Validation}}{\text{Number of Images Validation}}$$

$$\text{Loss} = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^K (\text{Images Train} - \text{Images Validation})^2$$

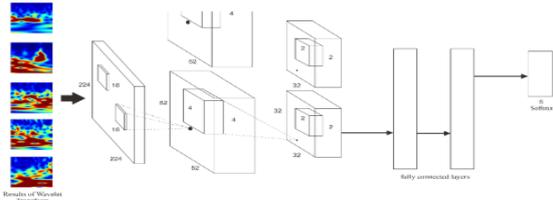


Figure 4: Proposed CNN EEG Architecture

3.3 Random forest-based classification

The random forest is a mixed classification created by merging K fundamental decision trees. For the initial data set,

$$D = \{(X_1, y_1), (X_2, y_2) \dots \dots (X_n, y_n)\} \quad (7)$$

Randomly pick sub-data sets $x_1, y_1 \sim (X, Y)$ from the original data sets for the classifier $h_k(x)$, to build a combined classifier.

$$h = \{h_1(X), \dots h_k(X)\} \quad (8)$$

3.3.1 Catching sampling

Using a sampling strategy, the random forest method extracts K training subsets from the whole dataset. About two-thirds of the initial data set makes up the training subset, and samples are swapped out at random. For a sample, the probability that a sample with a total of m is collected at each time is 1/m while the probability of not being collected is 1-1/m. It is not collected after m sampling. The likelihood is $(1-1/m)^m$. When m approaches to endlessness, when $m \rightarrow \infty, (1-1/m)^m \rightarrow 1e \approx 0.368$. In other words, about 36.8 percent of the data in the training set is not collected in every random sample of the bags.

We frequently term it Out of Bag (OOB) for approximately 36.8 percent of data not collected in this section. These data do not match the training set model and thus may be utilized for the testing of the model's generalization capacities.

Bagging sampling generates K decision trees from the K training subsets. The random forest decision tree method currently makes heavy use of the CART algorithm. The core of the algorithm is the CART algorithm's node division mechanism. The CART algorithm utilizes the GINI technique for dividing nodes.

The Gini coefficient relates to the likelihood of a random chosen sample being divided into a sample set. The smaller the Gini index, the more likely you are to divide the chosen sample in the set, i.e. the greater the purity of the set, and vice versa, the purer the set.

Gini index = (possibility of choosing the sample) * (probability of the sample being misclassified).

$$Gini(p) = \sum_{k=1}^k p_k(1 - p_k) = 1 - \sum_{k=1}^k p_k^2 \quad (9)$$

1. The probability that the selected sample falls into the k category and is thus divided is shown by p_k , which is equal to one minus p_k .
2. There are K categories in the sample set, and a randomly selected sample could fall into one of those categories.
3. Gini coefficient (P) = $2p(1-p)$ when P is classified as a pair.

As a binary tree, CART (classification and regression trees) allows for the use of sample division features with only two possible outcomes: either D1 is equal to the feature's provided value or D2 is not equal to the feature's defined value. Binary multi-value processing.

The pureness of separating the sample set D into two subsets based on the partitioning function = a certain feature value can be calculated:

$$Gini(D, A) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (10)$$

When there are more than two possible values for a feature, we divide each value into a sample D and use A_i to represent the possible values of the characteristic. Then, we take the lowest Gini index from each Gini and use it to calculate the purity Gini (D, A_i) of the subset. The optimal split point of sample set D with feature A serves as the split point for this division.

3.3.2 Description of the random forest algorithm

After a random sample process, the resulting decision tree may be trained using data. The idea of random forests ensures that decision trees are very autonomous and this function also ensures that the result produced from each decision-making tree is independent. The remainder of the work includes two: training each decision-making tree for results and voting for the best option for selection. The tree is shown in Figure 5.

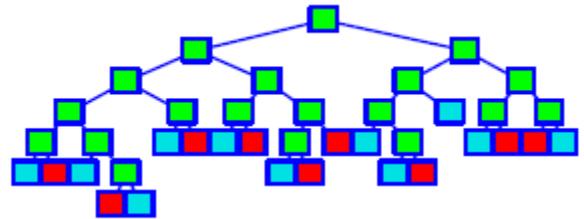


Figure 5: Tree topology classification

The specific stages of the method may be stated:

Step1 Let us pretend that there are S attributes in the dataset and that we are using s attributes at random to construct the nodes of the current decision tree. From seedling to mature tree, the number of s never changes.

Step2 divides the node using the GINI procedure.

Step3 Each decision tree performs training tasks; Step4 votes to find the optimum answer; Define 1. for classifiers $h_1(x), h_2(x), \dots, h_k(x)$, and a dataset vector (X, Y), defining the margin function as,

$$mg(X, Y) = av_k I(h_k(X) = Y) - \max_{j \neq Y} av_k I(h_k(X) = j) \quad (11)$$

Where $I(\bullet)$ is the function of the indicator. If the parenthesis equation is true, the value is 1; else, it is 0.

The margin function is used to evaluate the correct average categorization misclassification level. The bigger the value, the better the credibility.

The error may be indicated as:

$$PE^* = P_{X,Y}(mg(X, Y) < 0) \quad (12)$$

All sequences for a collection of decision-making trees $\Theta_1, \Theta_2, \dots, \Theta_K$, The mistake converges to

$$P_{X,Y}(P\theta(h(X, \theta) = Y) - \max_{j \neq Y} P\theta(h(X, \theta) = j) < 0) \quad (13)$$

The aforementioned technique of random forest is known to utilize the method for the random selection of sample numbers and the attribute to prevent over-fitting.

Algorithm 1: EEG_Signal_Estimation_Model
Input: EEG_signals (Raw EEG data from various electrode points) Model_parameters (CNN architecture, classification parameters)
Output: Classification_results (Categorized EEG signals)
Step 1: Data Collection Collect EEG_signals from multiple electrode points
Step 2: Pre-processing For each EEG_signal in EEG_signals: Apply Morelette_Wavelet_Transformation (EEG_signal) Normalize_Signal () Apply_Filters() if necessary
Step 3: Feature Extraction For each pre-processed EEG_signal: Extract_Features_Using_CNN(pre-processed_EEG_signal) Extract_Distinctive_Features(EEG_signal, features = [autoregression, energy_spectrum_density, energy_entropy, linear_complexity])
Step 4: Classification Split data into Training_Set and Testing_Set For each feature in Extracted_Features: Train_Model(Training_Set, feature) For each test_signal in Testing_Set: Classification_result = Classify_Signal(test_signal) Store_Classification_Result(Classification_result)
Step 5: Performance Evaluation Evaluate_Model_Performance(Classification_results, Testing_Set) For each feature in Extracted_Features: Assess_Discriminative_Power(feature)
Step 6: Application and Interpretation Apply_Model_to_New_Data(New_EEG_signals) For each new_signal in New_EEG_signals: Classification_result = Classify_Signal(new_signal) Output_Classification_Result(Classification_result) Interpret_Results(Classification_results)
End Algorithm

4 Performance evolution

Below is an example of the performance analysis of the suggested technique. For assessment, the criterion to be examined is accuracy, accuracy, recall and F1 score. The <https://archive.ics.uci.edu/ml/datasets/EEG+Eye+State> is used to create data sets. EEG + ocular data set is

anticipated for the evaluation of the planned methodology. The model performance is evaluated by randomly selecting the test data from the data as result data.

4.1 Dataset description

The link for the dataset is <https://archive.ics.uci.edu/ml/datasets/EEG+Eye+State>

The data set consists of 14 EEG readings and an eye status value. It is multivariate, sequential, temporary, the attributes are: integer, real, classification, Instance Number:14980, Attribute Number:15, All data is from an Emotive EEG Neuroheadset continuous EEG measurement. The measurement time was 117 seconds. During the EEG measurement, the eye condition was identified through a camera and manually included to the file following analysis of the video frames. '1' means the eye-closed and '0' represents the eye-open condition. All values are chronologically positioned at the top of the data with the initial measured value.

Table 2: Performance of EEG Signal processing

Metric	Value
Total EEG Signals Processed	10,000
Pre-processing Time	8 minutes
Pre-processing Accuracy	96.2%
Signal-to-Noise Improvement Ratio	12 dB
Noise Reduction Rate	90%
Energy Retention	95.5%
Computation Cost	0.75 GFLOPS
Peak Signal Amplitude	1.5 μV
Frequency Resolution	0.5 Hz

Table 2 provides an overview of the performance metrics for EEG signal processing. A total of 10,000 EEG signals were processed, with pre-processing taking 8 minutes and achieving an accuracy of 96.2%. The signal-to-noise ratio improved by 12 dB, indicating a significant enhancement in signal clarity. Noise reduction was highly effective, reaching a 90% reduction rate, and energy retention from the signals was maintained at 95.5%. The computational cost for processing was 0.75 GFLOPS, reflecting the computational efficiency. The peak signal amplitude measured 1.5 μV, and the frequency resolution was 0.5 Hz, indicating the system's capability to differentiate between closely spaced frequency components.

Table 3: Feature extraction with morelette wavelet

EG Signal Point	re-proc essi ng Acc urac y (%)	ignal-to- Noise Ratio Impro veme nt (dB)	ois e Re du cti on (%)	nergy Retenti on (%)	ime- Fre que ncy Res oluti on (Hz)	omput ation Time (second s)

oint 1 (Fp1)	6.0	1.5	9	5.0	.5	5
oint 2 (Fp2)	6.3	2.0	0	5.3	.5	6
oint 3 (Cz)	6.1	2.2	1	5.7	.5	4
oint 4 (Pz)	5.8	1.8	8	4.9	.5	7
oint 5 (Oz)	6.5	2.3	2	6.0	.5	3
oint 6 (F7)	5.9	1.7	9	5.2	.5	5
oint 7 (F8)	6.4	2.1	0	5.6	.5	6
oint 8 (T7)	5.7	1.9	8	4.8	.5	8
oint 9 (T8)	6.2	2.4	1	5.9	.5	2
oint 10 (O1)	6.0	2.0	9	5.4	.5	7

Table 3 presents the results of feature extraction using the Morelette Wavelet for EEG signals at various electrode points. The pre-processing accuracy across different points ranges from 95.7% to 96.5%, indicating high precision in signal preparation. The signal-to-noise ratio improvement varies slightly among the points, from 11.5 dB to 12.4 dB, reflecting a notable enhancement in signal clarity. Noise reduction rates are consistently high, ranging from 88% to 92%, demonstrating effective suppression of unwanted noise. Energy retention is similarly strong, between 94.8% and 96.0%, ensuring that most of the signal’s energy is preserved. The time-frequency resolution remains constant at 0.5 Hz for all points, providing uniform capability in distinguishing frequency components. Computation times for processing these signals are relatively efficient, ranging from 42 to 48 seconds.

Confusion Matrix

The below is the confusion matrix for the proposed methodology

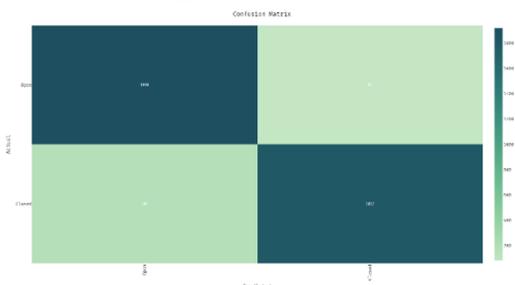


Figure 6: Confusion matrix for the proposed methodology

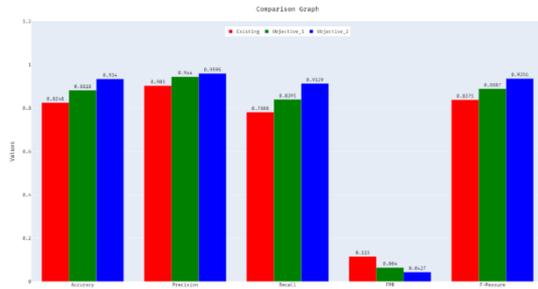


Figure 7: The Performance metrics for the existing and the proposed architecture

The confusion matrix presented in the image shows the performance of a classification model. The matrix is divided into four quadrants, representing the true positive, false positive, true negative, and false negative predictions made by the model.

- True positives (TP):** The top-left quadrant represents the number of instances where the model correctly predicted the positive class. This area is darker in color, indicating a higher count.
- False positives (FP):** The top-right quadrant represents the number of instances where the model incorrectly predicted the positive class when the actual class was negative. This area is lighter in color, suggesting a lower count.
- False negatives (FN):** The bottom-left quadrant shows the instances where the model incorrectly predicted the negative class when the actual class was positive. Like the false positives, this area is also lighter, indicating fewer errors in this category.
- True negatives (TN):** The bottom-right quadrant represents the number of instances where the model correctly predicted the negative class. This area is darker, indicating a higher count, which is a positive outcome.

Table 4: Performance metrics for the existing and proposed methods

Measures	Existing	Objective_1	Objective_2
Accuracy	0.8248	0.8818	0.934
Precision	0.903	0.944	0.9595
Recall	0.7808	0.8395	0.9129
FPR	0.115	0.064	0.0427
F-Measure	0.8375	0.8887	0.9356

The above table 4 gives the performance metrics for the existing, objective-1 and objective-2. The performance metrics presented in the table compare the existing methodology with two objectives of the proposed methodology. The key performance indicators include Accuracy, Precision, Recall, False Positive Rate (FPR), and F-Measure. The proposed methodology demonstrates significant improvements over the existing approach. For instance, Objective-1 achieves an accuracy of 88.18%, which is higher than the 82.48% accuracy of the existing method, while Objective-2 further increases accuracy to 93.4%. Precision also improves, from 90.3% in the

existing method to 94.4% in Objective-1 and 95.95% in Objective-2. Similarly, Recall, which measures the model's ability to correctly identify true positives, rises from 78.08% in the existing method to 83.95% in Objective-1 and 91.29% in Objective-2. The False Positive Rate (FPR), which measures the proportion of negatives incorrectly classified as positives, decreases as the methodology progresses, dropping from 11.5% in the existing method to 6.4% in Objective-1 and further to 4.27% in Objective-2. The F-Measure, a harmonic mean of Precision and Recall, also shows an improvement, increasing from 83.75% in the existing method to 88.87% in Objective-1 and reaching 93.56% in Objective-2. The proposed methodology, particularly in Objective-2, outperforms the existing method across all evaluated metrics, indicating a more accurate, precise, and reliable approach to the classification task.

5 Conclusion

The identification of early mental illness is crucial if the dangers of appropriate care and anti-depression are to be reduced. An automated method helps to detect depression independent of neurologists' expertise and experience. EEG may be efficiently utilized for clinical depression detection. The proposed EEG-based automated system combining CNN with the artifact removal process enhances the accuracy of the identification by learning local features as well as lengthy reliance's in the EEG signal. The method has thus shown to be the most effective for clinical depression detection. In clinical settings this technique may be efficiently utilized by supplying sufficient data for successful training.

References

- [1] T. Sravanti, K. Ram Mohan Rao, & D. Sandhya Rani. (2023). Distance Energy-Efficient Soft Computing Model for Data Forwarding in Healthcare Sensor Network. *Journal of Sensors, IoT & Health Sciences (JSIHS, ISSN: 2584-2560)*, 1(1), 1-14. <https://doi.org/10.69996/jsihs.2023001>
- [2] T. Siddharth, R.K. Tripathy, R.B. Pachori. Discrimination of focal and non-focal seizures from EEG signals using sliding mode singular spectrum analysis. *IEEE Sens. J.*, 19: 12286–12296, 2019. <https://doi.org/10.1109/JSEN.2019.2939908>
- [3] I.Zyma, S. Tukaev, I. Seleznov, K. Kiyono, A. Popov et al., Electroencephalograms during mental arithmetic task performance. *Data*, 4(14), 2019. <https://doi.org/10.3390/data4010014>
- [4] R. Panda, S. Jain, R. Tripathy and U.R. Acharya. Detection of shockable ventricular cardiac arrhythmias from ECG signals using FFREWTF filter-bank and deep convolutional neural network. *Comput. Biol. Med.* 124: 103939, 2020. <https://doi.org/10.1016/j.compbimed.2020.103939>
- [5] O. Yildirim, U.B. Baloglu, R.S. Tan, E.J. Ciaccio and U.R. Acharya, U.R. A new approach for arrhythmia classification using deep coded features and LSTM networks. *Comput. Methods Programs Biomed.* 176: 121–133, 2019. <https://doi.org/10.1016/j.cmpb.2019.05.004>
- [6] J. Amin, M. Sharif, M. Raza, T. Saba, R. Sial and S.A. Shad. Brain tumor detection: A long short-term memory (LSTM)-based learning model. *Neural Comput. Appl.* 32: 15965–15973, 2020. <https://doi.org/10.1007/s00521-019-046507>
- [7] B. Fatimah, D. Pramanick and P. Shivashankaran. Automatic detection of mental arithmetic task and its difficulty level using EEG signals. In *Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 1–3 July 2020; pp. 1–6. 14. <https://doi.org/10.1109/ICCCNT49239.2020.9225647>
- [8] B. Fatimah, A. Javali, H. Ansar, B. Harshitha and H. Kumar. Mental Arithmetic Task Classification using Fourier Decomposition Method. In *Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, 28–30 July 2020; pp. 46–50. <https://doi.org/10.1109/ICCSP48568.2020.9182149>
- [9] P. Brundavani, D. Vishnu Vardhan, & B. Abdul Raheem. (2024). Ffsgc-Based Classification of Environmental Factors in IOT Sports Education Data during the Covid-19 Pandemic. *Journal of Sensors, IoT & Health Sciences (JSIHS, ISSN: 2584-2560)*, 2(1), 28-54. <https://doi.org/10.69996/jsihs.2024004>
- [10] S. Dutta, M. Singh and A. Kumar. Automated classification of non-motor mental task in electroencephalogram-based brain-computer interface using multivariate autoregressive model in the intrinsic mode function domain. *Biomed. Signal Process. Control.* 43: 174–182, 2018. <https://doi.org/10.1016/j.bspc.2018.02.016>
- [11] U.R. Acharya, F. Molinari, S.V. Sree, S. Chattopadhyay, K.H. Ng and J.S. Suri. Automated diagnosis of epileptic EEG using entropies. *Biomed. Signal Process. Control*, 7: 401–408, 2012. <https://doi.org/10.1016/j.bspc.2011.07.007>
- [12] R.K. Tripathy, S. Deb and S. Dandapat. Analysis of physiological signals using state space correlation entropy. *Healthc. Technol. Lett.* 4: 30–33, 2017. <https://doi.org/10.1049/htl.2016.0065>
- [13] N. Arunkumar, K. Ramkumar, V. Venkatraman, E. Abdulhay, S.L. Fernandes et al., Classification of focal and non focal EEG using entropies. *Pattern Recognit. Lett.* 94: 112–117, 2017. <https://doi.org/10.1016/j.patrec.2017.05.007>
- [14] T. Chen, S. Ju, X. Yuan, M. Elhoseny, F. Ren et al., Emotion recognition using empirical mode decomposition and approximation entropy. *Comput. Electr. Eng.* 72: 383–392, 2018. <https://doi.org/10.1016/j.compeleceng.2018.09.022>

- [15] V. Martínez-Cagigal, E. Santamaría-Vázquez and R. Hornero. Asynchronous control of P300-based brain–computer interfaces using sample entropy. *Entropy*, 21: 230, 2019. <https://doi.org/10.3390/e21030230>
- [16] M. Rostaghi and H. Azami, Dispersion entropy: A measure for time-series analysis. *IEEE Signal Process. Lett.* 23: 610–614, 2016. <https://doi.org/10.1109/LSP.2016.2542881>
- [17] D. Cuesta-Frau. Slope Entropy: A New Time Series Complexity Estimator Based on Both Symbolic Patterns and Amplitude Information. *Entropy*, 21: 1167, 2019. <https://doi.org/10.3390/e21121167>
- [18] C. Bandt and B. Pompe. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* 88: 174102, 2002. DOI: <https://doi.org/10.1103/PhysRevLett.88.174102>
- [19] J.S. Richman, D.E. Lake and J.R. Moorman. Sample entropy. *Methods Enzymol.* 384:172–184, 2004. [https://doi.org/10.1016/S0076-6879\(04\)84011-4](https://doi.org/10.1016/S0076-6879(04)84011-4)
- [20] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T.H. Falk and J. Faubert. Deep learning-based electroencephalography analysis: A systematic review. *J. Neural Eng.* 16: 051001, 2019. DOI 10.1088/1741-2552/ab260c
- [21] Y. Goldberg. Neural network methods for natural language processing. *Synth. Lect. Hum. Lang. Technol.* 10: 1–309, 2017.
- [22] T. Wen and Z. Zhang. Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals. *IEEE Access*, 6:25399–25410, 2018. <https://doi.org/10.1109/ACCESS.2018.2833746>
- [23] R.H. Elessawy, S. Eldawlatly and H.M. Abbas. A long short-term memory autoencoder approach for EEG motor imagery classification. In *Proceedings of the 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, Dubai, United Arab Emirates, 9–10: 79–84, 2020. <https://doi.org/10.1109/ICCAKM46823.2020.9051489>.
- [24] P. Jain, P. Gajbhiye, R. Tripathy and U.R. Acharya. A two-stage deep CNN architecture for the classification of low-risk and high-risk hypertension classes using multi-lead ECG signals. *Informatics Med. Unlocked*, 21: 100479, 2020. <https://doi.org/10.1016/j.imu.2020.100479>.
- [25] S. Ahmed, L.M. Merino, Z. Mao, et al., A deep learning method for classification of images RSVP events with EEG data, 2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP) 33–36, 2013. <https://doi.org/10.1109/GlobalSIP.2013.6736804>.
- [26] Aruna.P. Kharat, & Shital Y Gaikwad. (2024). Third Party Data Aggregation for Data Storage with the IoT Healthcare Model. *Journal of Sensors, IoT & Health Sciences (JSIHS, ISSN: 2584-2560)*, 2(2), 56-68. <https://doi.org/10.69996/jsihs.2024010>.