IAES International Journal of Robotics and Automation (IJRA)
Vol. 11, No. 3, September 2022, pp. 250~262
ISSN: 2722-2586, DOI: 10.11591/ijra.v11i3.pp250-262

Feature extraction techniques for cognitive stimuli-based electroencephalogram signals: an experimental analysis

Jessy Jane Christy J., E. Grace Mary Kanaga
Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamilnadu, India.

Article Info

Article history:

Received Mar 15, 2021
Revised Jul 5, 2022
Accepted Jul 20, 2022

Keywords:

Brain computer interface
Cognitive stimuli classification
Electroencephalogram
Feature extraction

ABSTRACT

Over the past decades, brain-computer interface (BCI) has gained a lot of attention in various fields ranging from medicine to entertainment, and electroencephalogram (EEG) signals are widely used in BCI. Brain-computer interface made human-computer interaction possible by using information acquired from EEG signals of the person. The raw EEG signals need to be processed to obtain valuable information which could be used for communication purposes. The objective of this paper is to identify the best combination of features that could discriminate cognitive stimuli-based tasks. EEG signals are recorded while the subjects are performing some arithmetical based mental tasks. Statistical, power, entropy, and fractional dimension (FD) features are extracted from the EEG signals. Various combinations of these features are analyzed and validated using random forest classifier, K-nearest neighbors (KNN), multilayer perceptron, linear discriminant analysis, and support vector machine. The combination of entropy-FD features gives the highest accuracy of 90.47% with the KNN algorithm when compared to individual entropy and FD features which achieves 79.36% with random forest classifier, multilayer perceptron, and 82.53% with linear discriminant analysis, respectively. Our results show that the hybrid of entropy-FD features with KNN classifier can efficiently classify the cognition-based stimuli.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Jessy Jane Christy J.
Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences
Coimbatore, Tamilnadu, India
Email: jessyjane@karunya.edu.in

1. INTRODUCTION

Brain-computer interface (BCI) acts as an artificial and alternative output channel for the brain which is similar to the normal output channels like muscles and peripheral nerves. Hence, BCI is defined as "*a brain-computer interface is a communication system that does not depend on the brain's normal output pathways of peripheral nerves and muscles*" [1]. BCI requires two adaptive controllers: A brain from where the electrical activity is recorded and a system that converts this electrical activity into control commands. Gain [2] discussed the function of various lobes of the brain in human behaviors. The temporal lobe is responsible for language processing, the occipital lobe for visual processing, the parietal lobe for sensations, frontal lobe for cognition and emotions. BCI has an input, an output, and a translation algorithm. The input is the features of the signals recorded from the brain. Some of these features are the time-domain and frequency domain. The translation algorithm such as linear/nonlinear equations, neural networks, and converts these input features into control signals. These output control signals are used to control or operate any device [3]. Event-related potentials are responses from the brain under certain conditions like giving external stimuli. There are two types of event related potentials (ERPs): exogenous and endogenous. Exogenous ERPs are responses generated from the brain spontaneously as a result of external stimulus regardless of the subject's thinking or actions. Endogenous ERPs are responses that are generated while the subjects try to respond to external stimuli through thinking, imagination, or emotions. For example, solving the given mathematical problem. These kinds of ERPs are also called cognitive ERPs [4]. Most of the BCI application consists of the following steps: preprocessing, channel selection, feature extraction, feature optimization, and classification [5]. The main challenge while dealing with electroencephalogram (EEG) signals is extracting appropriate features because of the non-stationary property of EEG signals and the number of channels used. Since the EEG brain signals are non-stationary and recorded in the time-domain, it is necessary to analyze the EEG data from multiple domains which gives enhanced information about the time and frequency-related information of the recorded signals. The objective of this paper is to give an overview of various existing techniques feature extraction techniques and to extract features from time as well as frequency domain and analyze the impact of various combinations of features over the classification accuracy to find out which combination performs better on cognitive stimuli. The categories of features extracted in this study are statistical features (S), power features (P), entropy features (E), and fractional dimension (F) features. Two methods are employed to analyze the performance of the classifiers based on the features: i) combine each feature with respect to their categories and compare the accuracies of each category and ii) combine each category of features in multiple combinations and compare the accuracies of each combination of feature categories.

In this study, we use various classifiers such as random forest classifier (RFC), K-nearest neighbors (KNN), multilayer perceptron (MLP), linear discriminant analysis (LDA), and support vector machine (SVM) to classify EEG data into thirteen different classes of mental tasks (i.e., thirteen stimuli). Our results show that the combination of entropy-FD features employed in method (ii) with KNN gives the highest accuracy of 90.47%.

2. RELATED WORKS

Feature extraction methods are necessary to get the salient features from time-domain EEG signals which effectively classify the data. The features that can be obtained from the preprocessed signals belong to time-domain, frequency-domain, time-frequency domain, and spatial domain.

2.1. Time-domain features

In the time domain, power is analyzed with respect to time. Here, event-related potentials invoked by external stimuli act as a command. Examples for time-domain include P300 potentials, and slow cortical potentials [1]. Choi *et al.* [6] has examined the brain responses across different regions while classifying mathematical and baseline tasks which is an endogenous paradigm i.e., without external stimuli. Ear-EEG is used for recording these self-modulated signals from the brain while performing mathematical tasks. Here among the statistical features, mean, standard deviation, mean absolute value (MAV) of the first and second difference of raw and standardized signals are widely used. Nawaz *et al.* [7] proposed that time domain-based statistical features and SVM with RBF kernel give better accuracy when compared to power, wavelet, FD, entropy features. However, separating noise from the signal is considered a challenging task with time-domain features alone [8].

2.2. Frequency domain features

In the frequency domain, power is analyzed with respect to frequency. Here, the amplitude of frequency sub-bands acts as a command, examples for frequency domain include rhythms like α, and β [1]. Power-based feature extraction deals only with frequency sub bands that make it a frequency domain feature extraction [9]. Fast Fourier transform (FFT) is used for spectral analysis of a given signal which is stationary which makes it not suitable for EEG signals. FFT involves applying discrete FFT on the signal to find its frequency [10]. Wang *et al.* [11] utilizes frequency domain features using canonical correlation analysis (CCA) algorithm and power spectral density (PSD) techniques for more optimization. Stimulus frequency was identified using the above techniques. Frequency domain feature extraction is widely suggested, and PSD is a widely used technique for extracting frequency domain features. According to the Wiener-Khintchine theorem, PSD is calculated by applying Fourier transform on the autocorrelation function $\hat{R}_x(\mathrm{m})$ [12], or equivalently, PSD is calculated by taking the average of the squared magnitude of the Fourier transform [13]. Akrami *et al.* [14] proposed that logarithmic PSD is considered as the best method suitable for recognizing patterns from EEG. Shen *et al.* [15] proposed a method called WPT-BED to classify the cognitive tasks based on judgment where wavelet packet transform with db4 wavelet is used to decompose the signal into frequency bands and bispectrum features are extracted from the decomposed frequency bands. Then the sub-bands are reconstructed and bispectral eigenvalues of differential signals (BED) are used to optimize bispectral features from the resultant time-domain signal. The optimized features are then classified using SVM. PSD might ignore certain frequency features such as phase which is very important in processing EEG signals. BED features improve classification accuracy by considering an ample amount of information that is not considered in PSD.

2.3. Time-frequency domain features

Wavelet is suitable for analyzing non-stationary signals like EEG. Wavelet-based technique deals with both temporal and frequency ranges hence make it both time and frequency domain feature extraction technique [9]. The various types of wavelet transforms are discrete wavelet transform (DWT), continuous wavelet transform (CWT), tunable Q-factor wavelet transform (TQWT), dual tree-complex wavelet transform (DT-CWT) which is used in splitting signals into various frequency sub-bands (signal decomposition). The main drawback of FFT is it extracts the frequency features by taking the average over the entire signal without considering the difference in the time domain, which makes FFT only suitable for extracting frequency related feature from only signals which are stationary in the time domain. Since EEG is non-stationary, short time Fourier transform (STFT) is used for representing time-frequency features of the signal. The idea behind STFT is that the entire signals are divided into segments and apply FFT to segmented signals which are stationary in each segment. Hence, it provides frequency related information with respect to time interval [11]. STFT involves applying windows to the raw signals and the FFT is applied to the resultant signals [10]. The major drawback of STFT is that the window size is fixed which limits its capability to distinguish among various features and provides limited information regarding the location of frequency changes. On the other hand, wavelet transform or decomposition represents the features in a time-frequency domain called scalograms by decomposing the signals into various sub bands. Wavelet transform/decomposition helps to find the location of frequency changes in each sub band [8]. The drawback of STFT could be overcome by a CWT. In CWT, the window size can be changed based on the spectral component. CWT provides the “high localization of time in high-frequency EEG signals” as well as a large number of waveforms apart from sinusoidal waveform [16]. The major drawback of CWT is the scaling value ‘a’ and translation value ‘b’ change continuously which yields a lot of unrelated information. This drawback can be overcome by DWT which represents features at multiple levels [17]. DWT is used along with Daubechies 4^{th} order wavelet as mother wavelet to decompose signals into approximation and detailed coefficients. These coefficients are decomposed recursively which results in the high pass and low pass filters to get the frequency sub bands (between 0 and 50 Hz). Daubechies 4^{th} order wavelet (db4) is widely used because it resembles EEG waveforms. DT-CWT is similar to DWT but has better approximate shift variance and anti-aliasing than DWT [12]. Wavelet decomposition (WD) decomposed the raw signals only into lower frequency sub bands, but high frequencies are detected while performing mental tasks. Another drawback is the deterioration of feature quality due to the quick reduction of wavelet coefficients. To overcome this issue, the wavelet packet decomposition (WPD) is used to decomposed the raw signals into both lower and higher frequency sub-bands [18].

Mini *et al.* [19] adopted DWT, WPD, and DWPD which is a combination of the DWT and WPD wavelet decomposition techniques where DWT was applied to detailed coefficient and WPD was applied to the approximate coefficient for the further decomposition of signals. WPD gives high accuracy when compared to other methods. Wavelet packet node reconstruction (WPNR) and wavelet node reconstruction (WNR) are responsible for reconstructing signals from their respective nodes [18]. Shen *et al.* [15] proposed a method that utilizes hybrid EEG features for the identification of DRDS tasks. Features are extracted using the one-vs-one method from the particular channels that had been selected using CSG techniques. The extracted features are time-frequency domain features. Chatterjee *et al.* [9] proposed a method where features are extracted based on wavelets and power. The feature extraction techniques are wavelet-based energy-entropy, wavelet-based root means square, PSD-based band power, PSD-based average power, and their combinations. It is concluded that wavelet-based features such as Wavelet-based energy-entropy, wavelet-based root mean square lead to better performance than power-based features with classifiers such as logistics and SVM. Chatterjee and Bandyopadhyay [20] concluded that wavelet-based energy-entropy as a feature gives better accuracy when compared to statistical features and power features. Murugappan *et al.* [21] proposed certain energy-based features such as recoursing energy efficiency (REE), logarithmic REE (LREE), and absolute logarithmic REE (ALREE) and classified these features using two linear classifiers such as KNN and LDA. Here KNN performs better with a maximum accuracy of 83.26% with ALREE features. Hence it is concluded that energy features perform better than power and conventional features [22]. Harpale and Bairagi [13] suggested that wavelet-based analysis gives better accuracy for feature extraction techniques. Wavelet-based decomposition and features are considered as better than FFT and STFT because wavelet decomposition separates the signal into detailed and approximation coefficients iteratively where we can get improved details of signal and better time-frequency representation while the latter seems to give the least time/frequency information and least information about signals [22].

2.4. Hybrid features

Wei *et al.* [12] proposed a method in which time-domain, frequency-domain, and non-linear analysis features are extracted and used. In this method, raw EEG signals are preprocessed by filtering and decomposition into sub-bands using DT-CWT. Then the time-domain features are extracted using MAV, frequency-domain features by PSD, and non-linear analysis by fractional dimension (FD) and differential entropy (DE). Then these four features along with the best two frequency bands are given as input to the simple recurrent unit and by ensemble methods like voting and then the weighted average has been done to accomplish the classification task [12]. Suleiman and Fatehi [10] proposed that time-frequency-space analysis performed better than time/frequency domain and time-frequency domain. In multichannel EEGs, space-time-frequency (STF) is used for selecting signals from the appropriate regions or channels. This can be done by applying STFT on multiple electrodes to choose a channel. The selected channel is then combined with one of the channels and is sent as an input to the MLP which uses the back propagation algorithm. But in this method, no specific method was mentioned to select the best combination of channels which is necessary for extracting STF features [10]. Bajaj *et al.* [23] utilizes wavelet transform with statistical features. The raw EEG signals are decomposed into high and low pass sub bands using TQWT and features are extracted from these sub bands using statistical feature extraction methods like Horthy mobility (HM), minima, maxima, mean and standard deviation. Bandil and Wadhwan [24] proposed a method for epileptic classification in which DWT is used to decompose the EEG signals into 5 sub-bands with db4 mother wavelet. Then the signals are standardized to reduce the impact of higher estimated factors over the lesser ones. Morphological features like AR coefficient, and PSD, and statistical features like mean, median, mode, and entropy features are extracted. Harpale and Bairagi [13] proposed a method that classifies seizure and non-seizure EEG signals using hybrid features. Features are extracted from both time and frequency domains such as mean, coefficient of variation (COV), root mean square (RMS), kurtosis, and PSD respectively. By applying pattern adapted wavelet transform, features like mean, RMS, PSD, and standard deviation are extracted. Liu *et al.* [25] proposed a method in which features are extracted from the time domain, frequency domain, time-frequency domain, and multi-electrodes. Relevant features from all of these domains are selected based on maximum relevance and minimum redundancy as a feature selection method. Features are also extracted from the appropriate combination of channels that leads to better accuracy. Multi electrode features focus on extracting features based on the interconnections between electrodes that are attached to different brain regions [25]. Garg and Verma [8] proposed wavelet-based feature extraction techniques for classifying scalograms using neural networks. CWT is used to decompose signals into scalograms for better time-frequency representation of signals. Then scalogram images were fed into a convolutional neural network (CNN) where the spatial feature i.e., power of each frequency band in the scalogram images, is extracted in the pooling layer [16]. Various feature extraction techniques are summarized in Table 1. Based on various studies discussed above, hybrid features are considered to improve accuracy when compared to using a single feature or combination at a time.

Table 1. Various feature extraction techniques

Author	Preprocessing	Domain	Features	Feature Extraction Techniques	Classification Algorithms	Performance
[15]	Bandpass filter, ICA	Time-frequency domain	Wavelet	CSG and OVO	SVM with RBF kernel	Accuracies for five subjects - 94.67%, 91.33%, 0.00%, 87.67%, 73.83%
[11]	High-pass filter, noise removal	Frequency domain	Power	CCA and PSD	Voting mechanism	Accuracy exceeds 72.84%
[12]	DT-CWT	Time-domain, frequency domain, non-linear analysis	Hybrid	MAV, PSD, FD and DE	Simple recurrent units and ensemble methods such as voting and weighted average	MAV–79.22%, PSD – 78.29%, FD – 77.22%, DE – 80.02%
[6]	Bandpass filter, fourth-order Butterworth filter	Frequency-domain		CSP	sLDA	Accuracy- 75.6%

Table 1. Various feature extraction techniques *(continue)*

Author	Preprocessing	Domain	Features	Feature extraction techniques	Classification Algorithms	Performance
[9]	Elliptic bandpass filter	Time-frequency domain	Wavelet	Wavelet-based energy-entropy, wavelet-based root mean square	Logistic	ROC - 0.918 Recall - 0.821 Precision - 0.821 Accuracy - 82.14
					SVM	ROC - 0.850 Recall - 0.850 Precision - 0.852 Accuracy – 85
					MLP	ROC - 0.917 Recall - 0.836 Precision - 0.839 Accuracy - 83.57
[10]	Notch filters, high and low pass filters	Space-time-frequency-domain and time-frequency domain	Hybrid	FFT, STFT	MLP	Classification accuracy – 99% (two tasks) and 96%(three tasks)
[23]	TQWT	Time-domain	Hybrid	Hjorth mobility, minima, maxima, mean and standard deviation	ELM	Accuracy – 91.842%
[18]	Notch filter	Time-frequency domain	Hybrid	Interchannel correlation coefficient and statistical features	SVM with polynomial kernel	Accuracy – 86%
[24]	DWT	Time-domain and Frequency-domain	Hybrid	Morphological and statistical features	ANN	Accuracy-99%
[13]	ICA	Time-domain, frequency-domain, and time-frequency domain	Hybrid	Standard deviation, variance, RMS, kurtosis, SUM, POW, and PSD	Fuzzy inference system	Accuracy - 96.48%
[7]	Time-window segmentation	Time-domain	Statistical, FD	Mean, SD, MAV	SVM with RBF kernel	Accuracy – 77.62%, 78.96%, 77.60% (valence, arousal, dominance)
[25]	High pass filter	Time-domain, frequency-domain, and time-frequency domain	Hybrid	Mean, SD, MAV, HOC,FD, Hjorth, NSI, PSD, REE, RMS, entropy, multi electrode features such as DA, RA, MSCE	Random forest	Accuracy – 71.23%,69.9%(Arousal and Valence)
[14]	Bandpass filter	Frequency-domain	Hybrid	Logarithmic PSD	Neural network	Not mentioned
[8]	Bandpass filter	Time-frequency domain and spatial domain	Hybrid	Wavelet, CWT, spatial, feature extraction	GoogleNet based CNN	Maximum accuracy of 92.19%
[22]	Average mean reference (AMR)	Time-frequency domain	Wavelet and entropy	DWT	FCM, FKM	Not mentioned
[21]	Surface Laplacian	Time-frequency domain	Wavelet and energy	DWT, ALREE	KNN	83.26%
[5]	High pass filter, low pass filter and ICA	Frequency domain	Power	BED	SVM	84.38%
[26]	High pass filter, low pass filter, and ICA	Time-domain, frequency-domain, and time-frequency domain	Hybrid	Statistical features, FD, Hjorth features, PSD, Coif1 wavelet, energy, and entropy	Unsupervised Hyperplane partitioning	Maximum accuracy of 77.53%
[11]	Bandpass filter	Time-frequency domain	Wavelet	STFT	CNN	90.59%

3. METHODOLOGY

In the current study, we have implemented machine learning techniques to classify cognitive-based stimuli (arithmetic mental tasks) using the EEG data. A total of 13 stimuli are used for each subject and EEG signals are recorded while performing the mental calculation. The overall framework of the study is depicted in Figure 1. Firstly, the signals are segmented into segments of 10 seconds. Secondly, the four categories of features are extracted from the time and frequency domain. Thirdly, the different combinations are made and given as input to the classifiers. Finally, the best combination of features along with the classifier is noted to classify the cognitive stimuli-based EEG signals.

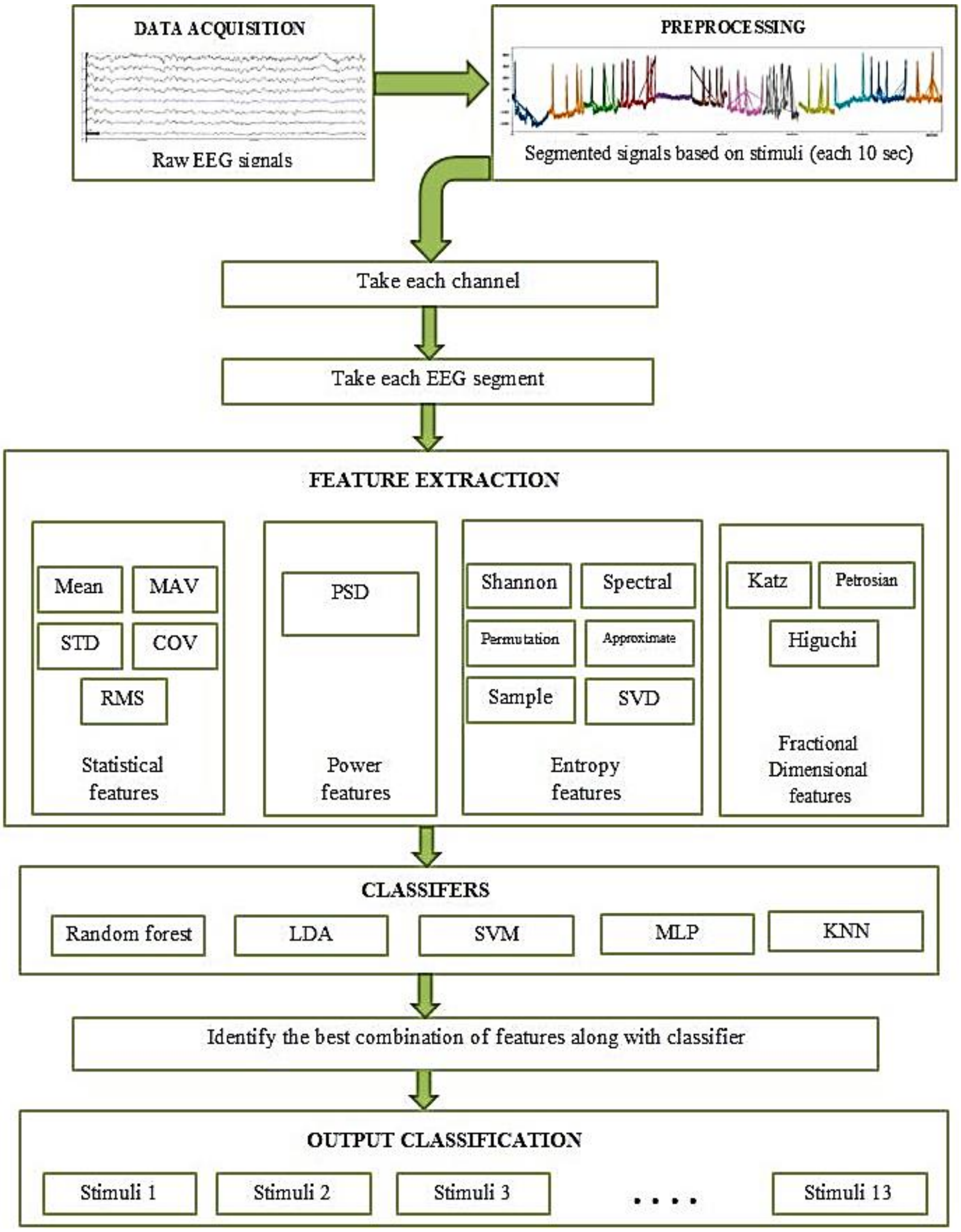

Figure 1. Framework of the study

3.1. Data acquisition and pre-processing

EEG is a non-invasive technology used to measure brain activity. In the project, a gTech recorder which consists of 16 electrodes was used to measure brain activity. The EEG signals were measured across 16 different channels such as FP2, F4, C4, P4, F8, T4, T6, O2, FP1, F3, C3, P3, F7, T3, T5, O1, and Ref. The electrodes were placed according to the international standard 10-20 positioning system [10]. A sampling frequency of 512 Hz was used. Sensitivity was set to 2.5μV/mm. The low pass filter of 1.0Hz was used to remove high-frequency noise [7]. The notch filter was set to 50 Hz to remove exceeded power supply [7].

The subject is made to sit on a chair in a comfortable position. The electrodes were attached to the scalp by using a gel (Ten20 Conductive gel). Then tapes were attached to the electrodes to prevent them from moving. The reference electrode was placed in the right ear. The readings were taken from various healthy subjects (age group between 22 and 25). The simple mental arithmetic tasks (i.e., basic addition, subtraction, multiplication, and division problems such as 10+5, 5-3, 5*5, and 6/2,) are shown and used as the cognitive stimuli and 2 trials were taken for each subject. Each event was recorded with a time duration of 10 seconds. The cognitive stimuli consisted of the 13 mental arithmetic tasks and each lasted for 10 seconds with a time break of 10 seconds. At the start of the experiment, there was a time break of 120 seconds. The raw EEG signals are processed in such a way that only the signals from the performance period are taken into account and the signals that are recorded during the resting state will be discarded. Therefore, the raw EEG data is segmented for every 10 seconds based on the target class which in this case is the stimuli. The feature extraction takes place on these segmented signals.

3.2. Feature extraction

The main aim of feature extraction is to obtain salient features from the EEG signals that could effectively classify the stimuli. In this study, four categories of features such as statistical, power, entropy, and FD features are extracted to analyze which feature set performs efficient classification. Each 10 seconds trails are further segmented into 2 seconds pieces and the following feature extraction techniques are applied.

3.2.1. Statistical features

In this study, six statistical features mean [7], standard deviation [7], mean absolute value [12], root mean square [13], coefficient of variation [13].

a. Mean

$$\mu_X = \frac{\sum_{n=1}^{N} X(n)}{N} \tag{1}$$

where 'μ_X denotes the mean of the data,'X(n)' denotes the data points and 'N' denotes the number of data points.

b. Standard deviation

$$\sigma_X = \sqrt{\frac{\sum_{n=1}^{N} (X(n)-\mu_X)^2}{N}} \tag{2}$$

where 'σ_X' denotes the standard deviation of the data, 'X(n)' denotes the data points, 'μ_X denotes the mean of the data points and 'N' denotes the number of data points.

c. Mean absolute value

MAV is calculated by taking the average of the absolute value of the data points.

$$M = \log\left(\frac{1}{N}\sum_{n=1}^{N} |x(n)|\right) \tag{3}$$

where 'x(n)' denotes the data points and 'N' denotes the number of data points.

Root mean square (RMS)

RMS is calculated by taking the square root of the averaged squared value of the data points [13].

$$RMS = \sqrt{\frac{1}{T}\int_0^T (x(n))^2\, dt} \tag{4}$$

d. Coefficient of Variation (COV)

$$COV = \frac{\sigma_X}{\mu_X} \tag{5}$$

where 'μ_X denotes the mean of the data and 'σ_X' denotes the standard deviation of the data.

3.2.2. Power features

Power features include features extracted from the frequency domain of EEG signals. One of the widely adopted techniques for extracting power features is PSD. PSD is calculated using Welch's method by taking the average of the Fourier transform of the segmented blocks of the original signal [11],

$$\hat{s}_x \triangleq \frac{1}{K}\sum_{K-1}^{m=0} P_{x_n}(m) \quad (6)$$

where $P_{x_n}(m)$ is called periodogram of each block which is the result of the FFT applied over the segmented signals and 'K' represents the total number of segmented blocks in the original signals.

3.2.3. Entropy features

Entropy is recommended to extract non-linear features of EEG signals [7]. In this study, six categories of entropy features are extracted for the non-linear analysis of EEG data.

a. Shannon entropy

Shannon entropy is a measure of uncertainty present in the value. It quantifies the amount of information that a particular variable or data holds over the result [27]. It is defined as (7),

$$H(X)=-\sum_{i=1}^{n} P_i \log_2 P_i \quad (7)$$

where 'n' denotes the number of data points and 'P_i' denotes the probability of a data point.

b. Spectral entropy

Spectral entropy (SE) represents the proportions of which power spectrum of the EEG signal is made which consists of 'flats' and 'peaks' distribution [28]. It is calculated by measuring Shannon's entropy for PSD [7] by (8),

$$SE= -\sum_{f_n}^{f=0} PSD(f) \log_2 (PSD(f)) \quad (8)$$

where 'f' is half of the sampling frequency [7].

c. Permutation entropy

Permutation entropy (PE) quantifies the information by analyzing the patterns of ranks of values present in the time series data [29]. It is defined by (9),

$$PE=-\sum_{i=1}^{n!} p'_i \log_2 (p'_i) \quad (9)$$

where p'_i denotes the number of times the pattern of a particular sequence occurs in a variable.

d. Singular value decomposition entropy (SVDE)

SVDE measures the dimensionality of the EEG data by analyzing the number of eigenvectors to represent the data [7]. It is defined by (10),

$$SVDE= =-\sum_{i=1}^{n} \sigma_i \log_2 \sigma_i \quad (10)$$

where 'σ_i' denotes the values of the embedding space matrix of the delayed vector (also known as singular spectrum) of the input EEG data and 'n' denotes the number of singular spectrums.

e. Approximate entropy and sample entropy

Approximate entropy (ApEn) measures the degree of irregularity present in the data [30]. According to Steve Pincus, ApEn is defined as the "likelihood that runs of patterns that are close remain close on next incremental comparisons" [31]. The study demonstrates that ApEn performs well with relatively small time-series data. Sample entropy is used to examine the sequence and regularity present in the data and assigns a non- negative number to the sequence in such a way that the larger value denotes more irregularity present in the data [30]. Sample entropy can be defined as (11),

$$SampEn(m,r) = \lim_{N\to\infty} \left\{- \ln \left[\frac{A^m(r)}{B^m(r)}\right]\right\} \quad (11)$$

where 'm' denotes the run length of data points, 'r' denotes tolerance window, $A^m(r)$ denotes the probability of two m+1 matched sequences and $B^m(r)$ denotes the probability of two m matched sequences.

3.2.4. Fractional dimension features

FD features are another non-linear analysis technique used for analyzing EEG signals. It is used to measure the FD of a geometric object [32].

a. Katz's FD

Katz's FD algorithms are calculated by derivating FD directly from the planar waveform [7], [32]. The Katz's FD is calculated as (12) [33],

$$FD= \frac{\log(N)}{\log(N)+\log\left(\frac{d}{L}\right)} \tag{12}$$

where 'd' denotes the diameter of the waveform and 'L' is the length of the waveform.

b. Petrosian FD

Petrosian FD is computed by applying Katz's FD over the binary sequences of the time-series data [7], [32]. The Petrosian FD is calculated as (13) [33],

$$FD= \frac{\log(N)}{\log(N)+\log\left(\frac{N}{N+0.4N_{\Delta}}\right)} \tag{13}$$

where 'N_{Δ}' is the number of unique segment pairs present in the binary sequence.

c. Higuchi's FD

Consider X(1), X(2), …, X(N) be the time-series data points and is constructed as (14),

$$X_k^m{:}X(m),X(m+k), \dots, X\left(m+\left[\frac{N-m}{k}\right].k\right) \tag{14}$$

where m=1,2, …, k and 'm' denotes the starting point. 'k' denotes intervals between data points. For each 'k', calculate the length of the curve by (15).

$$L_m(k)=\frac{1}{k}\left[\frac{\left(\sum_{i=1}^{\left[\frac{N-m}{k}\right]}|X(m+ik)-X(m+(i-1)k)|(N-1)\right)}{\left[\frac{N-m}{k}\right].k}\right] \tag{15}$$

where '$L_m(k)$' denotes the length of the curve. Then Higuchi's FD is calculated by applying (16).

$$FD= -\lim_{k\to\infty} \frac{\log\langle L(k)\rangle}{\log k} \tag{16}$$

4. RESULTS AND ANALYSIS

For finding out the better performance of EEG signals in classifying cognitive stimuli, we investigated which features or combination of features along with respective classifiers. Also, we suggested two methods and compared the feature extraction techniques based on that. Based on the comparison results, we provided our evaluation of the best set of features that could be used in the classification of cognitive stimuli-based EEG signals. In this study, the extracted features are given as input to the classifiers: RFC, KNN, MLP, LDA, and SVM and the accuracies are noted. Time and frequency domain features are then analyzed in two manners.

a. Analysis of individual domain features: Combine each feature with respect to their categories and compare the accuracies of each category
b. Analysis of hybrid domain features: Combine each category of features in multiple combinations and compare the accuracies of each combination of feature categories.

4.1. Analysis of individual domain features

The individual features that are extracted from the time-series data are combined and categorized in such a way that each feature belongs to one of the four categories named statistical, power, entropy, and FD features. Then each category is given as input to the classifiers and the accuracies are noted. In Table 2, accuracies of each feature category with all the five classifiers are shown and in Figure 2, the bar plot illustrates the accuracies of individual domain feature categories with all the five classifiers. It is shown that the FD feature has the highest accuracy of 82.53% with the LDA classifier.

Table 2. Performance analysis on individual domain features based on the accuracy

Features	RFC	LDA	SVM	MLP	KNN
Statistical features	63.5%	41.3%	69.8%	38.1%	50.8%
Power features	61.9%	17.5%	65.1%	38.1%	66.7%
Entropy features	79.4%	54.0 %	74.6%	79.4%	77.8%
FD features	76.2%	82.5%	66.7%	47.6%	74.6%

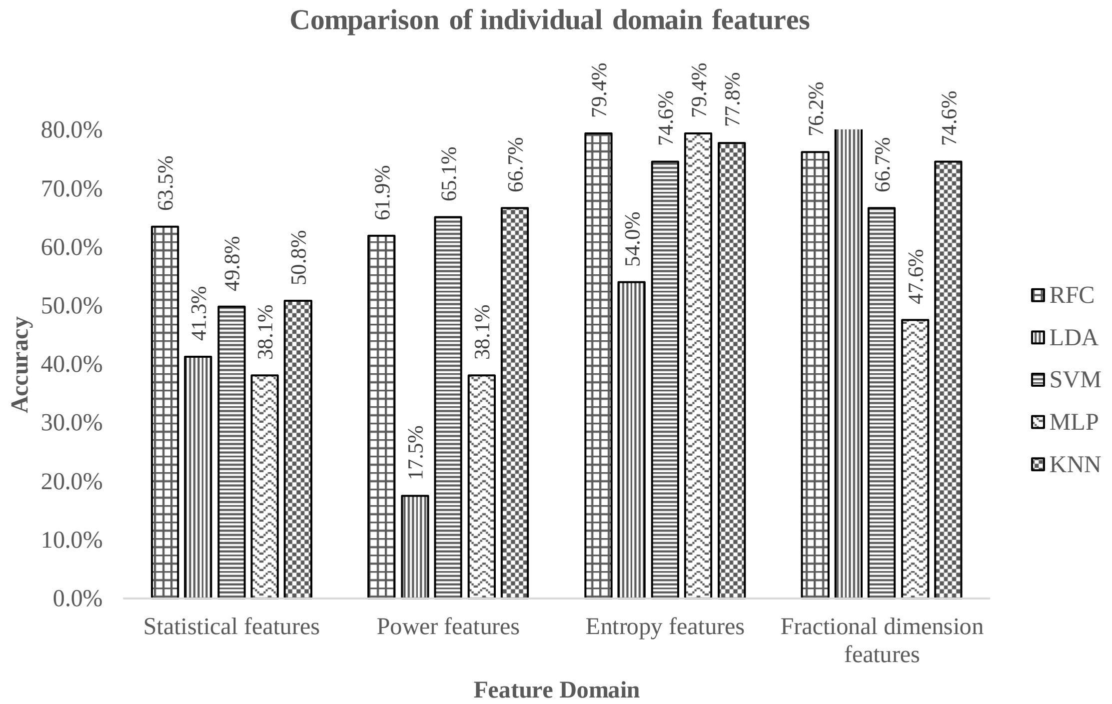

Figure 2. Performance analysis on individual domain features based on the accuracy

4.2. Analysis of hybrid domain features

The different combinations of four categories of features are made which consists of a total of 11 unique combinations of categories where 'S' represents statistical feature, 'P' represents power feature, 'E' represents entropy feature, 'F' represents FD feature. These different categorical combinations of features are given as input to the classifiers and the accuracies are noted. In Table 3, accuracies of every combination of the hybrid feature with all the five classifiers are shown and in Figure 3, the bar plot illustrates the accuracies of hybrid domain features with all the five classifiers. It is shown that the combination of entropy-FD features gives the highest accuracy of 90.47% with the KNN classifier.

Table 3. Performance analysis on hybrid domain features based on the accuracy

Features	RFC	LDA	SVM	MLP	KNN
S-P	68.2%	46.0%	61.9%	46.0%	50.8%
S-F	87.3%	28.6%	69.8%	52.4%	63.5%
S-E	84.1%	19.0%	69.8%	52.4%	74.6%
P-F	82.5%	87.3%	69.8%	46.0%	73.0%
P-E	79.4%	34.9%	69.8%	50.8%	77.8%
E-F	74.6%	69.8%	73.0%	81.0%	90.5%
S-P-F	82.5%	30.2%	61.9%	46.0%	65.1%
S-E-F	81.0%	27.0%	69.8%	47.6%	82.5%
S-E-P	82.5%	19.0%	61.9%	50.8%	76.2%
E-P-F	73.0%	47.6%	73.0%	61.9%	90.5%
S-E-P-F	81.0%	31.8%	63.5%	50.8%	82.5%

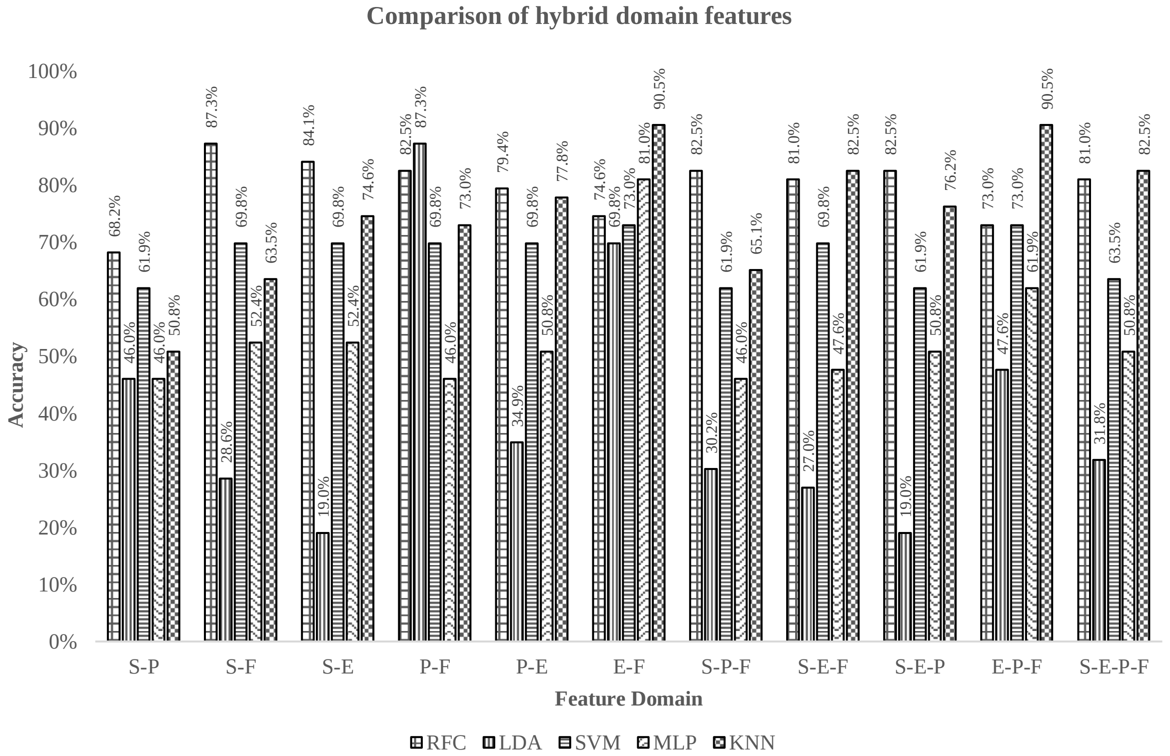

Figure 3. Performance analysis on hybrid domain features based on the accuracy

5. DISCUSSIONS

From the results obtained in the current study, we demonstrated that hybrid domain feature analysis gives the highest accuracy of 90.47% when using hybrid features of entropy-FD categories and hybrid features of entropy-power-FD categories with KNN outperforming individual domain feature analysis which achieves 79.36% with RFC, MLP and 82.53% with LDA, respectively. Dutta *et al.* [26] proposed feature extraction techniques for mental task-based EEG signals classification by combining multivariate empirical mode decomposition (MEMD) and phase-based decomposition. The features hence extracted are in the time domain which made this mode easy to implement in real-time applications. The LS-SVM classifier is used to classify the extracted features which achieve the highest accuracy of 83.33%. Our proposed approach outperformed this method with the highest accuracy of 90.47%. The hybrid features of entropy-FD categories are sufficient for efficient performance in classification since combining power feature with entropy-FD features has no impact on the classification accuracy and it provides the same accuracy as entropy-FD combination. Hence, we can conclude that the entropy-FD feature along with the KNN classifier can effectively be used in the classification of cognitive stimuli-based EEG signals.

6. CONCLUSION

Brain-computer interface is a way of communicating between the brain and an external device both sharing the same interface that can be controlled externally. The main agenda of the project is to enhance classification accuracy using hybrid features. In BCI, it is suggested that using features from multiple domains improves classification accuracy. There are various methods for extracting features from the raw EEG signals such as statistical approaches, power features, wavelet features, etc. In this study, statistical features such as mean, standard deviation, MAV, RMS, COV, and power features such as PSD, FD approaches, and entropy features are extracted from the raw EEG signals. Two methods of combinations are employed to find the best combination of features along with its classifier. From the experiment, it is shown that FD features with LDA classifier give the better accuracy of 82.53% when compared to other features, and the combination of entropy-FD features with KNN classifier gives the highest accuracy of 90.47% when compared to other combinations. From the above-obtained results, it is suggested that the combination of entropy and FD features with the KNN classifier can be used to effectively classify the target class. Hence the above combination with the respective model can be used for predicting the stimuli class of the tasks performed by the subjects from their brain signals very effectively. The above combination of hybrid feature sets might increase the computational complexity of the system as the data provided increases. Our work can be further extended by applying feature optimization and channel selection techniques to minimize the complexity of the existing model.

ACKNOWLEDGEMENTS

This research is financially supported by the Interdisciplinary Cyber Physical Systems (ICPS) Division of Department of Science and Technology (DST-ICPS).

REFERENCES

[1] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan, “Brain-computer interface research at the wadsworth center,” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 222–226, Jun. 2000, doi: 10.1109/86.847823.
[2] U. Gain, “The cognitive function and the framework of the functional hierarchy,” *Appl. Comput. Informatics*, vol. 16, no. 1/2, pp. 81–116, Mar. 2018, doi: 10.1016/j.aci.2018.03.003.
[3] J. R. Wolpaw *et al.*, “Brain-computer interface technology: a review of the first international meeting,” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–173, Jun. 2000, doi: 10.1109/TRE.2000.847807.
[4] D. H. R. Blackwood and W. J. Muir, “Cognitive brain potentials and their application,” *Br. J. Psychiatry*, vol. 157, no. S9, pp. 96–101, Dec. 1990, doi: 10.1192/S0007125000291897.
[5] L. Shen, X. Dong, and Y. Li, “Analysis and classification of hybrid EEG features based on the depth DRDS videos,” *J. Neurosci. Methods*, vol. 338, May 2020, doi: 10.1016/j.jneumeth.2020.108690.
[6] S.-I. Choi, G.-Y. Choi, H.-T. Lee, H.-J. Hwang, and J. Shin, “Classification of mental arithmetic and resting-state based on Ear-EEG,” in *2018 6th International Conference on Brain-Computer Interface (BCI)*, Jan. 2018, pp. 1–4. doi: 10.1109/IWW-BCI.2018.8311525.
[7] R. Nawaz, K. H. Cheah, H. Nisar, and V. V. Yap, “Comparison of different feature extraction methods for EEG-based emotion recognition,” *Biocybern. Biomed. Eng.*, vol. 40, no. 3, pp. 910–926, Jul. 2020, doi: 10.1016/j.bbe.2020.04.005.
[8] D. Garg and G. K. Verma, “Emotion recognition in valence-arousal space from multi-channel EEG data and wavelet based deep learning framework,” *Procedia Comput. Sci.*, vol. 171, no. 3, pp. 857–867, Jul. 2020, doi: 10.1016/j.procs.2020.04.093.
[9] R. Chatterjee, T. Bandyopadhyay, D. K. Sanyal, and D. Guha, “Comparative analysis of feature extraction techniques in motor imagery EEG signal classification,” in *Procedia Computer Science*, vol. 171, 2018, pp. 73–83. doi: 10.1007/978-981-10-5828-8_8.
[10] A. Suleiman and A.-H. Fatehi, “Features extraction techniques of EEG signal for BCI applications,” University of Mosul, Iraq, 2007.
[11] F. Wang *et al.*, “Emotion recognition with convolutional neural network and EEG-based EFDMs,” *Neuropsychologia*, vol. 146, Sep. 2020, doi: 10.1016/j.neuropsychologia.2020.107506.
[12] C. Wei, L. Chen, Z. Song, X. Lou, and D. Li, “EEG-based emotion recognition using simple recurrent units network and ensemble learning,” *Biomed. Signal Process. Control*, vol. 58, Apr. 2020, doi: 10.1016/j.bspc.2019.101756.
[13] V. Harpale and V. Bairagi, “An adaptive method for feature selection and extraction for classification of epileptic EEG signal in significant states,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 33, no. 6, pp. 668–676, Jul. 2021, doi: 10.1016/j.jksuci.2018.04.014.
[14] A. Akrami, S. Solhjoo, A. Motie-Nasrabadi, and M.-R. Hashemi-Golpayegani, “EEG-based mental task classification: Linear and nonlinear classification of movement imagery,” in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005, pp. 4626–4629. doi: 10.1109/IEMBS.2005.1615501.
[15] L. Shen, Z. Liu, and Y. Li, “EEG based dynamic RDS recognition with frequency domain selection and bispectrum feature optimization,” *J. Neurosci. Methods*, vol. 337, May 2020, doi: 10.1016/j.jneumeth.2020.108650.
[16] P. V RamaRaju, N. AnogjnaAurora, and V. M. Rao, “Relevance of wavelet transform for taxonomy of EEG signals,” in *2011 3rd International Conference on Electronics Computer Technology*, Apr. 2011, pp. 466–470. doi: 10.1109/ICECTECH.2011.5941838.
[17] A. S. Al-Fahoum and A. A. Al-Fraihat, “Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains,” *ISRN Neurosci.*, vol. 2014, pp. 1–7, Feb. 2014, doi: 10.1155/2014/730218.
[18] M. M. Rahman and S. A. Fattah, “An efficient feature extraction scheme for classification of mental tasks based on inter-channel correlation in wavelet domain utilizing EEG signal,” *Biomed. Signal Process. Control*, vol. 61, Aug. 2020, doi: 10.1016/j.bspc.2020.102033.
[19] M. P.P., T. Thomas, and R. Gopikakumari, “Wavelet feature selection of audio and imagined/vocalized EEG signals for ANN based multimodal ASR system,” *Biomed. Signal Process. Control*, vol. 63, Jan. 2021, doi: 10.1016/j.bspc.2020.102218.
[20] R. Chatterjee and T. Bandyopadhyay, “EEG based motor imagery classification using SVM and MLP,” in *2016 2nd International Conference on Computational Intelligence and Networks (CINE)*, Jan. 2016, pp. 84–89. doi: 10.1109/CINE.2016.22.
[21] M. Murugappan, N. Ramachandran, and Y. Sazali, “Classification of human emotion from EEG using discrete wavelet transform,” *J. Biomed. Sci. Eng.*, vol. 03, no. 04, pp. 390–396, 2010, doi: 10.4236/jbise.2010.34054.
[22] M. M, M. Rizon, R. Nagarajan, M. Rizon, and R. Nagarajan, “EEG feature extraction for classifying emotions using FCM and FKM,” *nternational J. Comput. Commun. Control*, vol. 2, no. 2, pp. 299–304, 2008.
[23] V. Bajaj, S. Taran, S. K. Khare, and A. Sengur, “Feature extraction method for classification of alertness and drowsiness states EEG signals,” *Appl. Acoust.*, vol. 163, p. 107224, Jun. 2020, doi: 10.1016/j.apacoust.2020.107224.
[24] M. K. Bandil and A. K. Wadhwani, “Multi-resolution EEG AND EEG sub-band features optimization for epileptic classification using hybrid evolutionary computing technique,” *Procedia Comput. Sci.*, vol. 152, pp. 243–251, 2019, doi: 10.1016/j.procs.2019.05.020.
[25] J. Liu, H. Meng, A. Nandi, and M. Li, “Emotion detection from EEG recordings,” in *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, Aug. 2016, vol. 152, pp. 1722–1727. doi: 10.1109/FSKD.2016.7603437.

[26] S. Dutta, M. Singh, and A. Kumar, “Classification of non-motor cognitive task in EEG based brain-computer interface using phase space features in multivariate empirical mode decomposition domain,” *Biomed. Signal Process. Control*, vol. 39, pp. 378–389, Jan. 2018, doi: 10.1016/j.bspc.2017.08.004.

[27] Y. Wu, Y. Zhou, G. Saveriades, S. Agaian, J. P. Noonan, and P. Natarajan, “Local shannon entropy measure with statistical tests for image randomness,” *Inf. Sci. (Ny).*, vol. 222, pp. 323–342, Feb. 2013, doi: 10.1016/j.ins.2012.07.049.

[28] H. Misra, S. Ikbal, H. Bourlard, and H. Hermansky, “Spectral entropy based feature for robust ASR,” in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. I--193--6. doi: 10.1109/ICASSP.2004.1325955.

[29] M. Riedl, A. Müller, and N. Wessel, “Practical considerations of permutation entropy,” *Eur. Phys. J. Spec. Top.*, vol. 222, no. 2, pp. 249–262, Jun. 2013, doi: 10.1140/epjst/e2013-01862-7.

[30] Y. Kumar and M. L. Dewal, “Complexity measures for normal and epileptic EEG signals using ApEn, sampen and SEN,” *Int. J. Comput. Commun. Technol.*, vol. 4, no. 2, pp. 95–101, Apr. 2013, doi: 10.47893/IJCCT.2013.1175.

[31] S. Pincus, “Approximate entropy (ApEn) as a complexity measure,” *Chaos An Interdiscip. J. Nonlinear Sci.*, vol. 5, no. 1, pp. 110–117, Mar. 1995, doi: 10.1063/1.166092.

[32] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt, “A comparison of waveform fractal dimension algorithms,” *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.*, vol. 48, no. 2, pp. 177–183, 2001, doi: 10.1109/81.904882.

[33] C.-T. Shi, “Signal pattern recognition based on fractal features and machine learning,” *Appl. Sci.*, vol. 8, no. 8, p. 1327, Aug. 2018, doi: 10.3390/app8081327.

BIOGRAPHIES OF AUTHORS

Jessy Jane Christy J. completed her B.Tech. in Computer Science and Engineering in 2019 from Karunya Institute of Technology and Sciences. She is currently pursuing M.Tech. in Computer Science and Engineering at Karunya Institute of Technology and Sciences. Her research interests include machine learning, data science, and image processing. Email: jessyjane@karunya.edu.in.

Dr. E. Grace Mary Kanaga is currently working as an associate professor in the Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences. Her area of research includes computational intelligence, big data analytics and brain computer interface. She has published more than 45 papers in reputable national and international conferences, international journals and book chapters. Currently she is working on a research project funded by DST- ICPS to the worth of Rs. 40 lakhs. Three candidates have successfully completed their Ph.D. under her guidance, and she is currently guiding 6 research scholars. Email: grace@karunya.edu.