CrossMark

# An efficient Concealed Information Test: EEG feature extraction and ensemble classification for lie identification

Annushree Bablani[1] · Damodar Reddy Edla[1] · Diwakar Tripathi[2] · Venkatanareshbabu Kuppili[1]

## Abstract

EEG-based lie detectors have become popular over polygraphs because it cannot be controlled by human intentions. Various studies have performed "Guilty Knowledge Test" or "Concealed Information Test" by creating a mock crime scenario to identify changes in brain potential. In this study, an individual's behavior during lying is analyzed and a new scenario is developed for "Concealed Information Test." This work involves a mock crime scenario using an EEG acquisition device for 10 participants. Data acquisition has been performed by placing 16 electrodes on the subjects' scalp. For this experiment, the subject has to recognize faces of some known and unknown personalities among 10 images flashed. These images behave as stimulus for the subject which generate corresponding brain responses. Various feature extraction approaches such as statistical, time domain, frequency domain and time–frequency domain are applied to the 16- channel EEG data. For classifying a subject as guilty or innocent, five classifiers have been applied on subject-wise EEG data. Moreover, the classifiers' ranking is considered based on the performance of classifiers. An ensemble framework is developed by aggregating the results of the best three classifiers out of the tested five classifiers. The classifiers' results are aggregated using a weighted voting approach and have been compared with popular conventional approaches using various classification performance measures. Results present a comparative performance of different feature extraction approaches and classifiers using subject-wise single-trial EEG data. The wavelet approach performs better for EEG data of most of the subjects. A comparison between base classifiers and ensemble framework is provided with the ensemble approach outperforming over the base classifiers. Further proposed framework is compared with some existing approaches, and a highest accuracy of 92.4% has been achieved.

**Keywords** Electroencephalography · Event-related potential · Lie detection · Feature extraction · Ensemble classification

## 1 Introduction

Over the years, significant research has been conducted aiming at improving the interface between humans and machines. An interface aiming at developing coordination and correlation between man and his brain has been evolved by the name brain–computer interface (BCI). When our brain responds to an event or activity, a difference in potential is generated. This difference in potential is nothing but transmission of a message from one synapse to another through the release of a chemical. Movement of chemical inside brain cells generates brain potential or brain signal, which is then recorded via acquisition devices. There are several different techniques to record the brain signals or potentials.

Few commonly used noninvasive acquisition techniques include electroencephalograph (EEG), magnetoencephalography and functional magnetic resonance imaging. Invasive techniques such as electrocorticography (ECoG) place electrodes beneath the scalp, whereas noninvasive techniques as discussed above place electrodes on the scalp and are safe and easy to use [1]. Various types of brain potentials are generated depending on the type of events/stimuli/commands given to the brain. One of them is event-related potential (ERP) which is a psychological response, originated due to the reflex

✉ Annushree Bablani
  annubablani@nitgoa.ac.in

  Damodar Reddy Edla
  dr.reddy@nitgoa.ac.in

  Diwakar Tripathi
  diwakarnitgoa@gmail.com

  Venkatanareshbabu Kuppili
  venkatanaresh@nitgoa.ac.in

1  National Institute of Technology, Goa, Ponda, India

2  Madanapalle Institute of Technology and Science, Madanapalle, A.P., India

 Springer

generated in the brain on performing some mental activity or observing something familiar. Using the advantage of ERP, lie detectors have been developed which, by analyzing variation in brain signals, identifies whether a subject is guilty or not. In the past, polygraph-based lie detectors were available which analyzes human behavior like sweating, increased heart rate. But this indirect view was not successful in classifying a subject as guilty or innocent. ERP-based lie detectors provide a direct view to understand and detect deception.

An ERP component called P300 is usually identified by researchers for lie detection [2,3]. P300 is a response which is generated between 300 and 1500 ms on the occurrence of some meaningful stimulus or "oddball stimulus" [2]. "Guilty Knowledge Test (GKT)" or "Concealed Information Test (CIT)" is conducted to detect whether the given subject is guilty or innocent. To perform CIT, subjects are trained for a mock crime scene where three different stimuli are randomly given to the subject, i.e., target, irrelevant and probe. Target stimulus is recognized by guilty and innocent; the probe is crime-related stimulus which is identified by guilty only, and irrelevant stimulus is not related to crime and is unidentified. Out of the total stimuli given to a subject, 70% is irrelevant and rest includes probe and target [2]. Probe and target are rarely occurring meaningful stimuli; hence, they will generate P300 wave and the irrelevant stimulus will not generate any P300 component.

For the guilty subject, probe and target stimuli will generate a P300 response, whereas, for the innocent subject, only target stimulus will generate a P300 response. Many methods have been applied to analyze probe, irrelevant and target responses and identify whether responses are generated from guilty or innocent subject. We have categorized these methods into two approaches— statistical approaches and machine learning approaches.

## 1.1 Statistical approach

### 1.1.1 Bootstrapping techniques

Bootstrapping is a statistical technique to measure the similarity between the brain responses generated by stimuli presented to any subject. It is applied by randomly sampling the parameters with various iteration. A "double centered correlation" method [3] has been applied by Farwell et al., where they found a correlation between probe and irrelevant and between probe and target response. A conclusion has been made that if the correlation between probe and target stimuli is greater than the correlation between probe and irrelevant stimuli, then the subject is guilty and vice versa. A different mock crime scenario has been developed where instead of two groups, three group mock crime scenario have been conducted. One group was guilty, the other was innocent, and the third group was a countermeasure group, which

had the freedom to perform any covert responses for irrelevant stimuli. To identify concealed information, "bootstrap amplitude method" has been applied by authors. Average of ERP amplitude is calculated from probe set (randomly with replacement) and from the irrelevant set (randomly with replacement) [4,5]. The average of ERP from probe set and the irrelevant set is subtracted, and the process is iterated for 100 times. With 95% confidence, if the difference of the average of probe set with irrelevant set is greater than zero, a guilty decision is made. Similar bootstrapping method, known as "bootstrap reaction time," was used by authors where instead of amplitude difference, reaction time was considered [4].

### 1.1.2 Analysis of variance

Analysis of variance (ANOVA) is used to analyze the difference between the means (i.e., variance) of various groups. The variance is partitioned into components attributable to different sources of variation; hence, it is useful for comparing the variance of several groups. Using its advantage, ANOVA has been applied to ERP responses [6]. A CIT using "cards" has been conducted by authors, where different suit cards were either probe or irrelevant. From a pack of cards, five selected cards and a joker card were presented as target stimulus.

Peak values of EEG waves have been considered as a measure of identification of deception in many research works [5–7]. Signal peak alone is not enough to explain various characteristics of ERP responses generated while testing deception. EEG data are also affected by the noise like ocular artifacts or muscular artifacts making it difficult to analyze the ERP response generated by stimuli. To overcome the issue of noise removal, machine learning techniques are applied which reduces noise from EEG data. Many authors have worked on removal of artifact and classification of EEG data, which is being discussed in the succeeding paragraph.

## 1.2 Machine learning approach

Machine learning approaches have become another tool for CIT as it analyzes the signal for each trial. A GKT [3] has been applied by Abootalebi et al. [2] in their work. Instead of analyzing only peak values, a set of relevant features from EEG data have been extracted. Morphological features like latency, amplitude, the ratio of latency and amplitude; frequency features like mean, mode, the median of frequency; and wavelet features are extracted from EEG waveform. But all the features do not give exact information about the data and do not provide better classification accuracy. Hence, genetic algorithm [2] has been applied for feature selection and further statistical classifier is applied to classify data as innocent or guilty. In another work, genetic algorithm for feature selection has been applied [8]. "Empirical

Mode Decomposition (EMD)" has been applied for feature extraction on P300 responses. EMD decomposes a signal into components called "Intrinsic Mode Functions (IMFs)." The IMFs are frequency or amplitude modulated waveforms which provide the physical characteristics of a wave. IMFs have been used as features and post-feature selection LDA has been applied as a classifier.

To resolve the issue of denoising EEG data, independent component analysis (ICA) has been used to remove noisy components [9]. ICA partitions data into various independent components (ICs) and spatial map sets or categorizes EEG data into task related and task unrelated. But there has been an issue in identifying the difference between task-related and the unrelated components. To overcome this issue, a template matching method has been devised [9]. The spatial map set which is similar to the template is selected from different guilty subjects for different stimuli and these are regarded as P300 ICs, by which the P300 and non-P300 responses with high SNR are reconstructed at sensors. A birthday paradigm for CIT has been applied considering time constraint with accuracy to predict the deception [10]. Nonparametric-based feature extraction with LDA and KNN as classifier has been used. A spare representation method has been used on irrelevant and target ERP responses for single subject [11]. In [12], genetic SVM as classifier has been applied to identify guilty subject using a novel CIT method. The author in [13] also proposed a novel association-based CIT having similarities with reaction time–CIT, which considers reaction time differences between irrelevant and probe stimuli.

A single classifier gives a good performance, but combining various classifiers in an ensemble framework provides near to optimal results. Hence, we have applied ensemble classification approach in our work to obtain better results. Ensemble classification methods like bagging, boosting, random forest have been applied by many authors on different datasets. Bagging [14] combines various machine learning approaches to design one model by decreasing its variance, whereas boosting [15] decreases the bias values. Boosting and bagging were applied by [16] on 23 datasets using decision tree and neural networks. They showed that ensemble methods provide consistent results with single classifiers. Another ensemble approaches random forest (RF) [17], constructed multiple decision trees at training time, and gives a class label as output which takes mean or mode of multiple decision trees.

An ensemble framework has been proposed, where five classifiers, namely LDA, SVM, MLFFNN, KNN and naïve Bayes, have been applied to EEG data. Later, classification ranking has been applied and results of three best classifiers have been aggregated using weighted voting (WV) approach. In this study, we have recorded EEG signals by conducting a CIT and detailed explanation of it is given in Sect.

2. Brain signals are affected by a lot of noise; therefore, noise removal followed by feature extraction and proposed classification framework has been discussed in Sect. 3. The result of the proposed framework is given in Sect. 4. Later sections provide conclusion and future work followed by references.

## 2 Data acquisition

### 2.1 Subjects

An EEG-based CIT experiment has been conducted where 10 participants bearing age between 23 and 35 participated. No medical record of any psychological disorder was found with subjects, and they were having normal or corrected vision. Subjects are given a brief description of the complete experimental procedure. Before beginning the experimental procedure, subjects have given in written for their consent to this experiment. The EEG data recording is done by placing Ag/AgCl electrodes at Fz, FC1, FC2, C3, Cz, C4, CP5, CP1, CP2, CP6, P3, Pz, P4, O1, Oz and O2 sites (10–20 international system).

To record vertical electro-occulograph (VEOG) and horizontal electro-occulograph (HEOG), a signal was recorded from right eye, for VEOG above and below the eye and for HEOG from outer canthus. An electrode placed on mastoid for reference and an electrode on the forehead as ground. For signal acquisition, we have used EasyCap [18] (a 32-Channel EEG Standard Cap Set (Munich, Germany)), a V-amp amplifier, set of 16 electrodes and brain vision recorder [19]. Electrode placement protocol is similar as in [9].

Before execution of experiment, subjects are randomly grouped under to the "guilty" and "innocent" groups for two experimental sessions. All 10 participants behave as innocent for one session consisting of 30 trials and guilty for second session consisting of 30 trials. For a total of 60 trials, subjects are presented with a set of images of some known and unknown personalities for 31 seconds. Three types of stimulus are presented to the subjects:

- Probe stimulus: It is a crime-related stimulus which is presented rarely and generates a P300 response for guilty subjects. Here, probe is an image of a known person (images of the person from institute)
- Target stimulus: This stimulus is familiar to all subjects, as it is given to maintain the concentration of the subject in the experiment. Target will generate a P300 response for both guilty and innocent subjects. For the experiment, images of well-known personalities or celebrities have been used as target
- Irrelevant stimulus: This stimulus is not related to crime and does not generate any ERP response that is guilty
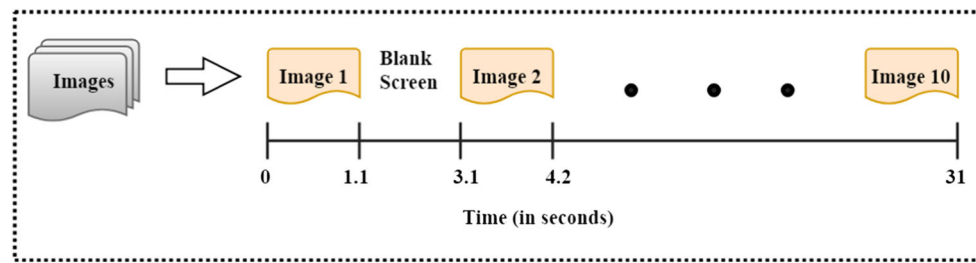
**Fig. 1** Trial structure

or innocent. These are random images downloaded from internet sources. For the innocent subject, responses generated from probe and irrelevant stimuli will be same.

In the initial phase of the experiment, subjects are trained for a mock crime scenario. They are trained as they have committed a crime with the person whom they know (i.e., a well-known colleague from college). There are 10 images to be presented to subjects where one image is a probe, two of them are target and rest are irrelevant. These images are presented to subjects on a 15.4-inch display screen. The images of known persons will act as a probe, and this will generate P300 response in the brain. Also, target images (celebrity images) will generate P300 response, whereas the images of random unknowns, i.e., irrelevant stimuli, will generate a non-P300 response. These generated responses are hence recorded and analyzed using brain vision analyzer.

Images were presented for 31 seconds; each stimuli image remained on screen for 1.1 second with 2 second interstimulus period (Fig. 1) [2]. On observing the stimulus, subject is instructed to respond "yes" or "no." For probe stimuli, subjects who replied "no" indicates that they are lying. The probe is presented rarely, and these images are related to crime and hence will generate P300. For irrelevant stimuli, also subject has to reply "no" which represents that they are speaking the truth. As irrelevant stimuli are mostly occurring and are not related to crime, hence it will not generate P300 responses. For target stimuli, subjects are instructed to reply "yes" as they identify the target stimuli image. As the target is also rarely occurring and is not related to crime, but familiar to the subject, hence P300 responses will generate for the same.

Subjects go through a practice session of 5 min in which they perform few trials of the task identical to the full experiment as described above. After the training session, experiments are conducted of 30 trials as innocent and 30 trials as guilty. The complete experimental procedure developed in our work is depicted in Fig. 2. The experiment has been conducted for two sessions for each subject.

## 2.2 Signal preprocessing

Before proceeding for analysis of CIT data, EEG signal is preprocessed for artifact removal. VEOG- and HEOG-based ocular artifacts are corrected using brain vision analyzer. Further, to remove other noise mixed with the signal, we have applied bandpass filter on our data. The bandpass filter is applied from a range of 0.3–30 Hz which will eliminate high-frequency bands from the raw signal. This frequency range is mostly observed during mental tasks [4]; hence, bandpass filter is applied for this range.

Signals acquired using EEG device are hence converted into (.)mat form using Brain Vision Analyzer 2.1 [19]. Further, the feature extraction and classification are done using MATLAB R2015a on Intel i7 processor with 8 GB RAM and 64-bit Windows 10 Pro platform.

## 3 Proposed approach

This section describes about the proposed ensemble framework and discusses about the various feature extraction approaches applied.

### 3.1 Feature extraction

The EEG data have been recorded from 16 channels of 10 subjects (S-1 to S-10), but data of subject 6 (S-6) has not been considered for whole study process, due to the presence of lots of artifacts. Hence, we analyzed 16-channel data of 9 subject ($9 \times 16$) for 30 trials from two sessions (truth session and lie session). Various statistical and machine learning approaches for feature extraction have been applied. Statistical approaches alone are not sufficient to provide all the information of the signal; hence, we have combined it with various machine learning approaches and have tried to extract features from every domain of the signal. Following set of feature extraction methods have been used for the study:
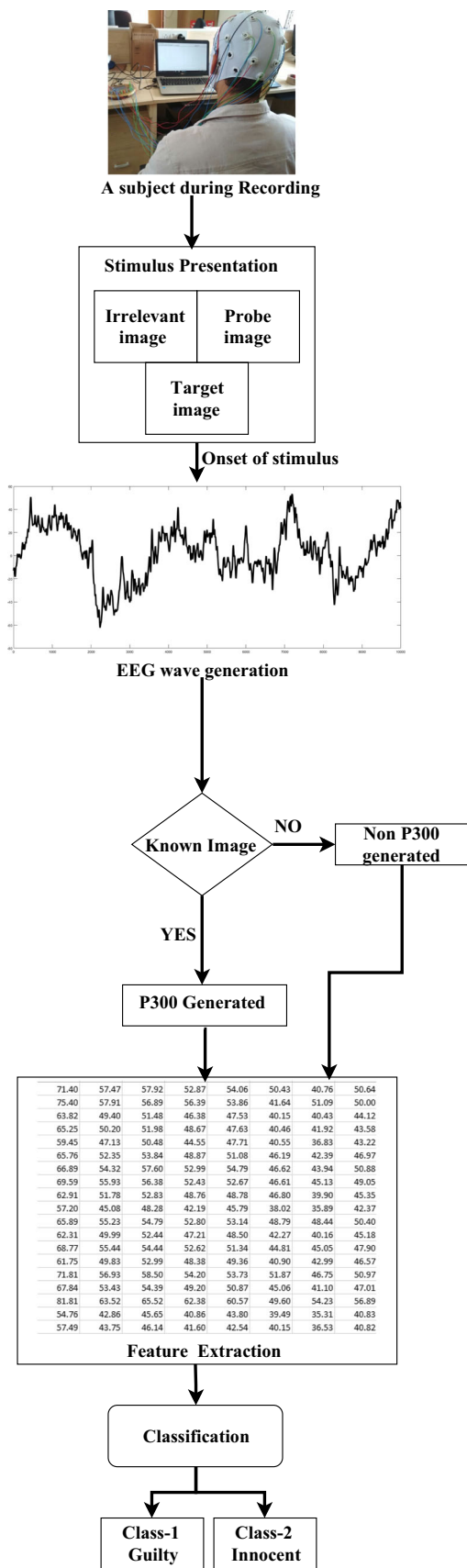
**Fig. 2** Experimental procedure

### 3.1.1 Potential or amplitude

Potential gives the maximum peak values at any time instant $t$ [2,6]. Let $x(t)$ represents the EEG signal at any time instance $t$. Potential ($P(x(t))$) is given as:

$$P(x(t)) = maxpeak(x(t)) \tag{1}$$

### 3.1.2 Power

Power gives the energy of the signal which is given by the square of potential value ($E(x(t))$)

$$E(x(t)) = (P(x(t)))^2 \tag{2}$$

### 3.1.3 Frequency response

Frequency response $\mu(x(t))$ [20] calculated using fast Fourier transform (FFT) gives complex values. Hence, magnitude of frequency is calculated as given in Eqs. 3–5.

$$X[N] = \Sigma_{m=0}^{M-1} x(m) W^{mN} \tag{3}$$

$$x(m) = \frac{1}{S} \Sigma_{N=0}^{M-1} X[N] W^{-mN} \tag{4}$$

where $N = 0, 1, \ldots, M - 1$ and $S$ is number of samples in $t$ time

$$\mu(x(t)) = abs(x(m)) \tag{5}$$

### 3.1.4 Hjorth features

Hjorth [21] developed three statistical parameters using time domain. We have used two of them, i.e., *mobility* and *complexity*, for extracting time features of EEG signals. The parameters are used for EEG feature extraction earlier for emotion recognition by [20]. *Mobility* gives root of ratio of "derivative of variance of signal" with "variance of signal," and *complexity* provides ratio of derivative of mobility to mobility of signal. Value of complexity lies in range of [0, 1], and 1 shows that signal is similar to sine wave

**Mobility**

$$M(x(t)) = \sqrt{\frac{\sigma(x'(t))}{\sigma(x(t))}} \tag{6}$$

where $x'(t)$ represents differential of $x(t)$ and $\sigma$ represents variance of the data.

### Complexity

$$C(x(t)) = \frac{M(x'(t))}{M(x(t))} \qquad (7)$$

#### 3.1.5 Wavelet transform

All the features explained above consider only time domain characteristics of ERP response. As ERP waveform has both time and frequency characteristics, nowadays authors are considering wavelet-like approaches [2]. Discrete wavelet transform (DWT) decomposes a signal into a variable frequency range which is represented in approximation or detail levels. For first-level decomposition, "approximation coefficients" consider high-frequency components, while "detail coefficients" consider low-frequency components of the signal. For next level, detail values are again decomposed into approximation and detail coefficients, and this process is continued for the number of levels assigned. DWT conserves time information while considering frequency components of signals. DWT has been applied to 4 levels using "db2" wavelet. After applying wavelet transform for 4 levels, the output vector consists of a set of approximation coefficients and four sets of detail coefficients which are used as the feature vector. For each trial, we received a large number of approximation and detail coefficient in the range of 0.5 to 30 Hz. Hence, root-mean-square value [20] of wavelet coefficients has been calculated and used as a feature.

$$Wv = \sqrt{\Sigma_{i=1}^{N} \frac{C_i^2}{N_i}} \qquad (8)$$

where $C_i$ are wavelet coefficients and $N$ gives a total number of coefficients generated from all levels.

### 3.2 Classification

For EEG-based CIT data, subjects are classified into two classes: *innocent* and *guilty*. A classifier can perform well on a particular dataset, while the same classifier may give unproductive results with some other dataset. Hence, one cannot predict the performance of a classifier for a dataset [22]. Many classifiers are available in the literature that outperform while dealing with binary classes. Previous CIT- based studies use LDA [8], SVM [9], KNN [10] as classifiers. In order to improve the performance of our system, we proposed an ensemble framework for lie detection, in which the decision of innocent or guilty is taken by aggregating the performance of different classifiers. Ensemble classification provides better performance as compared to single classifier performance [23]. Ensemble framework can be structured either by combining the same type of classifiers, i.e., homogeneous

framework, or by combining different types of classifiers, i.e., heterogeneous framework. Here, in our work we have applied heterogeneous ensemble framework, by combining 5 classifiers, namely LDA, SVM, multilayer feedforward neural network (MLFFNN), KNN and naïve Bayes (NB).

#### 3.2.1 Linear discriminant analysis (LDA)

LDA [24] is a typical linear classification technique, which provides separability by drawing decision region between data of two classes. To assure the maximum separability, it tries to optimize the ratio of within-class scatter to the between-class scatter. LDA searches for a linear solution to separate the data into classes.

#### 3.2.2 Support vector machines (SVM)

SVM [25] is a non-probabilistic linear classification approach which chooses an optimal separating hyperplane such that it maximizes the distance between data points of different classes. To construct the best solution for separating hyperplane, training data points are used which are considered as vectors or support vectors. These support vectors help in determining the width of the hyperplane. SVM can be extended for the cases where data are not separated by a hard margin. Hence, a trade-off parameter is used which allows margin to be flexible and separates nonlinearly separable data. In addition to that, SVM can be used to classify nonlinear data using kernel trick. This is the reason behind extensive use to SVM as it achieves to an optimal solution for both linear and nonlinear data.

#### 3.2.3 Multilayer feedforward neural network (MLFFNN or NN)

MLFFNN is successfully used in classification tasks, feature extraction, pattern mapping, etc. It takes input, processes it and produces an output. Various layers called hidden layers are added between input and the output layer to improve performance of the system. The nodes from input to output carry some weights. These weights are adjusted until the network reaches to an optimal solution. MLFFNN [26,27] consists of one input layer, one output layer and many hidden layers. At the output layer, an activation function is applied which processes the final output. The activation function can be a linear or a nonlinear function according to the specified problem.

#### 3.2.4 *k*-nearest neighbor

KNN [28] is a nonparametric classification approach that classifies test data into a particular class based on class of majority of its neighbors. To identify the class of given data

point, $k$ number of neighbors are chosen using different distance-based approaches like Euclidean distance, Mahalanobis distance, Manhattan distance. The $k$ is a constant, which can be selected based on the number of data points to be classified. There is no predefined technique to calculate the value of $K$, and it is selected heuristically. KNN is best suitable for low-dimensional data.

### 3.2.5 Naïve Bayes

NB [29] is a probabilistic classification approach which uses the concept *maximum likelihood* and is based on Bayesian theorem. It considers the conditional probability model to classify a data into various classes. It is generally suitable for high- dimensional data.

### 3.3 Ensemble framework

For any particular dataset, classifiers discussed above perform differently. Hence, we are unable to predict which classifier may have better classification performance. Ensemble is a strong approach to achieve the optimal solution for every dataset [30]. Above discussed, five classifiers are combined to form a heterogeneous ensemble framework (Fig. 3) for classification of CIT-based EEG data. LDA provides the best solution for linearly separable data, SVM gives optimal solution by using a trade-off parameter, NN provides nonlinear data classification by making use of activation functions and synaptic weights, naíve Bayes classifies data on basis of its prior probability, and KNN uses distance as a measure to classify data. Hence, by combining these different properties of five different classifiers, we have tried to obtain a near to optimal solution for our dataset. There are three approaches which aggregate results of base classifiers, as follows:

- Majority voting: It takes the decision in favor of a particular class if the majority (more than 50%) of classifiers classify it as that particular class.
- Unanimous voting: For a given class, say, *Class-Guilty*, if any classifiers' results classify data as *Class-Guilty*, then unanimous voting takes decision as *Class-Guilty*
- Weighted voting: In this approach, an aggregation function is applied which assigns higher weight values to the classifier with better performance. This will increase its participation in ensemble framework, in turn increasing the performance of overall system.

Majority voting and unanimous voting approach utilizes homogeneous ensemble framework providing better classification performance toward a particular classifier. It assigns equal weights to all classifiers; hence, the classifier which is performing better is given same weight as the least performing classifier. This approach will reduce the performance of the system if heterogeneous classifiers are used. However, weighted voting gives better classification results as it assigns more weights to the better performing classifier. Hence in this work, ensemble framework with heterogeneous classifiers is aggregated using weighted voting approach. The output of base classifiers is aggregated using aggregation function as in Eq. 9.

$$y_i = \Sigma_{i=1}^{m} w_i C_i \qquad (9)$$

where $m$ denotes number of classes and $w_i$ and $C_i$ denotes weights and output predicted by $i$th classifier.

For assigning the weight to classifier, initially equal weights are assigned to each base classifier, and on basis of their performance, weights are updated according to the Eq. 10.
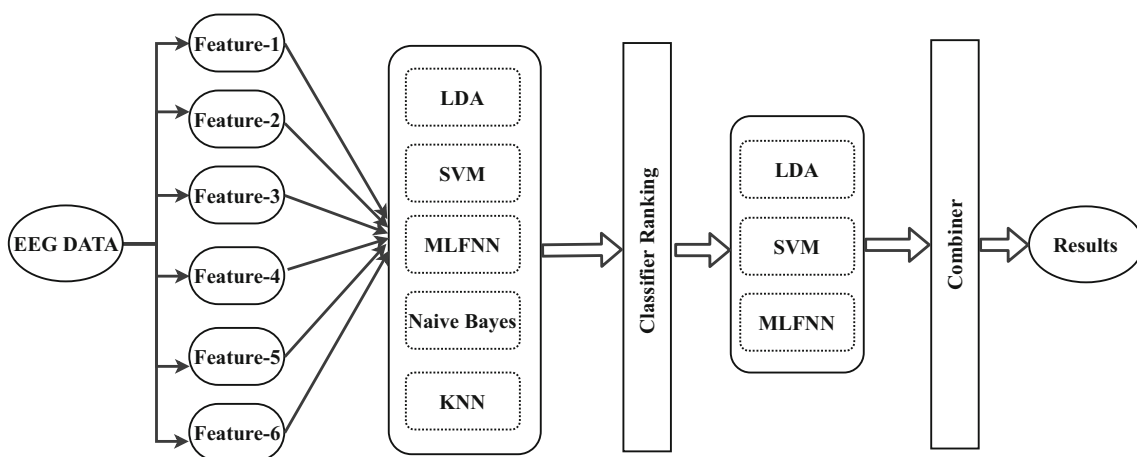


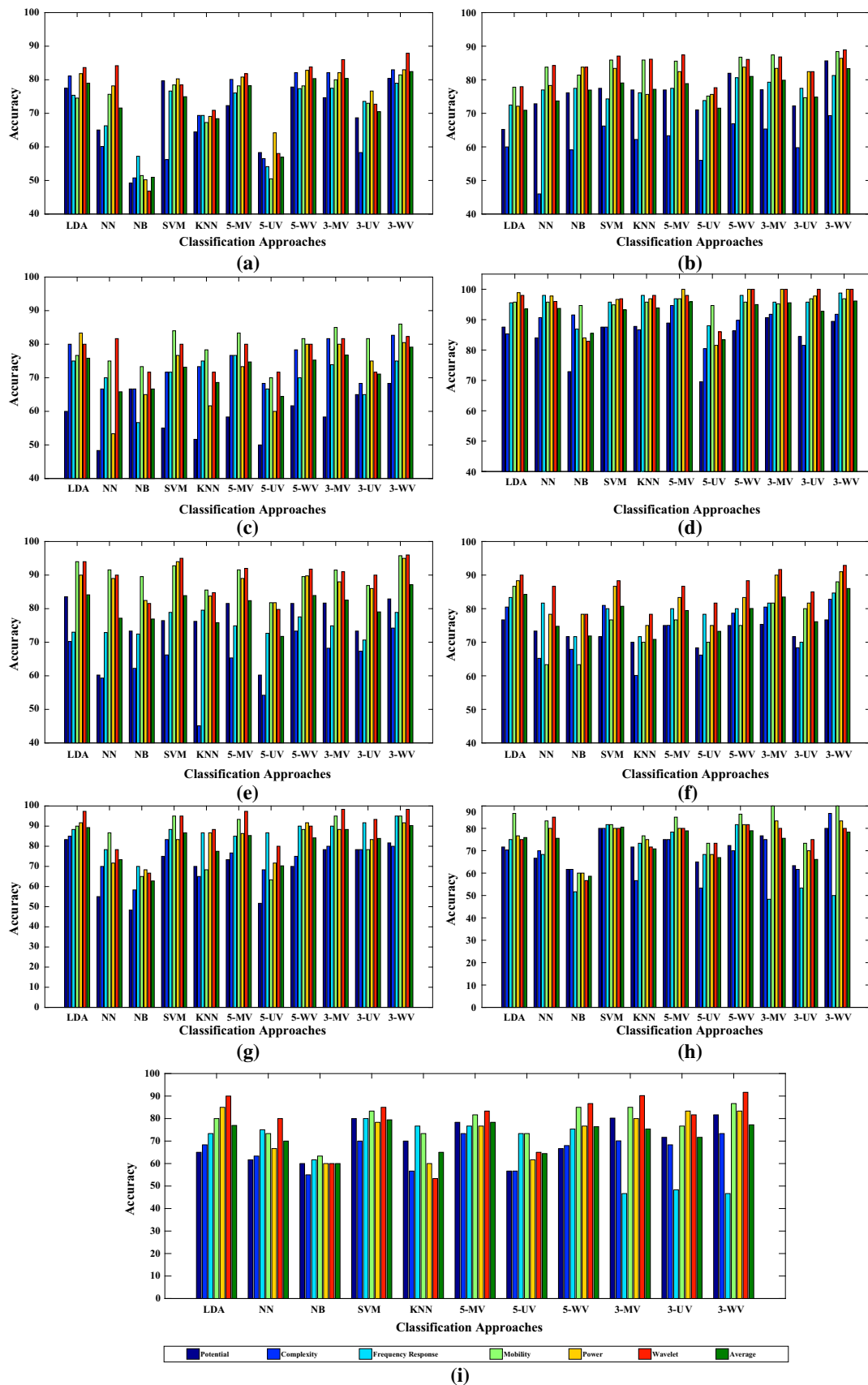**Fig. 3** Proposed ensemble framework for lie detection

Fig. 4 a–i Subject-wise results using various classifiers and different voting strategies for 3-classifier and 5-classifier system

**Table 1** Subject-wise performance using various classifiers and ensemble framework applied to potential values

| Measures | LDA | NN | NB | SVM | KNN | 5-MV | 5-UV | 5-WV | 3-MV | 3-UV | 3-WV (proposed method) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject 1 | | | | | | | | | | | |
| Accuracy | 77.5 | 65.0 | 49.2 | 79.7 | 64.5 | 72.3 | 58.3 | 77.8 | 74.7 | 68.7 | 80.4 |
| Sensitivity | 86.7 | 62.9 | 54.4 | 93.3 | 86.7 | 86.7 | 96.7 | 60.9 | 90.0 | 93.3 | 70.0 |
| Specificity | 67.4 | 63.5 | 46.6 | 70.3 | 37.1 | 55.4 | 16.0 | 186.7 | 59.4 | 40.6 | 81.1 |
| G-measure | 75.3 | 63.1 | 50.2 | 80.6 | 46.9 | 68.7 | 23.7 | 100.0 | 72.2 | 53.5 | 75.3 |
| Subject 2 | | | | | | | | | | | |
| Accuracy | 65.2 | 72.8 | 76.1 | 77.5 | 77.0 | 77.0 | 71.0 | 81.9 | 77.0 | 72.2 | 85.6 |
| Sensitivity | 67.1 | 77.6 | 78.1 | 73.3 | 89.8 | 75.6 | 96.0 | 93.8 | 65.3 | 92.0 | 73.3 |
| Specificity | 63.3 | 76.3 | 77.0 | 80.0 | 63.3 | 76.7 | 46.7 | 70.0 | 76.7 | 46.7 | 76.7 |
| G-measure | 63.9 | 76.3 | 77.1 | 74.1 | 74.2 | 75.0 | 66.6 | 80.3 | 68.1 | 64.8 | 73.0 |
| Subject 3 | | | | | | | | | | | |
| Accuracy | 60.0 | 48.3 | 66.7 | 55.0 | 51.7 | 58.3 | 50.0 | 61.7 | 58.3 | 65.0 | 68.3 |
| Sensitivity | 56.7 | 55.7 | 65.5 | 43.3 | 60.0 | 56.7 | 83.3 | 76.7 | 50.0 | 73.3 | 56.7 |
| Specificity | 63.3 | 58.2 | 71.6 | 66.7 | 43.3 | 60.0 | 33.3 | 43.3 | 70.0 | 23.3 | 60.0 |
| G-measure | 58.7 | 56.9 | 68.2 | 44.7 | 48.4 | 57.8 | 51.7 | 57.1 | 56.4 | 36.0 | 57.1 |
| Subject 4 | | | | | | | | | | | |
| Accuracy | 87.6 | 84.0 | 72.9 | 87.6 | 87.8 | 88.9 | 69.6 | 86.3 | 90.7 | 84.4 | 89.4 |
| Sensitivity | 81.3 | 84.0 | 68.6 | 79.6 | 89.8 | 85.8 | 96.0 | 92.0 | 83.6 | 89.8 | 83.6 |
| Specificity | 93.8 | 84.0 | 83.3 | 95.6 | 85.8 | 92.0 | 43.1 | 81.8 | 97.8 | 79.1 | 93.3 |
| G-measure | 87.1 | 84.0 | 75.4 | 86.5 | 87.7 | 88.7 | 63.7 | 86.6 | 90.0 | 83.9 | 87.8 |
| Subject 5 | | | | | | | | | | | |
| Accuracy | 83.6 | 60.2 | 73.3 | 76.4 | 76.2 | 81.6 | 60.2 | 81.6 | 81.7 | 73.3 | 82.9 |
| Sensitivity | 87.6 | 64.9 | 81.3 | 80.9 | 89.3 | 85.3 | 95.6 | 91.6 | 87.6 | 87.6 | 87.6 |
| Specificity | 79.6 | 49.3 | 74.1 | 72.0 | 63.1 | 77.8 | 24.9 | 59.6 | 77.8 | 59.1 | 75.6 |
| G-measure | 80.3 | 52.5 | 76.9 | 71.9 | 74.9 | 80.5 | 36.2 | 71.7 | 79.5 | 69.8 | 78.4 |
| Subject 7 | | | | | | | | | | | |
| Accuracy | 76.7 | 73.3 | 71.7 | 71.7 | 70.0 | 75.0 | 68.3 | 75.0 | 75.3 | 71.7 | 76.7 |
| Sensitivity | 80.0 | 67.5 | 73.3 | 76.7 | 73.3 | 76.7 | 90.0 | 90.0 | 80.0 | 86.7 | 83.3 |
| Specificity | 73.3 | 90.0 | 73.9 | 66.7 | 66.7 | 73.3 | 46.7 | 60.0 | 76.7 | 56.7 | 70.0 |
| G-measure | 75.5 | 77.7 | 73.2 | 70.6 | 68.9 | 74.3 | 63.6 | 72.2 | 77.3 | 69.7 | 75.5 |
| Subject 8 | | | | | | | | | | | |
| Accuracy | 83.3 | 55.0 | 48.3 | 75.0 | 70.0 | 73.3 | 51.7 | 70.0 | 78.3 | 78.3 | 81.7 |
| Sensitivity | 93.3 | 54.3 | 48.9 | 73.3 | 100.0 | 100.0 | 100.0 | 100.0 | 80.0 | 100.0 | 93.3 |
| Specificity | 73.3 | 56.9 | 47.9 | 76.7 | 40.0 | 46.7 | 3.3 | 40.0 | 76.7 | 56.7 | 70.0 |
| G-measure | 82.2 | 55.4 | 48.3 | 74.5 | 54.7 | 65.3 | 8.2 | 55.0 | 78.1 | 73.2 | 80.3 |
| Subject 9 | | | | | | | | | | | |
| Accuracy | 71.7 | 66.7 | 61.7 | 80.0 | 71.7 | 75.0 | 65.0 | 72.3 | 76.7 | 63.3 | 80.0 |
| Sensitivity | 80.0 | 74.2 | 71.0 | 80.0 | 93.3 | 83.3 | 96.7 | 90.0 | 80.0 | 90.0 | 80.0 |
| Specificity | 63.3 | 73.8 | 58.8 | 80.0 | 50.0 | 66.7 | 33.3 | 46.7 | 73.3 | 36.7 | 80.0 |
| G-measure | 66.6 | 73.9 | 64.1 | 74.3 | 57.7 | 69.1 | 41.4 | 59.1 | 70.2 | 47.4 | 74.3 |
| Subject 10 | | | | | | | | | | | |
| Accuracy | 65.0 | 61.7 | 60.0 | 80.0 | 70.0 | 78.3 | 56.7 | 66.7 | 80.2 | 71.7 | 81.7 |
| Sensitivity | 56.7 | 50.0 | 65.0 | 76.7 | 86.7 | 76.7 | 93.3 | 83.3 | 80.0 | 83.3 | 80.0 |
| Specificity | 73.3 | 72.4 | 66.1 | 83.3 | 53.3 | 80.0 | 20.0 | 50.0 | 83.3 | 60.0 | 83.3 |
| G-measure | 52.8 | 56.0 | 65.5 | 79.2 | 67.5 | 77.3 | 37.3 | 60.9 | 81.0 | 68.3 | 81.0 |

**Table 1** continued

| Measures | LDA | NN | NB | SVM | KNN | 5-MV | 5-UV | 5-WV | 3-MV | 3-UV | 3-WV (proposed method) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Average | | | | | | | | | | | |
| Accuracy | 74.5 | 65.2 | 64.4 | 75.9 | 71.0 | 75.5 | 61.2 | 74.8 | 77.1 | 72.1 | 80.7 |
| Sensitivity | 76.6 | 65.7 | 67.3 | 74.0 | 85.4 | 80.7 | 94.2 | 86.5 | 77.4 | 88.4 | 78.6 |
| Specificity | 72.3 | 69.4 | 66.6 | 76.4 | 55.9 | 69.8 | 29.7 | 70.9 | 76.9 | 51.0 | 76.7 |
| G-measure | 71.4 | 66.2 | 66.5 | 72.9 | 64.5 | 73.0 | 43.6 | 71.4 | 74.7 | 62.9 | 75.8 |

$$w_i = \frac{acc_i}{\Sigma_{i=1}^{m} acc_i} \tag{10}$$

where $acc_i$ is accuracy at $i$th classifier

For evaluating the performance of all five classifiers, we have utilized a classifier ranking approach, which selects the best- performing classifier among the five classifiers. In order to find the rank of the classifier, we have considered *G*-measure as the measure of calculation. Data have been partitioned into two parts training and testing in 9:1 ratio randomly, this procedure is repeated for *n* iterations, and mean of *n* iterations is considered for ranking of the classifiers. Mean of weights of respective classifier for all iterations is assigned to the classifier.

# 4 Experimental results and analysis

EEG data for deception detection have been recorded using brain vision analyzer and recorder (brain products, Germany [19]). Recorded data were analyzed on the system with 8 GB RAM Intel i7 processor and implemented on MATLAB 2015a. Results of various feature extraction approaches like potential, mobility, complexity, power, frequency response and wavelet are discussed in this section. Also, comparison of proposed 3-classifier ensemble framework with the 5-classifier framework and with the base classifier has been presented.

## 4.1 Feature extraction

EEG is a signal waveform, and its amplitude provides the value of a signal at each peak; power provides the strength of the signal; the frequency component of a signal is extracted using FFT and statistical parameter using Hjorth features (mobility and complexity); and wavelet is used to extract time–frequency components of EEG signal. Hence, instead of using a single type of feature extraction technique like [8,11], we have used various feature extraction techniques to analyze EEG data more precisely. After signal preprocessing, each feature extraction approach (as discussed in Sect. 3) has been applied on single subjects' 16 channel for 30

trials of one session ($1 \times 16 \times 30$). The experiment is conducted for two sessions where subjects behaved as guilty for one session and innocent for another session. While taking results, data of 9 subjects out of 10 subjects have been considered. One subject (i.e., subject number 6) data have not been acquired properly and have lots of artifacts; hence, it is not considered for this study. For comparative analysis, various feature extraction approaches with various classification techniques and ensemble frameworks (with 5-classifier ensemble framework and with 3-classifier ensemble framework) are aggregated by majority voting, unanimous voting and weighted voting with respective subjects. The results are depicted in Fig. 4. These results show mean of the performance evaluated by applying subject-wise fivefold cross-validation (5-FCV). Here, 5-MV represents majority voting using 5 classifiers, 5-UV represents unanimous voting using 5 classifiers, and 5-WV represents weighted voting using 5 classifiers. Similarly, 3-MV, 3-UV and 3-WV represent majority voting, unanimous voting and weighted voting using 3 top-ranked classifiers, respectively. From Fig. 4, it is inferred that in most of the cases mobility and wavelet have performed better than other feature extraction techniques. It is observed that the combined result of proposed approach (3-WV) is higher than others for most of the feature extraction approaches.

A similar set of features have been utilized by R. Jenke et al. for emotion recognition dataset [20]. After applying various feature selection approaches, they have achieved an average accuracy of 35.9% using LDA as classifier, with highest accuracy of 45% for subject 6 using mRMR feature selection approach. Using proposed 3-WV framework, an average accuracy of 84.6% has been achieved using comparatively less channels.

## 4.2 Classification

For classification, data have been labeled into two classes: guilty as *Class-1* and innocent as *Class-2*. The performance of classifiers is measured using various performance measures such as accuracy, sensitivity, specificity and *G*-measure. Specificity and sensitivity provide negative class and positive class accuracy where *G*-measures consider

**Table 2** Subject-wise performance using various classifiers and ensemble framework applied on complexity values

| Measures | LDA | NN | NB | SVM | KNN | 5-MV | 5-UV | 5-WV | 3-MV | 3-UV | 3-WV (proposed method) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Subject 1** | | | | | | | | | | | |
| Accuracy | 81.1 | 60.1 | 50.8 | 56.2 | 69.4 | 80.1 | 56.5 | 82.1 | 82.1 | 58.3 | 83.0 |
| Sensitivity | 83.3 | 62.0 | 65.8 | 16.7 | 70.0 | 73.3 | 100.0 | 73.3 | 80.0 | 100.0 | 83.3 |
| Specificity | 81.1 | 65.0 | 49.5 | 100.0 | 68.0 | 92.0 | 8.0 | 92.0 | 85.1 | 12.0 | 81.1 |
| G-measure | 78.0 | 63.4 | 57.1 | 18.3 | 58.2 | 72.3 | 12.6 | 72.3 | 78.8 | 15.5 | 78.0 |
| **Subject 2** | | | | | | | | | | | |
| Accuracy | 60.0 | 46.0 | 59.1 | 66.2 | 62.2 | 63.3 | 56.0 | 66.9 | 65.3 | 59.8 | 69.3 |
| Sensitivity | 52.9 | 50.0 | 58.1 | 58.7 | 77.3 | 58.7 | 95.6 | 85.3 | 52.9 | 79.1 | 64.9 |
| Specificity | 67.1 | 51.2 | 62.1 | 73.8 | 47.1 | 69.8 | 10.2 | 40.9 | 77.8 | 40.4 | 73.8 |
| G-measure | 58.6 | 45.5 | 59.8 | 63.2 | 58.4 | 62.2 | 18.5 | 51.7 | 63.0 | 49.8 | 67.7 |
| **Subject 3** | | | | | | | | | | | |
| Accuracy | 80.0 | 66.7 | 66.7 | 71.7 | 73.3 | 76.7 | 68.3 | 78.3 | 81.7 | 68.3 | 82.7 |
| Sensitivity | 86.7 | 63.5 | 67.2 | 76.7 | 93.3 | 80.0 | 100.0 | 93.3 | 93.3 | 100.0 | 93.3 |
| Specificity | 73.3 | 71.5 | 67.3 | 66.7 | 53.3 | 73.3 | 36.7 | 63.3 | 70.0 | 36.7 | 70.0 |
| G-measure | 79.2 | 67.2 | 67.0 | 70.7 | 68.9 | 75.6 | 59.5 | 76.2 | 80.5 | 54.0 | 80.5 |
| **Subject 4** | | | | | | | | | | | |
| Accuracy | 85.3 | 90.7 | 91.6 | 87.6 | 86.7 | 94.7 | 80.4 | 89.8 | 91.8 | 81.6 | 91.8 |
| Sensitivity | 80.9 | 91.0 | 93.3 | 83.6 | 93.3 | 93.3 | 95.6 | 95.6 | 87.6 | 89.8 | 87.6 |
| Specificity | 89.8 | 95.0 | 93.8 | 91.6 | 80.0 | 96.0 | 65.3 | 84.0 | 96.0 | 73.3 | 96.0 |
| G-measure | 83.8 | 92.5 | 93.2 | 86.1 | 84.8 | 94.2 | 76.7 | 88.9 | 91.0 | 79.3 | 91.0 |
| **Subject 5** | | | | | | | | | | | |
| Accuracy | 70.2 | 59.3 | 62.2 | 66.2 | 45.1 | 65.3 | 54.2 | 73.3 | 68.2 | 67.3 | 74.2 |
| Sensitivity | 71.1 | 43.8 | 71.8 | 56.9 | 56.9 | 62.7 | 97.8 | 67.1 | 55.1 | 83.6 | 73.3 |
| Specificity | 69.3 | 48.9 | 60.7 | 75.6 | 33.3 | 65.8 | 16.9 | 63.6 | 69.3 | 51.1 | 67.1 |
| G-measure | 65.6 | 41.7 | 65.6 | 55.4 | 38.5 | 61.0 | 30.5 | 63.3 | 57.7 | 60.0 | 65.6 |
| **Subject 7** | | | | | | | | | | | |
| Accuracy | 80.5 | 65.2 | 67.9 | 81.0 | 60.1 | 75.0 | 66.2 | 78.7 | 80.5 | 68.4 | 82.8 |
| Sensitivity | 73.3 | 63.7 | 68.8 | 83.3 | 63.3 | 80.0 | 96.7 | 86.7 | 83.3 | 90.0 | 83.3 |
| Specificity | 89.8 | 81.3 | 66.0 | 77.8 | 57.3 | 69.8 | 36.0 | 69.8 | 77.8 | 46.7 | 77.8 |
| G-measure | 80.3 | 71.0 | 67.3 | 80.1 | 56.3 | 74.2 | 51.0 | 77.5 | 80.3 | 63.3 | 80.3 |
| **Subject 8** | | | | | | | | | | | |
| Accuracy | 85.0 | 70.0 | 58.3 | 83.3 | 65.0 | 76.7 | 68.3 | 75.0 | 80.0 | 78.3 | 80.0 |
| Sensitivity | 93.3 | 69.9 | 61.7 | 90.0 | 80.0 | 86.7 | 96.7 | 93.3 | 90.0 | 96.7 | 90.0 |
| Specificity | 76.7 | 69.2 | 56.9 | 76.7 | 50.0 | 66.7 | 40.0 | 56.7 | 70.0 | 60.0 | 70.0 |
| G-measure | 83.8 | 64.4 | 59.2 | 80.3 | 52.3 | 67.8 | 54.3 | 64.2 | 77.1 | 67.6 | 77.1 |
| **Subject 9** | | | | | | | | | | | |
| Accuracy | 70.3 | 70.0 | 61.7 | 80.0 | 56.7 | 75.0 | 53.3 | 70.0 | 75.0 | 61.7 | 86.7 |
| Sensitivity | 93.3 | 80.5 | 60.0 | 86.7 | 86.7 | 90.0 | 93.3 | 90.0 | 86.7 | 93.3 | 86.7 |
| Specificity | 63.3 | 92.0 | 61.4 | 73.3 | 26.7 | 60.0 | 13.3 | 50.0 | 63.3 | 30.0 | 86.7 |
| G-measure | 74.4 | 85.0 | 56.4 | 76.2 | 35.7 | 69.6 | 21.0 | 60.3 | 69.6 | 42.8 | 84.4 |
| **Subject 10** | | | | | | | | | | | |
| Accuracy | 68.3 | 63.3 | 55.0 | 70.0 | 56.7 | 73.3 | 56.7 | 68.0 | 70.1 | 68.3 | 73.3 |
| Sensitivity | 66.7 | 55.9 | 53.1 | 63.3 | 73.3 | 83.3 | 93.3 | 90.0 | 73.3 | 86.7 | 73.3 |
| Specificity | 70.0 | 71.6 | 63.0 | 76.7 | 40.0 | 63.3 | 20.0 | 40.0 | 73.3 | 50.0 | 73.3 |
| G-measure | 66.0 | 58.5 | 57.5 | 66.5 | 44.2 | 69.1 | 32.8 | 57.2 | 70.9 | 63.4 | 70.9 |

**Table 2** continued

| Measures | LDA | NN | NB | SVM | KNN | 5-MV | 5-UV | 5-WV | 3-MV | 3-UV | 3-WV (proposed method) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Average | | | | | | | | | | | |
| Accuracy | 73.18 | 65.7 | 63.7 | 73.6 | 63.9 | 75.6 | 62.2 | 75.8 | 77.2 | 68.0 | 80.4 |
| Sensitivity | 78.0 | 64.5 | 66.6 | 69.0 | 77.1 | 78.7 | 96.5 | 80.4 | 77.9 | 91.0 | 81.8 |
| Specificity | 75.6 | 71.7 | 64.5 | 79.1 | 50.6 | 73.0 | 27.4 | 62.2 | 75.9 | 44.5 | 77.3 |
| G-measure | 74.4 | 65.4 | 64.0 | 66.3 | 55.2 | 71.7 | 46.5 | 67.9 | 74.3 | 55.0 | 77.2 |

both negative and positive class accuracies to calculate performance score. *G*-measure gives geometric mean of specificity and sensitivity. Five classifiers have been considered here for evaluating the performance of system. Using a classifier ranking approach, three best classifiers among five classifiers have been identified. Results of classifiers are aggregated following an ensemble framework. For aggregation, weighted voting approach has been used and its performance is compared with other aggregation approaches as majority voting and unanimous voting. For weight updation instead of using conventional weight update formula 11, we updated weight as Eq. 10.

$$w_{i0} = w_{i1} + \frac{1}{2} \log \left( \frac{acc_i}{1 - acc_i} \right) \qquad (11)$$

where $w_{i0}$ and $w_{i1}$ represent old weight and updated weight, respectively, for $i$th classifier at $n$th iteration and $acc_i$ represents accuracy of $i$th classifier. Equation 11 provides logarithm of the accuracy of classifier, which makes weight update as zero, provided accuracy is 0 or 100%. Hence, in order to overcome that limitation normalization approach is considered.

Using various classification performance measures, results are calculated and are presented in Tables 1, 2, 3, 4, 5 and 6. Results are evaluated by applying 5-FCV on subject-wise EEG data. Results of base classifiers with 5-classifier ensemble framework and 3-classifier ensemble framework are compared. Each table represents results evaluated from various feature extraction approaches. Result after applying 5-FCV on potential as the feature is given in Table 1. Table 1 compares performance to single classifiers with 5-classifier ensemble framework and 3-classifier ensemble framework. Performance of classification for each subject is compared using various measures like accuracy, sensitivity, specificity and *G*-measure. Using amplitude as feature, proposed framework achieves accuracy of 89% for subject 4 which is highest among all subjects. In most of the cases, *G*-measure value is higher for proposed framework and compromised performance of sensitivity and specificity. As sensitivity and specificity provides measure toward a particular class, its value is higher for unanimous voting approach, which provides decision in favor of one class.

In Table 2, performance of various subjects using complexity as feature is tabulated. All classifiers are applied on features extracted using complexity, and the results are compared with proposed ensemble framework. Using complexity as feature, an accuracy of 91.8% with subject 4 and an average accuracy of 80.4% has been achieved. Table 3 represents comparative performance of ensemble framework with other classifiers using frequency response. With frequency response as a feature, proposed framework has attained a highest accuracy of 98.8% for subject 4 and an average accuracy of all subjects as 76.3%.

Table 4 shows results of various classification approaches using one of the Hjorth's parameter, viz. mobility. An accuracy of 96.9% has been achieved using weighted voting approach for 3-classifiers and an average accuracy of 89.8%.

In Table 5, power values extracted from EEG data have been classified and compared achieving highest accuracy of 100% for subject 4 and an average accuracy of 88.2%.

Table 6 provides a comparative analysis of various classifiers and ensemble framework using wavelet as feature. After applying classifiers on wavelet features, an accuracy of 100% for subject 3 and an average accuracy of 92.4% are attained, which is highest among all.

From the results depicted in Tables 1, 2, 3, 4, 5 and 6, it can be inferred that from base classifier performance 5-classifier system (5-MV, 5-UV and 5-WV) and 3-classifier system (3-MV, 3-UV and 3-WV) have performed better for almost all feature extraction approaches. On comparing the performance of different feature extraction approaches, classification accuracy is improved when wavelets as feature is used (as depicted in Table 6). Also, performance of classifiers is similar for Hjorth's parameters. Among various subjects, subject-4 has responded best for all features, thus providing highest classification accuracy for almost all classifiers. The classifier ranking approach is applied to aggregate the best-performing classifiers. A graph depicting the comparison between two systems is given in Fig. 5. It is observed that the ensemble framework with ranking approach has a great impact toward the improvement of classification accuracy. Results of 3-classifier ensemble system are aggregated using three different approaches namely majority voting, unanimous voting and weighted voting. From the results, it is

An efficient Concealed Information Test: EEG feature extraction and ensemble classification...

825

**Table 3** Subject-wise performance using various classifiers and ensemble framework applied on frequency response values

| Measures | LDA | NN | NB | SVM | KNN | 5-MV | 5-UV | 5-WV | 3-MV | 3-UV | 3-WV (proposed method) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Subject 1** | | | | | | | | | | | |
| Accuracy | 75.4 | 66.3 | 57.2 | 76.6 | 69.4 | 76.1 | 54.1 | 77.4 | 77.5 | 73.6 | 79.0 |
| Sensitivity | 80.0 | 67.4 | 55.6 | 73.3 | 73.3 | 73.3 | 86.7 | 73.3 | 78.5 | 73.6 | 78.6 |
| Specificity | 71.4 | 58.9 | 68.3 | 81.1 | 64.0 | 80.0 | 17.7 | 80.0 | 73.3 | 86.7 | 80.0 |
| G-measure | 73.9 | 58.1 | 61.1 | 74.4 | 64.6 | 73.8 | 34.9 | 73.8 | 85.1 | 59.4 | 67.4 |
| **Subject 2** | | | | | | | | | | | |
| Accuracy | 72.5 | 77.0 | 77.5 | 74.3 | 76.1 | 77.5 | 73.8 | 80.6 | 79.3 | 77.5 | 81.3 |
| Sensitivity | 71.6 | 76.9 | 77.7 | 71.1 | 89.8 | 79.6 | 96.0 | 96.0 | 79.3 | 77.5 | 79.3 |
| Specificity | 70.0 | 88.3 | 77.0 | 76.7 | 60.0 | 73.3 | 50.0 | 63.3 | 79.6 | 89.8 | 79.6 |
| G-measure | 67.1 | 81.4 | 77.1 | 72.2 | 64.0 | 74.0 | 60.3 | 74.3 | 76.7 | 63.3 | 76.7 |
| **Subject 3** | | | | | | | | | | | |
| Accuracy | 75.0 | 70.0 | 56.7 | 71.7 | 75.0 | 76.7 | 66.7 | 70.0 | 73.9 | 65.0 | 75.0 |
| Sensitivity | 80.0 | 74.4 | 60.0 | 70.0 | 86.7 | 83.3 | 96.7 | 90.0 | 75.0 | 65.0 | 75.0 |
| Specificity | 70.0 | 66.7 | 55.0 | 73.3 | 63.3 | 70.0 | 36.7 | 50.0 | 76.7 | 86.7 | 76.7 |
| G-measure | 71.8 | 65.2 | 57.3 | 69.2 | 73.0 | 73.5 | 45.4 | 56.0 | 73.3 | 43.3 | 73.3 |
| **Subject 4** | | | | | | | | | | | |
| Accuracy | 95.6 | 98.0 | 86.9 | 95.8 | 98.0 | 96.9 | 88.0 | 98.0 | 95.8 | 95.8 | 98.8 |
| Sensitivity | 91.1 | 96.7 | 83.5 | 95.6 | 96.0 | 97.8 | 100.0 | 100.0 | 95.8 | 95.8 | 95.8 |
| Specificity | 100.0 | 100.0 | 98.0 | 96.0 | 100.0 | 96.0 | 76.0 | 96.0 | 95.6 | 95.6 | 95.6 |
| G-measure | 94.9 | 98.3 | 89.9 | 95.5 | 97.9 | 96.7 | 86.0 | 97.9 | 96.0 | 96.0 | 96.0 |
| **Subject 5** | | | | | | | | | | | |
| Accuracy | 73.0 | 72.9 | 72.4 | 78.9 | 79.6 | 74.9 | 72.7 | 77.6 | 74.9 | 70.7 | 78.9 |
| Sensitivity | 96.0 | 90.0 | 85.7 | 97.8 | 89.8 | 93.8 | 100.0 | 97.8 | 94.9 | 90.7 | 96.9 |
| Specificity | 96.0 | 98.0 | 81.7 | 100.0 | 69.3 | 96.0 | 65.3 | 77.3 | 93.8 | 100.0 | 97.8 |
| G-measure | 96.0 | 93.7 | 83.4 | 98.9 | 78.4 | 94.6 | 80.7 | 87.0 | 96.0 | 81.3 | 96.0 |
| **Subject 7** | | | | | | | | | | | |
| Accuracy | 83.3 | 81.7 | 71.7 | 80.0 | 71.7 | 80.0 | 78.3 | 80.0 | 81.7 | 70.0 | 84.7 |
| Sensitivity | 80.0 | 80.3 | 70.0 | 80.0 | 76.7 | 86.7 | 100.0 | 90.0 | 81.7 | 70.0 | 81.7 |
| Specificity | 86.7 | 90.5 | 74.0 | 80.0 | 66.7 | 73.3 | 56.7 | 70.0 | 83.3 | 90.0 | 83.3 |
| G-measure | 82.9 | 84.7 | 71.9 | 78.8 | 70.4 | 78.6 | 74.7 | 78.9 | 80.0 | 50.0 | 80.0 |
| **Subject 8** | | | | | | | | | | | |
| Accuracy | 88.3 | 78.3 | 70.0 | 88.3 | 86.7 | 85.0 | 86.7 | 90.0 | 90.0 | 91.7 | 95.0 |
| Sensitivity | 80.0 | 78.6 | 79.8 | 80.0 | 86.7 | 76.7 | 100.0 | 96.7 | 90.0 | 91.7 | 95.0 |
| Specificity | 96.7 | 81.4 | 69.3 | 96.7 | 86.7 | 93.3 | 73.3 | 83.3 | 83.3 | 93.3 | 93.3 |
| G-measure | 87.2 | 79.7 | 74.4 | 86.3 | 86.0 | 82.8 | 85.2 | 89.5 | 96.7 | 90.0 | 96.7 |
| **Subject 9** | | | | | | | | | | | |
| Accuracy | 75.0 | 68.3 | 51.7 | 81.7 | 73.3 | 78.3 | 68.3 | 81.7 | 48.3 | 53.3 | 50.0 |
| Sensitivity | 93.3 | 71.3 | 56.2 | 90.0 | 83.3 | 90.0 | 100.0 | 100.0 | 48.3 | 53.3 | 50.0 |
| Specificity | 56.7 | 70.0 | 60.1 | 73.3 | 63.3 | 66.7 | 36.7 | 63.3 | 50.0 | 86.7 | 50.0 |
| G-measure | 64.4 | 70.6 | 53.4 | 70.9 | 65.6 | 67.9 | 52.8 | 70.5 | 46.7 | 20.0 | 50.0 |
| **Subject 10** | | | | | | | | | | | |
| Accuracy | 73.3 | 75.0 | 61.7 | 80.0 | 76.7 | 76.7 | 73.3 | 75.3 | 46.7 | 48.3 | 46.7 |
| Sensitivity | 70.0 | 80.3 | 62.5 | 73.3 | 90.0 | 76.7 | 96.7 | 80.0 | 46.7 | 48.3 | 46.7 |
| Specificity | 76.7 | 76.3 | 62.1 | 86.7 | 63.3 | 76.7 | 50.0 | 66.7 | 60.0 | 70.0 | 60.0 |
| G-measure | 61.7 | 77.6 | 57.0 | 75.2 | 74.2 | 74.0 | 66.2 | 68.9 | 33.3 | 26.7 | 33.3 |

**Table 3** continued

| Measures | LDA | NN | NB | SVM | KNN | 5-MV | 5-UV | 5-WV | 3-MV | 3-UV | 3-WV (proposed method) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Average | | | | | | | | | | | |
| Accuracy | 79.0 | 76.4 | 67.3 | 80.8 | 78.5 | 80.2 | 73.5 | 81.2 | 74.2 | 71.8 | 76.6 |
| Sensitivity | 82.4 | 79.6 | 70.1 | 81.2 | 85.8 | 84.2 | 97.3 | 91.5 | 76.0 | 74.0 | 77.6 |
| Specificity | 80.5 | 81.1 | 71.7 | 84.9 | 70.7 | 80.6 | 51.4 | 73.0 | 77.3 | 88.7 | 79.6 |
| G-measure | 77.7 | 78.8 | 69.5 | 80.1 | 75.0 | 79.5 | 65.1 | 77.4 | 76.0 | 58.9 | 74.4 |

**Table 4** Subject-wise performance using various classifiers and ensemble framework applied on mobility values

| Measures | LDA | NN | NB | SVM | KNN | 5-MV | 5-UV | 5-WV | 3-MV | 3-UV | 3-WV (proposed method) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject 1 | | | | | | | | | | | |
| Accuracy | 74.5 | 75.7 | 51.5 | 78.5 | 67.3 | 78.2 | 50.5 | 78.2 | 80.0 | 73.0 | 81.5 |
| Sensitivity | 76.7 | 80.8 | 53.9 | 76.7 | 80.0 | 76.7 | 86.7 | 76.7 | 76.7 | 83.3 | 76.7 |
| Specificity | 72.0 | 86.0 | 44.0 | 81.1 | 52.0 | 80.0 | 9.7 | 80.0 | 84.0 | 61.1 | 81.1 |
| G-measure | 67.2 | 81.8 | 44.0 | 74.7 | 49.0 | 73.8 | 15.9 | 73.8 | 76.2 | 65.3 | 74.7 |
| Subject 2 | | | | | | | | | | | |
| Accuracy | 77.8 | 83.8 | 81.4 | 85.9 | 85.9 | 85.5 | 75.2 | 86.7 | 87.4 | 74.7 | 88.4 |
| Sensitivity | 81.3 | 86.7 | 88.8 | 85.3 | 92.0 | 89.8 | 92.0 | 92.0 | 87.6 | 92.0 | 87.6 |
| Specificity | 76.7 | 88.3 | 90.0 | 90.0 | 80.0 | 90.0 | 56.7 | 83.3 | 86.7 | 56.7 | 86.7 |
| G-measure | 78.6 | 86.6 | 89.0 | 87.1 | 81.9 | 89.1 | 68.7 | 86.0 | 86.4 | 69.9 | 86.4 |
| Subject 3 | | | | | | | | | | | |
| Accuracy | 76.7 | 75.0 | 73.3 | 84.0 | 78.3 | 83.3 | 70.0 | 81.7 | 85.0 | 81.7 | 86.0 |
| Sensitivity | 76.7 | 78.5 | 77.8 | 80.0 | 70.0 | 83.3 | 93.3 | 90.0 | 80.0 | 90.0 | 80.0 |
| Specificity | 76.7 | 86.3 | 77.3 | 90.0 | 86.7 | 83.3 | 46.7 | 73.3 | 90.0 | 73.3 | 90.0 |
| G-measure | 75.4 | 80.9 | 76.7 | 82.9 | 76.3 | 80.7 | 62.8 | 79.8 | 82.9 | 80.3 | 82.9 |
| Subject 4 | | | | | | | | | | | |
| Accuracy | 95.8 | 95.8 | 94.7 | 94.9 | 95.8 | 96.9 | 94.7 | 95.8 | 95.2 | 96.9 | 96.9 |
| Sensitivity | 91.6 | 97.8 | 95.0 | 92.0 | 96.0 | 96.0 | 96.0 | 96.0 | 93.8 | 96.0 | 93.8 |
| Specificity | 100.0 | 94.4 | 96.7 | 97.8 | 95.6 | 97.8 | 93.3 | 95.6 | 100.0 | 97.8 | 100.0 |
| G-measure | 95.5 | 96.0 | 95.6 | 94.6 | 95.5 | 96.7 | 94.2 | 95.5 | 96.7 | 96.7 | 96.7 |
| Subject 5 | | | | | | | | | | | |
| Accuracy | 94.0 | 91.6 | 89.6 | 92.8 | 85.6 | 91.6 | 81.8 | 89.6 | 91.5 | 86.9 | 95.8 |
| Sensitivity | 100.0 | 92.2 | 88.9 | 95.6 | 91.6 | 95.6 | 100.0 | 95.6 | 95.6 | 100.0 | 95.6 |
| Specificity | 96.0 | 92.2 | 92.2 | 100.0 | 79.6 | 95.6 | 63.6 | 95.6 | 100.0 | 73.8 | 100.0 |
| G-measure | 97.9 | 92.1 | 90.3 | 97.6 | 84.9 | 95.6 | 79.2 | 95.6 | 97.6 | 83.6 | 97.6 |
| Subject 7 | | | | | | | | | | | |
| Accuracy | 86.7 | 63.3 | 63.3 | 76.7 | 70.0 | 76.7 | 70.0 | 75.0 | 81.7 | 80.0 | 88.0 |
| Sensitivity | 86.7 | 62.0 | 58.0 | 66.7 | 66.7 | 76.7 | 86.7 | 83.3 | 76.7 | 90.0 | 86.7 |
| Specificity | 86.7 | 73.9 | 63.7 | 86.7 | 73.3 | 76.7 | 53.3 | 66.7 | 86.7 | 70.0 | 83.3 |
| G-measure | 83.9 | 66.8 | 56.8 | 71.0 | 67.2 | 73.4 | 66.3 | 72.1 | 78.9 | 68.7 | 82.2 |
| Subject 8 | | | | | | | | | | | |
| Accuracy | 90.0 | 86.7 | 65.0 | 95.0 | 68.3 | 93.3 | 63.3 | 88.3 | 95.0 | 78.3 | 95.0 |
| Sensitivity | 96.7 | 83.7 | 68.3 | 96.7 | 86.7 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Specificity | 83.3 | 96.7 | 55.8 | 93.3 | 50.0 | 86.7 | 26.7 | 76.7 | 90.0 | 56.7 | 90.0 |
| G-measure | 88.8 | 89.5 | 57.9 | 94.6 | 63.6 | 92.8 | 42.8 | 87.4 | 94.6 | 72.1 | 94.6 |

**Table 4** continued

| Measures | LDA | NN | NB | SVM | KNN | 5-MV | 5-UV | 5-WV | 3-MV | 3-UV | 3-WV (proposed method) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject 9 | | | | | | | | | | | |
| Accuracy | 86.7 | 83.3 | 60.0 | 81.7 | 76.7 | 85.0 | 73.3 | 86.3 | 90.0 | 73.3 | 90.0 |
| Sensitivity | 96.7 | 84.2 | 51.2 | 96.7 | 96.7 | 96.7 | 100.0 | 100.0 | 96.7 | 100.0 | 96.7 |
| Specificity | 76.7 | 95.0 | 60.2 | 66.7 | 56.7 | 73.3 | 46.7 | 66.7 | 83.3 | 46.7 | 83.3 |
| G-measure | 83.1 | 88.4 | 55.5 | 69.8 | 63.1 | 80.6 | 58.8 | 71.5 | 86.4 | 59.4 | 86.4 |
| Subject 10 | | | | | | | | | | | |
| Accuracy | 80.0 | 73.3 | 63.3 | 83.3 | 73.3 | 81.7 | 73.3 | 85.0 | 85.0 | 76.7 | 86.7 |
| Sensitivity | 80.0 | 80.0 | 65.8 | 83.3 | 80.0 | 80.0 | 100.0 | 96.7 | 86.7 | 90.0 | 66.7 |
| Specificity | 80.0 | 74.4 | 68.3 | 83.3 | 66.7 | 83.3 | 46.7 | 73.3 | 83.3 | 63.3 | 86.7 |
| G-measure | 77.4 | 77.1 | 66.5 | 81.1 | 68.7 | 79.3 | 66.9 | 83.7 | 83.3 | 71.6 | 71.0 |
| Average | | | | | | | | | | | |
| Accuracy | 84.6 | 80.9 | 71.3 | 85.8 | 77.9 | 85.7 | 72.4 | 85.2 | 87.8 | 80.2 | 89.8 |
| Sensitivity | 87.3 | 82.8 | 71.9 | 85.8 | 84.4 | 88.3 | 94.9 | 92.2 | 88.2 | 93.4 | 87.1 |
| Specificity | 83.1 | 87.5 | 72.0 | 86.5 | 71.2 | 85.1 | 49.2 | 79.0 | 89.3 | 66.5 | 87.0 |
| G-measure | 83.1 | 84.4 | 70.3 | 82.3 | 72.2 | 84.6 | 61.7 | 80.8 | 87.0 | 74.1 | 85.8 |

**Table 5** Subject-wise performance using various classifiers and ensemble framework applied on power values

| Measures | LDA | NN | NB | SVM | KNN | 5-MV | 5-UV | 5-WV | 3-MV | 3-UV | 3-WV (proposed method) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject 1 | | | | | | | | | | | |
| Accuracy | 81.8 | 78.2 | 50.2 | 80.3 | 69.1 | 80.8 | 64.2 | 82.8 | 82.1 | 76.6 | 83.0 |
| Sensitivity | 90.0 | 78.7 | 52.5 | 76.7 | 86.7 | 83.3 | 93.3 | 83.3 | 80.0 | 90.0 | 80.0 |
| Specificity | 72.0 | 84.9 | 49.4 | 84.0 | 48.0 | 80.0 | 30.3 | 80.0 | 84.0 | 61.1 | 84.0 |
| G-measure | 75.7 | 81.2 | 50.7 | 75.4 | 55.4 | 77.8 | 45.8 | 77.8 | 77.0 | 69.1 | 77.0 |
| Subject 2 | | | | | | | | | | | |
| Accuracy | 72.1 | 78.3 | 83.8 | 83.4 | 75.6 | 82.4 | 75.6 | 83.8 | 83.4 | 82.4 | 86.4 |
| Sensitivity | 68.9 | 79.8 | 86.5 | 83.1 | 89.8 | 89.8 | 92.0 | 92.0 | 85.3 | 92.0 | 85.3 |
| Specificity | 73.3 | 87.6 | 83.3 | 83.3 | 60.0 | 73.3 | 56.7 | 73.3 | 80.0 | 70.0 | 80.0 |
| G-measure | 69.5 | 82.8 | 84.8 | 82.5 | 69.7 | 79.6 | 62.0 | 80.3 | 81.4 | 76.9 | 81.4 |
| Subject 3 | | | | | | | | | | | |
| Accuracy | 83.3 | 53.3 | 65.0 | 76.7 | 61.7 | 73.3 | 60.0 | 80.0 | 80.0 | 75.0 | 80.5 |
| Sensitivity | 76.7 | 60.7 | 64.0 | 76.7 | 63.3 | 80.0 | 96.7 | 63.3 | 80.0 | 86.7 | 80.0 |
| Specificity | 90.0 | 72.1 | 67.3 | 76.7 | 60.0 | 66.7 | 23.3 | 96.7 | 80.0 | 63.3 | 76.7 |
| G-measure | 82.2 | 66.1 | 65.6 | 76.3 | 59.8 | 72.9 | 40.8 | 78.0 | 79.7 | 73.6 | 77.7 |
| Subject 4 | | | | | | | | | | | |
| Accuracy | 98.9 | 97.8 | 84.0 | 96.7 | 96.9 | 100.0 | 81.6 | 100.0 | 100.0 | 97.8 | 100.0 |
| Sensitivity | 97.8 | 96.4 | 81.0 | 97.8 | 93.8 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Specificity | 100.0 | 100.0 | 96.0 | 95.6 | 100.0 | 100.0 | 63.1 | 100.0 | 100.0 | 95.6 | 100.0 |
| G-measure | 98.9 | 98.1 | 87.6 | 96.6 | 96.7 | 100.0 | 78.1 | 100.0 | 100.0 | 97.6 | 100.0 |
| Subject 5 | | | | | | | | | | | |
| Accuracy | 90.0 | 89.0 | 82.4 | 94.0 | 83.8 | 89.0 | 81.8 | 89.8 | 88.0 | 86.0 | 95.0 |
| Sensitivity | 88.0 | 91.0 | 87.7 | 96.0 | 88.0 | 96.0 | 100.0 | 96.0 | 96.0 | 100.0 | 96.0 |
| Specificity | 92.0 | 100.0 | 81.0 | 96.0 | 79.6 | 96.0 | 63.6 | 83.6 | 96.0 | 88.0 | 96.0 |
| G-measure | 89.8 | 95.2 | 83.9 | 96.0 | 82.8 | 96.0 | 79.2 | 89.4 | 96.0 | 92.6 | 96.0 |

**Table 5** continued

| Measures | LDA | NN | NB | SVM | KNN | 5-MV | 5-UV | 5-WV | 3-MV | 3-UV | 3-WV (proposed method) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Subject 7** | | | | | | | | | | | |
| Accuracy | 88.3 | 78.3 | 78.3 | 86.7 | 75.0 | 83.3 | 75.0 | 83.3 | 90.0 | 81.7 | 91.0 |
| Sensitivity | 96.7 | 78.1 | 74.0 | 96.7 | 80.0 | 90.0 | 100.0 | 96.7 | 96.7 | 100.0 | 96.7 |
| Specificity | 80.0 | 84.3 | 85.3 | 76.7 | 70.0 | 76.7 | 50.0 | 70.0 | 83.3 | 63.3 | 83.3 |
| G-measure | 86.1 | 80.6 | 79.4 | 85.8 | 74.3 | 82.8 | 69.1 | 82.2 | 89.5 | 78.3 | 89.5 |
| **Subject 8** | | | | | | | | | | | |
| Accuracy | 91.7 | 71.7 | 68.3 | 83.3 | 86.7 | 86.3 | 71.7 | 91.7 | 88.3 | 83.3 | 91.7 |
| Sensitivity | 93.3 | 65.0 | 70.0 | 70.0 | 96.7 | 80.0 | 100.0 | 100.0 | 83.3 | 96.7 | 83.3 |
| Specificity | 90.0 | 73.9 | 69.6 | 96.7 | 76.7 | 96.7 | 43.3 | 83.3 | 100.0 | 70.0 | 100.0 |
| G-measure | 91.1 | 69.3 | 69.7 | 79.6 | 85.5 | 86.3 | 57.2 | 90.9 | 89.8 | 81.5 | 89.8 |
| **Subject 9** | | | | | | | | | | | |
| Accuracy | 76.7 | 80.0 | 60.0 | 80.0 | 75.0 | 80.0 | 68.3 | 81.7 | 83.3 | 70.0 | 83.3 |
| Sensitivity | 93.3 | 81.0 | 68.9 | 86.7 | 93.3 | 90.0 | 100.0 | 100.0 | 93.3 | 100.0 | 93.3 |
| Specificity | 60.0 | 80.0 | 57.6 | 73.3 | 56.7 | 70.0 | 36.7 | 63.3 | 73.3 | 40.0 | 73.3 |
| G-measure | 70.5 | 80.4 | 62.5 | 68.7 | 54.9 | 69.2 | 44.5 | 70.9 | 72.7 | 53.9 | 72.7 |
| **Subject 10** | | | | | | | | | | | |
| Accuracy | 85.0 | 66.7 | 60.0 | 78.3 | 60.0 | 76.7 | 61.7 | 76.7 | 80.0 | 83.3 | 83.3 |
| Sensitivity | 83.3 | 68.0 | 69.7 | 63.3 | 73.3 | 73.3 | 100.0 | 93.3 | 86.7 | 100.0 | 100.0 |
| Specificity | 86.7 | 72.4 | 59.6 | 93.3 | 46.7 | 80.0 | 23.3 | 60.0 | 93.3 | 66.7 | 66.7 |
| G-measure | 84.0 | 70.1 | 64.4 | 72.9 | 52.9 | 73.0 | 36.0 | 70.9 | 89.1 | 78.6 | 78.6 |
| **Average** | | | | | | | | | | | |
| Accuracy | 85.3 | 77.0 | 70.5 | 84.3 | 76.0 | 83.0 | 71.1 | 85.5 | 86.1 | 81.8 | 88.2 |
| Sensitivity | 87.5 | 77.6 | 72.7 | 83.0 | 85.0 | 86.9 | 98.0 | 91.6 | 89.0 | 96.1 | 90.5 |
| Specificity | 82.6 | 83.9 | 72.1 | 86.2 | 66.4 | 82.1 | 43.4 | 78.9 | 87.7 | 68.7 | 84.4 |
| G-measure | 83.1 | 80.4 | 72.1 | 81.5 | 70.2 | 81.9 | 57.0 | 82.3 | 86.1 | 78.0 | 84.7 |

observed that 3-WV (proposed) framework performs the best for almost all feature extraction approaches on subject- wise single-trial EEG data.

Further, the results obtained by proposed approach are compared with some state-of-the-art methods. All the approaches are applied on the EEG data recorded for this work. Ten subjects, each subjects' 16- channel EEG data, have been recorded by conducting a CIT. Analysis of variance (ANOVA) is the most commonly used statistical technique for CIT-based studies such as [5,6,13]. In this work, ANOVA has been applied on subject-wise EEG data for single trial and it has been observed that means of two groups or two classes overlaps. ERP data generally overlap [31] (as shown in Fig. 6); hence, statistical approaches like ANOVA are not sufficient to identify the human behavior.

Figure 6 shows that the mean values calculated using a two-way ANOVA and overlapping of means of two classes show certain similarity in data recorded for both classes. Authors like Wang et al. [10] have applied nonparametric weighted feature extraction technique based on LDA and KNN as classifier. Nonparametric feature extraction techniques are useful if some specific number of features are

needed to be extracted. It also reduces the effect of outliers present in data. We have applied the same approach on recorded EEG data for different subjects. After applying the same, accuracy of 76.8%, specificity of 70.0%, sensitivity as 73.1% and *G*-measure as 76.5% have been achieved. Arasteh et al. [8] used a machine learning approach known as empirical mode decomposition which divides signal into various intrinsic mode functions. EMD was used as feature extraction approach and LDA as classifier in their work. This framework has been applied on EEG data recorded, and results were obtained. An accuracy of 80.1%, specificity of 75.7%, sensitivity of 75.7% and *G*-measure of 77.8% have been obtained. Authors have also applied genetic algorithm for feature selection. As this work is focused on feature extraction and classification, feature selection will be performed as future work. In an other work [9], P300 components are separated from non-P300 components by ICA. In order to identify independent components, topographic template matching has been performed on extracted P300 components. SVM as classification approach has been used. In the same framework, we have applied on recorded EEG data and an accuracy of 60.17%, specificity of 51.33%, sensitivity of 67.83% and *G*-measure

**Table 6** Subject-wise performance using various classifiers and ensemble framework applied on wavelet coefficients

| Measures | LDA | NN | NB | SVM | KNN | 5-MV | 5-UV | 5-WV | 3-MV | 3-UV | 3-WV (proposed method) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Subject 1** | | | | | | | | | | | |
| Accuracy | 83.6 | 84.2 | 46.9 | 78.5 | 70.9 | 81.8 | 58.0 | 83.8 | 86.0 | 72.7 | 87.9 |
| Sensitivity | 93.3 | 91.7 | 48.1 | 76.7 | 86.7 | 83.3 | 96.7 | 83.3 | 90.0 | 100.0 | 90.0 |
| Specificity | 72.0 | 87.1 | 52.9 | 80.0 | 52.0 | 80.0 | 14.9 | 80.0 | 84.0 | 40.0 | 84.0 |
| G-measure | 80.0 | 88.4 | 49.6 | 73.3 | 62.3 | 76.6 | 33.6 | 76.6 | 85.4 | 54.2 | 85.4 |
| **Subject 2** | | | | | | | | | | | |
| Accuracy | 77.9 | 84.2 | 83.8 | 87.1 | 86.1 | 87.4 | 77.6 | 86.1 | 86.8 | 82.4 | 88.9 |
| Sensitivity | 75.6 | 82.3 | 86.5 | 80.9 | 85.8 | 89.8 | 92.0 | 92.0 | 87.6 | 92.0 | 87.6 |
| Specificity | 76.7 | 86.3 | 83.3 | 83.3 | 63.3 | 83.3 | 56.7 | 76.7 | 83.3 | 70.0 | 83.3 |
| G-measure | 72.3 | 84.1 | 84.8 | 81.4 | 64.1 | 85.3 | 62.0 | 80.2 | 84.5 | 76.9 | 84.5 |
| **Subject 3** | | | | | | | | | | | |
| Accuracy | 98.0 | 96.0 | 82.9 | 96.9 | 98.0 | 98.0 | 86.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Sensitivity | 73.3 | 81.4 | 70.8 | 76.7 | 70.0 | 80.0 | 96.7 | 93.3 | 76.7 | 90.0 | 76.7 |
| Specificity | 86.7 | 83.1 | 73.6 | 83.3 | 73.3 | 80.0 | 46.7 | 66.7 | 86.7 | 53.3 | 86.7 |
| G-measure | 78.5 | 82.2 | 72.1 | 79.5 | 70.7 | 79.6 | 65.7 | 77.9 | 81.1 | 67.3 | 81.1 |
| **Subject 4** | | | | | | | | | | | |
| Accuracy | 94.0 | 90.0 | 81.6 | 95.0 | 84.8 | 92.0 | 79.8 | 91.8 | 91.0 | 90.0 | 96.0 |
| Sensitivity | 100.0 | 100.0 | 81.0 | 93.8 | 96.0 | 96.0 | 100.0 | 96.0 | 100.0 | 100.0 | 100.0 |
| Specificity | 100.0 | 93.3 | 94.0 | 100.0 | 100.0 | 100.0 | 72.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| G-measure | 100.0 | 96.5 | 86.6 | 96.7 | 97.9 | 97.9 | 83.2 | 97.9 | 100.0 | 100.0 | 100.0 |
| **Subject 5** | | | | | | | | | | | |
| Accuracy | 94.0 | 90.0 | 81.6 | 95.0 | 84.8 | 92.0 | 79.8 | 91.8 | 91.0 | 90.0 | 96.0 |
| Sensitivity | 96.0 | 87.0 | 85.6 | 100.0 | 92.0 | 92.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Specificity | 92.0 | 96.0 | 81.5 | 92.0 | 71.6 | 92.0 | 59.6 | 75.6 | 92.0 | 88.0 | 92.0 |
| G-measure | 93.9 | 91.2 | 83.1 | 95.5 | 79.7 | 91.8 | 76.3 | 86.5 | 95.5 | 92.6 | 95.5 |
| **Subject 7** | | | | | | | | | | | |
| Accuracy | 90.0 | 86.7 | 78.3 | 88.3 | 78.3 | 86.7 | 81.7 | 88.3 | 91.7 | 85.0 | 92.9 |
| Sensitivity | 93.3 | 90.0 | 81.0 | 90.0 | 83.3 | 93.3 | 100.0 | 100.0 | 90.0 | 100.0 | 90.0 |
| Specificity | 86.7 | 90.5 | 77.7 | 86.7 | 73.3 | 80.0 | 63.3 | 76.7 | 93.3 | 70.0 | 93.3 |
| G-measure | 88.9 | 89.5 | 79.1 | 87.8 | 77.3 | 86.0 | 78.3 | 87.2 | 90.9 | 82.9 | 90.9 |
| **Subject 8** | | | | | | | | | | | |
| Accuracy | 97.3 | 78.3 | 66.7 | 95.0 | 88.3 | 97.3 | 80.0 | 90.0 | 98.3 | 93.3 | 98.3 |
| Sensitivity | 96.7 | 85.3 | 64.5 | 93.3 | 96.7 | 96.7 | 100.0 | 96.7 | 96.7 | 100.0 | 96.7 |
| Specificity | 100.0 | 79.4 | 68.7 | 96.7 | 80.0 | 100.0 | 60.0 | 83.3 | 100.0 | 86.7 | 100.0 |
| G-measure | 98.3 | 68.3 | 66.5 | 94.8 | 87.4 | 98.3 | 69.2 | 89.4 | 98.3 | 92.8 | 98.3 |
| **Subject 9** | | | | | | | | | | | |
| Accuracy | 75.0 | 85.0 | 56.7 | 80.0 | 71.7 | 80.0 | 73.3 | 81.7 | 80.0 | 75.0 | 80.0 |
| Sensitivity | 80.0 | 85.0 | 68.9 | 86.7 | 86.7 | 86.7 | 100.0 | 96.7 | 86.7 | 90.0 | 86.7 |
| Specificity | 70.0 | 84.5 | 52.6 | 73.3 | 56.7 | 73.3 | 46.7 | 66.7 | 73.3 | 60.0 | 73.3 |
| G-measure | 71.1 | 84.7 | 59.9 | 67.9 | 50.5 | 67.9 | 52.7 | 71.3 | 67.9 | 63.4 | 67.9 |
| **Subject 10** | | | | | | | | | | | |
| Accuracy | 90.0 | 80.0 | 60.0 | 85.0 | 53.3 | 83.3 | 65.0 | 86.7 | 90.2 | 81.7 | 91.7 |
| Sensitivity | 90.0 | 86.7 | 62.3 | 76.7 | 60.0 | 73.3 | 93.3 | 86.7 | 86.7 | 90.0 | 66.7 |
| Specificity | 90.0 | 82.5 | 59.7 | 93.3 | 46.7 | 93.3 | 36.7 | 86.7 | 96.7 | 73.3 | 96.7 |
| G-measure | 88.7 | 83.7 | 60.9 | 82.2 | 43.7 | 78.8 | 47.6 | 84.4 | 89.8 | 79.0 | 78.5 |

**Table 6** continued

| Measures | LDA | NN | NB | SVM | KNN | 5-MV | 5-UV | 5-WV | 3-MV | 3-UV | 3-WV (proposed method) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Average | | | | | | | | | | | |
| Accuracy | 88.8 | 86.0 | 71.0 | 89.0 | 79.5 | 88.7 | 75.7 | 88..9 | 90.5 | 85.5 | 92.4 |
| Sensitivity | 88.7 | 87.7 | 72.0 | 86.0 | 84.1 | 87.9 | 97.6 | 93.8 | 90.5 | 95.7 | 88.2 |
| Specificity | 86.0 | 87.0 | 71.5 | 87.6 | 68.5 | 86.8 | 50.7 | 79.1 | 89.9 | 71.2 | 89.9 |
| G-measure | 85.7 | 85.4 | 71.4 | 84.3 | 70.4 | 84.7 | 63.2 | 83.5 | 88.2 | 78.8 | 86.9 |
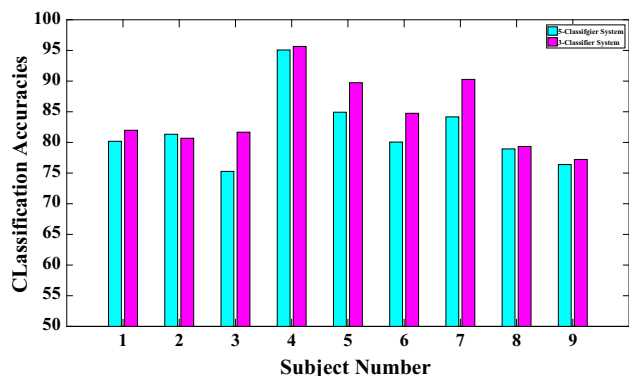


**Fig. 5** Comparison of the performance of 5-classifier ensemble framework and 3-classifier ensemble framework



**Fig. 6** Comparison of the means of two classes using ANOVA

of 34.27% have been obtained. All the comparative results are tabulated in Table 7. From the results, it is observed that proposed approach gives better performance in terms of accuracy, sensitivity, specificity and *G*-measure. The approach attained an average overall accuracy of 84.7%, specificity of 83.9%, sensitivity of 82.5% and *G*- measure of 80.8%. From results, it can be inferred that among various feature extraction approaches, wavelets providing time–frequency information of signals gives the best performance. In comparison with applying base classifiers for classification, ensemble technique provides better results by combining the best-performing classifiers. The selection of classifier is tedious task as it depends on the type of dataset applied [22]. Therefore, with the knowledge of dataset and using ensemble approach better performance results can be obtained.

## 5 Conclusion

In this paper, an ensemble framework has been proposed by aggregating three best-performing classifiers. A classifier ranking approach has been applied to select three classifiers among five classifiers, viz. LDA, SVM, MLFFNN, KNN and NB. The main aim is developing the ensemble framework to provide a better approach for classification of guilty and innocent subjects. The proposed framework is applied on EEG data recorded for a Concealed Information Test to analyze the human behavior while lying. During data acquisition, set of images are flashed in front of subjects in two different sessions. Signals acquired during the experimental sessions are analyzed to identify the guilty or innocent subject. A wide range of available feature extraction techniques has been applied on acquired EEG data. For classification of EEG data, five classifiers are applied and their results are aggregated using majority voting, weighted voting and unanimous voting. To improve the performance of ensemble framework, classifier ranking is applied. The best three performing classifiers are aggregated using weighted voting approach for developing the ensemble framework. The proposed ensemble framework (3-WV) results are compared with different base classifiers' performance and with 5-classifier system. Among the various feature extraction approaches, applied results show that wavelet performs best. Using wavelet with proposed ensemble framework, subject 3 has attained a highest of 100% accuracy and 92.4% average accuracy. Results of proposed framework are compared with some existing approaches. An improved overall classification accuracy has been achieved using proposed ensemble framework (3-WV). Further in future, different time–frequency domain feature extraction techniques can be applied to extract useful information from EEG signals. Feature selection approaches can be applied in future, so as to feed the best set of features to the classifiers. Also, optimization using bio-inspired approaches

**Table 7** Comparison with existing approaches

| Classification approach | Feature extraction Technique | EEG dataset | Number of channels | Accuracy | Specificity | Sensitivity | G-measure |
|---|---|---|---|---|---|---|---|
| KNN [10] | Nonparametric LDA | Current study dataset | 16 channels | 76.8 | 70.0 | 73.1 | 76.5 |
| QDA [20] | Various (power, wavelet, Hjorth parameters, etc.) | Emotion dataset [20] | 64 channels | 35.9 | – | – | – |
| LDA [8] | EMD | Current study dataset | 16 channels | 80.1 | 75.7 | 75.7 | 77.8 |
| SVM [9] | ICA | Current study dataset | 16 channels | 60.1 | 51.3 | 67.8 | 34.2 |
| Ensemble classification-3-WV (proposed) | Various (amplitude, complexity, mobility frequency, power, wavelet) | Current study dataset | 16 channels | 84.7 | 83.9 | 82.5 | 80.8 |

like PSO can be applied to reach the optimal solution with faster rate and better accuracy.

# References

1. Ramadan, R.A., Vasilakos, A.V.: Brain computer interface: control signals review. Neurocomputing **223**, 26–44 (2017)
2. Abootalebi, V., Moradi, M.H., Khalilzadeh, M.A.: A new approach for EEG feature extraction in p300-based lie detection. Comput. Methods Progr. Biomed. **94**(1), 48–57 (2009)
3. Farwell, L.A., Donchin, E.: The truth will out: interrogative polygraphy ("lie detection") with event-related brain potentials. Psychophysiology **28**(5), 531–547 (1991)
4. Rosenfeld, J.P., Soskins, M., Bosh, G., Ryan, A.: Simple, effective countermeasures to p300-based tests of detection of concealed information. Psychophysiology **41**(2), 205–219 (2004)
5. Rosenfeld, J.P., Labkovsky, E., Winograd, M., Lui, M.A., Vandenboom, C., Chedid, E.: The Complex Trial Protocol (CTP): a new, countermeasure-resistant, accurate, p300-based method for detection of concealed information. Psychophysiology **45**(6), 906–919 (2008)
6. Kubo, Kenta, Nittono, Hiroshi: The role of intention to conceal in the p300-based concealed information test. Appl. Psychophysiol. Biofeedback **34**(3), 227–235 (2009)
7. Meixner, J.B., Rosenfeld, J.P.: A mock terrorism application of the p300-based concealed information test. Psychophysiology **48**(2), 149–154 (2011)
8. Arasteh, A., Moradi, M.H., Janghorbani, A.: A novel method based on empirical mode decomposition for p300-based detection of deception. IEEE Trans. Inf. Forensics Secur. **11**(11), 2584–2593 (2016)
9. Gao, Junfeng, Liang, Lu, Yang, Yong, Gang, Yu., Na, Liantao, Rao, NiNi: A novel concealed information test method based on independent component analysis and support vector machine. Clin. EEG Neurosci. **43**(1), 54–63 (2012)
10. Wang, Deng, Miao, Duoqian, Blohm, Gunnar: A new method for EEG-based concealed information test. IEEE Trans. Inf. Forensics Security **8**(3), 520–527 (2013)
11. Akhavan, A., Moradi, M.H., Vand, S.R.: Subject-based discriminative sparse representation model for detection of concealed information. Comput. Methods Progr. Biomed. **143**, 25–33 (2017)
12. Farahani, E.D., Moradi, M.H.: Multimodal detection of concealed information using genetic-SVM classifier with strict validation structure. Inform. Med. Unlocked **9**, 58–67 (2017)
13. Lukács, Gáspár, Gula, Bartosz, Szegedi-Hallgató, Emese, Csifcsák, Gábor: Association-based concealed information test: a novel reaction time-based deception detection method. J. Appl. Res. Mem. Cognit. **6**(3), 283–294 (2017)
14. Breiman, Leo: Bagging predictors. Mach. Learn. **24**(2), 123–140 (1996)
15. Freund, Y., Schapire, R.E., et al.: Experiments with a new boosting algorithm. In: ICML, vol. 96, pp. 148–156. Bari, Italy (1996)
16. Opitz, D.W., Maclin, R.: Popular ensemble methods: an empirical study. J. Artif. Intell. Res. (JAIR) **11**, 169–198 (1999)
17. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. **20**(8), 832–844 (1998)
18. easycap. http://www.easycap.de/e/products/products.htm15 (2017)
19. Brain products. http://www.brainproducts.com/ (2017)
20. Jenke, Robert, Peer, Angelika, Buss, Martin: Feature extraction and selection for emotion recognition from EEG. IEEE Trans. Affect. Comput. **5**(3), 327–339 (2014)
21. Hjorth, Bo: EEG analysis based on time domain properties. Electroencephalogr. Clin. Neurophysiol. **29**(3), 306–310 (1970)
22. Edla, D.R., Tripathi, D., Cheruku, R., Kuppili, V.: An efficient multi-layer ensemble framework with BPSOGSA-based feature selection for credit scoring data analysis. Arab. J. Sci. Eng. 1–20 (2017). https://doi.org/10.1007/s13369-017-2905-4
23. Abellán, J., Castellano, J.G.: A comparative study on base classifiers in ensemble methods for credit scoring. Expert Syst. Appl. **73**, 1–10 (2017)
24. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Ann. Eugen. **7**(2), 179–188 (1936)
25. Cortes, Corinna, Vapnik, Vladimir: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
26. Hornik, Kurt, Stinchcombe, Maxwell, White, Halbert: Multilayer feedforward networks are universal approximators. Neural Netw. **2**(5), 359–366 (1989)
27. Svozil, Daniel, Kvasnicka, Vladimir, Pospichal, Jiri: Introduction to multi-layer feed-forward neural networks. Chemom. Intell. Lab. Syst. **39**(1), 43–62 (1997)
28. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. Am. Stat. **46**(3), 175–185 (1992)
29. Russell, S., Norvig, P.: Artificial Intelligence. A Modern Approach, pp. 25–27. Prentice-Hall, Englewood Cliffs (1995)
30. Tripathi, D., Edla, D.R., Cheruku, R.: Hybrid credit scoring model using neighborhood rough set and multi-layer ensemble classification. J. Intell. Fuzzy Syst. **34**(3), 1543–1549 (2018)
31. Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M.J., Iragui, V., Sejnowski, T.J.: Removing electroencephalographic artifacts by blind source separation. Psychophysiology **37**(2), 163–178 (2000)

**Annushree Bablani** is presently working as full-time research scholar in Department of Computer Science and Engineering at National Institute of Technology Goa, India. She received M.Tech. degree in computer science and engineering (software engineering) from Government Engineering College, Ajmer, in 2016, B.Tech. degree in computer science and engineering from Rajasthan College of Engineering, Jaipur, in 2013. Her current research area includes brain-computer interface, soft computing and artificial intelligence.

**Damodar Reddy Edla** is an assistant professor in Department of Computer Science and Engineering, National Institute of Technology Goa. He received M.Sc. degree from University of Hyderabad in 2006, M.Tech. and Ph. D. degree in computer science and engineering from Indian School of Mines, Dhanbad, in 2009 and 2013, respectively. His research interests are data mining and wireless sensor networks. He has published more than 50 research papers in reputed journals and international conferences. He is senior member of IACSIT. He is also editorial board member of several International journals.

**Diwakar Tripathi** received BE degree in computer science and engineering from Institution of Electronics and Telecommunication Engineering, New Delhi, India, in 2009, ME degree in computer engineering (software engineering) from the Institute of Engineering and Technology, Devi Ahilya Vishwavidyalaya, Indore, India, in 2014 and Ph.D. (thesis submitted) from National Institute of Technology Goa, India. He is currently working as assistant professor at Madanapalle Institute of Technology and Science, Madanapalle, A.P., India. His current research includes data mining, soft computing and artificial intelligence.

**Venkatanareshbabu Kuppili** is an assistant professor in Department of Computer Science and Engineering, National Institute of Technology Goa, India. He was with Evalueserve Pvt. Ltd, as a Senior Research Associate. He received M.Tech. and Ph.D. degree in computer science and engineering from Indian Institute of Technology Delhi. He has authored a number of research papers published in reputed international journals in the area of neural networks, classification, and clustering.