

A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD)

Wajid Mumtaz¹  · Syed Saad Azhar Ali¹ · Mohd Azhar Mohd Yasin² · Aamir Saeed Malik¹

Received: 30 January 2017 / Accepted: 3 July 2017
© International Federation for Medical and Biological Engineering 2017

Abstract Major depressive disorder (MDD), a debilitating mental illness, could cause functional disabilities and could become a social problem. An accurate and early diagnosis for depression could become challenging. This paper proposed a machine learning framework involving EEG-derived synchronization likelihood (SL) features as input data for automatic diagnosis of MDD. It was hypothesized that EEG-based SL features could discriminate MDD patients and healthy controls with an acceptable accuracy better than measures such as interhemispheric coherence and mutual information. In this work, classification models such as support vector machine (SVM), logistic regression (LR) and Naïve Bayesian (NB) were employed to model relationship between the EEG features and the study groups (MDD patient and healthy controls) and ultimately achieved discrimination of study participants. The results indicated that the classification rates were better than chance. More specifically, the study resulted into SVM classification accuracy = 98%, sensitivity = 99.9%, specificity = 95% and *f*-measure = 0.97; LR classification accuracy = 91.7%, sensitivity = 86.66%, specificity = 96.6% and *f*-measure = 0.90; NB classification accuracy = 93.6%, sensitivity = 100%, specificity = 87.9% and *f*-measure = 0.95. In conclusion, SL could be a promising method for diagnosing depression. The findings could be

generalized to develop a robust CAD-based tool that may help for clinical purposes.

Keywords Major depressive disorder · Machine learning-based diagnosis · Computer-aided diagnosis (CAD) for depression · EEG-based machine learning techniques for depression · Synchronization likelihood (SL) features

1 Introduction

Major depressive disorder (MDD), commonly termed as depression, is a mental illness. It has been considered as the most prevalent disease among elderly patients in the USA [16, 54]. The symptoms for depression persist for more than 2 weeks rendering the patients functionally disabled. Identifying depression at an early stage could prevent its further development to more severe conditions such as mild to severe depression [55]. In addition, an early diagnosis may further help during treatment process. Unfortunately, the early diagnosis has been a challenge due to its comorbid nature [13]. Conventionally, the MDD patients must meet the diagnostic criteria defined in the Diagnostic and Statistical Manual for Mental Disorders V (DSM-V) [2]. The subjectivity in diagnosing MDD patients may not be completely avoided due to human factors such as missing patient's family history or exaggerated reports on patient's symptoms. However, the subjectivity could be reduced by incorporating scientific evidences generated by neuroscience data such as electroencephalogram (EEG).

Recently, EEG-based machine learning (ML) techniques have received considerable attention due to their capability to mine noninvasive neuroimaging data to establish the computer-aided diagnosis (CAD) solutions to facilitate the diagnosis of depression [9, 37, 39, 41]. In addition, mining

All authors have equal contribution for this research work.

✉ Aamir Saeed Malik
aamir_saeed@utp.edu.my

¹ Center for Intelligent Signal and Imaging Research, Electrical and Electronic Engineering Department, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Malaysia

² Department of Psychiatry, Hospital Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia

functional magnetic resonance imaging (fMRI) data with ML methods has shown promising research results [32, 56]. Specifically, the support vector machine (SVM) has been emphasized as a method of choice for the diagnosis of depression [44]. On the other hand, the automated EEG-based ML methods are proved feasible to discriminate the depressed patients from healthy controls [5, 6, 21, 27, 36]. In addition to depression, the classification algorithms are found useful for neurological diseases such as schizophrenia and Alzheimer's disease [25]. In the context of depression, classifier such as artificial neural networks (ANN) is trained to classify the MDD patients and the healthy controls [14, 46]. Recently, a depression diagnostic index is proposed based on nonlinear features extracted from EEG data [6]. The ML techniques have shown promising research results, yet their clinical translation has been less clear. Hence, there is a need to further replicate the finding with newer methods that could achieve high efficiencies and could be considered feasible for clinical applications.

The MDD patients have demonstrated abnormal functional connectivity among different scalp locations in the frontal and temporal regions when compared with healthy controls [42]. The literature has shown that the EEG-based assessment of functional connectivity between different brain regions could be promising marker for depression [10, 11, 15, 18]. For example, the aberrant neuronal interactions between different brain regions have been associated with MDD patients when compared with healthy controls. In addition, the synchronization likelihood (SL) was used to diagnose Alzheimer's disease [50].

In the literature, a group of researchers has tested the feasibility of phase synchronization for diagnosing depression with 12 MDD and 16 healthy controls [45]. The study concluded that the patients exhibited significant deficits in information transmission among cortical regions during rest conditions, and that these deficits could be estimated by the SL measure of the EEGs recorded from cerebral cortex. The researchers have focused on the investigation of SL and its association with MDD but could not employ the automated selection of useful patterns involving ML concepts. On the contrary, the current work is an improvement based on the same concept of utilizing SL in combination with three different classification models. The proposed scheme could become a stepping stone for the development of a CAD tool suitable for clinical applications. Hence, the current study advances the current knowledge to a next level by proposing the novel methodology. Therefore, in this study, the quantification of functional connectivity was performed by computing SL values between different sensor pairs at sparse locations [49].

In this study, the SL features are used as input data to the proposed ML method to automate the discrimination of the MDD patients and healthy controls. For the sake of

comparison, the proposed method is compared with the interhemispheric coherence and mutual information. In this paper, classification accuracy, sensitivity, specificity and the *f*-measures are computed as classification performance metrics involving logistic regression (LR), support vector machine (SVM) and Naïve Bayesian (NB) classifiers.

2 Materials and methods

2.1 Study participants

In this study, the experimental data were acquired from 34 MDD patients (females = 18) (age, mean = 40.33, SD ± 12.861) and an age-matched group of 30 healthy subjects (females = 9) (age, mean = 38.227, SD ± 15.64). The MDD patients, who met the diagnostic criteria for MDD according to Diagnostic and Statistical Manual-IV (DSM-IV) [1], were recruited from outpatient clinic of Hospital Universiti Sains Malaysia (HUSM). The MDD patients have passed the diagnostic criteria for unipolar depression without any psychotic symptoms. Both the MDD patients and healthy participants were volunteers and have signed the consent form of participation. The ethics committee of the hospital has approved the experiment design that was briefed to the participants. The healthy participants were examined for clinical symptoms to exclude the possibility of any physical and mental disability and were found normal. The participants were instructed to abstain from coffee, nicotine, alcohol before EEG recordings. To maintain consistency of the data, all participants have performed the experiment data acquisition at same time of the day.

2.2 Experimental data acquisition

The experimental data acquisition included clinical scores and EEG data. The clinical data involved questionnaires for the assessment of disease severity, such as Beck Depression Inventory-II (BDI-II) and Hospital Anxiety and Depression Scale (HADS) [6, 33]. The EEG data acquisition included 5 min of EEG data recordings during eyes closed (EC) and 5 min of EEG recordings during eyes open (EO) conditions. The EEG data acquisition involved 19-channel EEG cap sensors placed according to the 10–20 electrode placement standard [23] with linked ear (LE) as reference. For the sake of EEG analysis, the EEG data were re-referenced to infinity reference (IR) as suggested by [47]. The 19 electrodes covering the scalp include frontal (Fp1, Fp2, F3, F4, F7, F8, Fpz), temporal (T3, T4, T5, T6), parietal (P3, P4, P7, P8), occipital (O1, O2), central (C3, C4). The 19-channel EEG cap was attached with an amplifier from Brain Master Systems. The hardware configuration was set described here. The 0.5–70-Hz filter, with a 50-Hz notch

filter, and a sample rate of 256 per second were used to discretize the EEG data before storage to a computer disk for later analysis such as preprocessing and EEG analysis.

2.3 EEG noise reduction

The EEG data were confounded with noise or artifacts such as eye blinks, movements and muscular activities, e.g., heart beats. The data confounded with artifacts may not truly represent the underlying brain activities; therefore, reduction of noise from the recorded EEG data was desirable and termed as EEG data preprocessing. In this paper, the noise from the EEG data reduced multiple source modeling technique [12], implemented in the standard brain electric source analysis (BESA) software [19]. According to the technique, the noise topographies and a head model were utilized to effectively clean the EEG noise. For example, to remove eye blinks, the noise vectors for eye blinks were constructed from the recorded EEG data. Based on these noise vectors, eye-blink topographies were constructed. These noise topographies and a head model, already provided in BESA, were then utilized to successfully clean the EEG data for eye blinks. In addition, noise topographies for the head movement, heart or muscular activities were constructed. Finally, a head model and artifacts topographies due to muscular and heart artifacts were utilized to clean head movement, heart or muscular activities from the data. The construction of noise topographies including eye blinks, muscular and heart activities was based on the recorded EEG data and then can be applied to whole recording. This method preserves the information in the data as it allows the EEG data affected by artifacts to be utilized in the data analysis. In comparison, the methods based on artifact rejection also delete the EEG data which may have result into loss of information.

2.4 Proposed machine learning (ML) method

Figure 1 shows an overview of the proposed ML method involving EEG feature extraction, selection, classification and validation. As detailed in previous subsection, the EEG artifact reduction implicated into clean EEG data which was followed by EEG feature extraction. For feature extraction, the artifact-free EEG data were segmented into 1-min epochs per study subject. As a result, a large number (N_c) of candidate features were extracted and arranged column-wise in a matrix, termed as the *EEG data matrix*. Columns of the matrix represented features, and each column was denoted as x_i , where $i = 1 \dots N_c$. Rows of the matrix represented the study participants (MDD patients and healthy controls) with 2 physiological conditions per participants, denoted by $L = [(x_i, y_i), i = 1 \dots N_c]$ including both the features and the corresponding output class labels or targets,

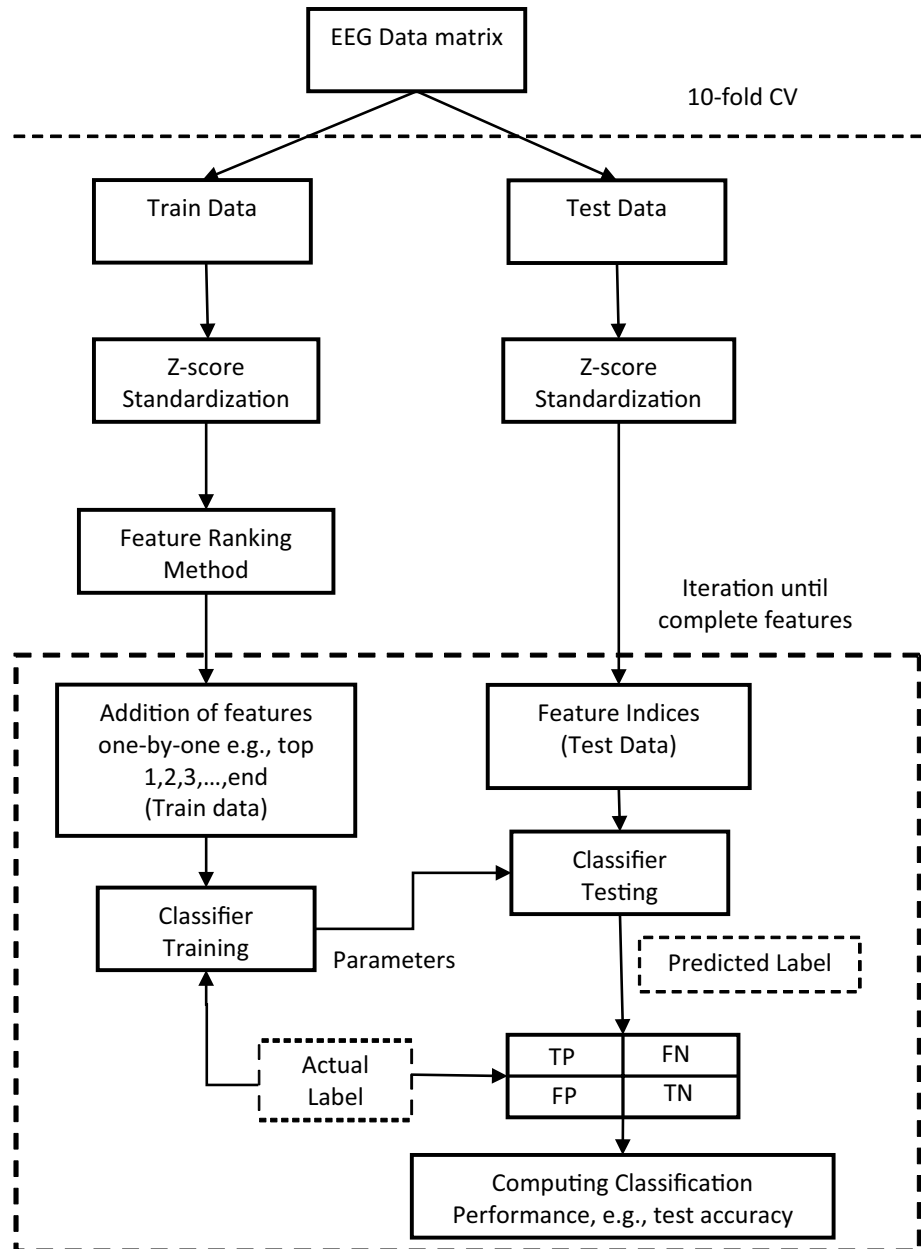
i.e., $y = [MDD \text{ or } Controls]$. A detailed description for each subprocess is provided in the respective subsections.

In this study, the feature selection was based on a rank-based feature selection method [34]. In brief, each feature in the *EEG data matrix* was assigned a weight and rank-ordered accordingly. The weights were computed according to the area under curve (AUC) of each feature, termed as z -values. Next, the features were rank-ordered such as the feature with highest z -value was placed at the top of the list and other features were arranged in descending order of their respective z -values. A detailed description on the feature selection method is provided in the respective subsection.

The classifier training and testing was performed after rank-ordering the features. For computing the classifier test accuracies as a function of number of features, each feature was progressively added into a reduced set of features, e.g., starting with the highest z -value feature among the list of rank-ordered features. Upon addition of each feature, the size of reduced set of feature matrix kept increasing. For example, during first iteration, the matrix size was 64×1 (where 64 corresponded to number of samples and the 1 corresponded to the feature itself). For this matrix, classifier performance was computed. During second iteration, adding another feature into the reduced set of feature matrix implicated into a matrix size of 64×2 , and the classifier performance was computed. The process kept repeating until all features were utilized, i.e., 64×513 . As a result, the classifier performance was plotted as a function of the number of features. The feature set that showed best accuracy among others was selected and considered as selected features. Since the process is repeated for each fold of 10-CV, the final values were computed by taking average across each fold. As a result, all features set were tested accordingly. To compute the classifier accuracy, sensitivity, specificity and f -measure, a confusion matrix was constructed involving the classifier predictions and the actual labels. In addition, the classifier accuracies were compared based on the ROC plots as well. In the rest of this section, a detailed description on feature extraction, selection, classification and validation is provided.

2.4.1 EEG feature extraction: synchronization likelihood (SL)

The SL is a measure to quantify synchronization of 2 times series data [49]. Because of the quantification the SL value may have values between 0 and 1: A minimum value such as zero represents complete nonsynchronization. On the other hand, a maximum value, i.e., one, corresponds to the completely synchronized EEG data at the specified brain location. For example, the EEG data recorded during the eyes closed (EC) rest state and eye open (EO) rest state are found

Fig. 1 The proposed ML framework

to have small synchronization values. On the other hand, the SL resulted into a high value near 1 during the EEG data recorded during epileptic seizures. In short, the SL is a good measure to quantify the EEG data to discriminate the mental states.

In this study, the recorded EEG data consisted of multiple channels which were denoted as M simultaneously recorded time series $x_{k,i}$, where k denoted the channel number and ($k = 1, 2, 3, \dots, M$) and i denoted discrete time ($i = 1, 2, 3, \dots, N$). As described in Eq. (1), the EEG data corresponding to a channel were used to construct embedding vectors $X_{k,i}$ by using the time-delay embedding [51]:

$$X_{k,i} = (x_{k,i}, x_{k,i+l}, x_{k,i+2l}, \dots, x_{k,i+(m-1)l}) \quad (1)$$

where l is lag and m is the embedding dimension. For each channel k and each time i , a probability $P_{k,i}^\epsilon$ was computed such that embedding vector is closer to each other than a distance ϵ as described in Eq. (2):

$$P_{k,i}^\epsilon = \frac{1}{2(\omega_2 - \omega_1)} \sum_{\substack{j=1 \\ \omega_1 < |i-j| < \omega_2}}^N \theta(\epsilon - |X_{k,i} - X_{k,j}|) \quad (2)$$

Here the $||$ is Euclidean distance and θ is the Heaviside step function, $\theta(x) = 0$ if $x \leq 0$ and $\theta(x) = 1$ for $x > 0$. Here ω_1 and ω_2 are two windows; ω_1 is Theiler correction for autocorrelation effects and should be at least of the order

of autocorrelation time [52]; ω_2 is a window that sharpens the time resolution of the synchronization measure and is chosen such that $\omega_1 \ll \omega_2 \ll N$.

Now for each k and each i the critical distance $\epsilon_{k,i}$ is determined for which $P_{k,i}^e = p_{\text{ref}}$, where $p_{\text{ref}} \ll 1$. In this study, the value of the critical distance $\epsilon_{k,i}$ was found via empirical process such as it should fulfill the requirement $P_{k,i}^e = p_{\text{ref}}$.

Now for each time pair (i,j) and within the considered window ($\omega_1 < |i - j| < \omega_2$), it was easy to determine the number of channels $H_{i,j}$ where the embedding vectors $X_{k,i}$ and $X_{k,j}$ will be closer together than the critical distance $\epsilon_{k,i}$:

$$H_{i,j} = \sum_{k=1}^M \theta(\epsilon_{k,i} - |X_{k,i} - X_{k,j}|) \quad (3)$$

This number lies in a range between 0 and M and reflects how many of the embedded signals ‘resemble’ each other. Now, for each channel k and each discrete time pair (i,j) , the synchronization likelihood is defined as

$$\begin{aligned} \text{if } |X_{k,i} - X_{k,j}| < \epsilon_{k,i} : S_{k,i,j} &= \frac{H_{i,j}-1}{M-1} \\ \text{if } |X_{k,i} - X_{k,j}| \geq \epsilon_{k,i} : S_{k,i,j} &= 0 \end{aligned}$$

As mentioned in Eq. (3), by averaging over all j synchronization likelihood $S_{k,i}$ is obtained as follows:

$$S_{k,i} = \frac{1}{2(\omega_2 - \omega_1)} \sum_{\substack{j=1 \\ \omega_1 < |j-i| < \omega_2}}^N S_{k,i,j} \quad (4)$$

As described in Eq. (4), the synchronization likelihood $S_{k,i}$ is a measure which describes how strongly channel k at time i is synchronized to all the other $M-1$ channels. The SL takes on values between P_{ref} and 1. $S_{k,i} = P_{\text{ref}}$ corresponds to the case where all M channels are uncorrelated, and $S_{k,i} = 1$ corresponds to maximal synchronization to all M channels. The value of P_{ref} should be set to low value, and it did not depend on the properties of the time series or embedding parameters.

In this study, the SL was computed for each channel pair involving frontal (Fp1, Fp2, F3, F4, F7, F8, Fpz), temporal (T3, T4, T5, T6), parietal (P3, P4, P7, P8), occipital (O1, O2) and central (C3, C4). For example, the synchronization likelihood computed for Fp1 included channel pairs such as Fp1-Fp2; Fp1-F4; Fp1-F8; Fp1-T4; Fp1-T6; Fp1-P4; Fp1-P8; Fp1-O2; Fp1-C4. The extracted features were arranged column-wise in a matrix, termed as the EEG data matrix. Furthermore, the EEG data matrix was subjected to feature selection because the features could be redundant or irrelevant.

2.4.2 EEG data matrix and z-score standardization

The features extraction has resulted into *EEG data matrix* involving the number of rows (*data points* = 63). The matrix might not be centered and unequally distributed. Hence, the data standardization was performed involving z-score standardization. The standardization was performed by computing values column-wise by subtracting each element value with its column-wise mean and divided by the corresponding standard deviation.

A true assessment of any classification model requires independence between the test and train samples. In this study, the independence was maintained during EEG features extraction, standardization and classification. More specifically, the z-scores were computed for each feature involving the feature mean and standard deviation. Therefore, separate computation of z-score did not affect the performance of the proposed ML scheme. Moreover, the separate treatment of the test and train samples during each step implicated into test accuracy, sensitivity, specificity and f-measure. Therefore, in this study, the z-scores were computed separately for test and train samples.

2.4.3 Rank-based feature selection method

Feature selection remains as challenging research topic and carries critical importance during data analysis. The extracted features might be either redundant or irrelevant. From the classification point of view, high-dimensional datasets may easily over-fit or under-fit a classification model. Therefore, feature selection was desirable to reduce the feature space, i.e., from N_c to a lower dimension N_r .

In this study, the feature selection involved a rank-based feature selection method based on receiver operating curve (ROC) [31, 34]. According to the method, the area under curve (AUC) for each feature was computed termed as z-value. The z-value represents relevance of a feature with the class labels. A high z-value (equal or near 0.5) corresponded to the ability of a feature to discriminate within classes. The z-value could take any value between 0 and 0.5 indicating bad to good classification ability, accordingly. In this study, the features were ranked in descending order according to the z-values such as the most noteworthy features were top-listed. To find minimum number of features that would be sufficient to train the classifier model without over-fitting, an empirical process was adopted. The minimum number of features was determined based on iteratively observing performance of the classification models for each feature subsets selected from top-listed features such as 1st, 2nd, 3rd, 4th, 5th until complete features in the *EEG data matrix*. The feature set that showed best accuracy among others was selected and considered as selected features.

Table 1 Pseudocode for feature ranking method

```

patterns = [x y];
patterns = sortrows(patterns,-1);
y = patterns(:,2);
p = cumsum(y==1);
tp = p/sum(y==1);
n = cumsum(y== -1);
fp = n/sum(y== -1);
n = length(tp);
Y = (tp(2:n) + tp(1:n - 1))/2;
X = fp(2:n) - fp(1:n - 1);
auc = sum(Y.*X) - 0.5;

```

Table 2 Sample EEG data

Sample ID	...	i	j	...	Label
1		-0.2	+0.5		(-)
2		-1.4	-1.4		(-)
3		+0.8	-0.9		(-)
4		-0.8	+0.2		(+)
5	...	+0.1	-2.5	...	(+)
6		+0.5	+1.4		(-)
7		+1.6	-0.3		(+)
8		-2.1	-1.2		(-)
9		-0.3	+2.2		(+)
10		+3.4	-1.7		(-)

Table 3 Intermediate variable values

Feature values (sorted in descending order)	Labels	p	n	tp	fp
3.4	(-)	0	1	0	0.1667
1.6	(+)	1	1	0.25	0.1667
0.8	(-)	1	2	0.25	0.333
0.5	(-)	1	3	0.25	0.5
0.1	(+)	2	3	0.5	0.5
-0.2	(-)	2	4	0.5	0.6667
-0.3	(+)	3	4	0.75	0.6667
-0.8	(+)	4	4	1	0.6667
-1.4	(-)	4	5	1	0.8333
-2.1	(-)	4	6	1	1

Table 4 Computation of $Y = (tp(2:n) + tp(1:n - 1))/2$

$tp(2:n)$	0.25	0.25	0.25	0.5	0.5	0.75	1	1	1
$tp(1:n - 1)$	0	0.25	0.25	0.25	0.5	0.5	0.75	1	1
$(tp(2:n) + tp(1:n - 1))/2$	0.25	0.5	0.5	0.75	1	1.25	1.75	2	2
Y	0.125	0.25	0.25	0.375	0.5	0.625	0.875	1	1

To generate a sufficient statistical distribution of classifier performance metrics such as the accuracy, sensitivity and specificity for each subgroup, 100 times simulations were performed and box plots were plotted.

Table 1 provides pseudocode for the rank-based feature selection method. Let x be a vector that represents a feature, and the vector y represents the target labels (-1, +1). In this study, both x and y have the same dimensions.

Table 2 enlists the sample EEG data to further explain the pseudocode for the rank-based feature selection method. The table listed the sample data for 10 examples as shown in the first column. In addition, the columns 'i' and 'j' represent 2 unique features, accordingly. The last column shows the corresponding class labels.

Tables 3, 4 and 5 list the intermediate values of different variables during the computation of the AUC for the feature 'i' (as listed in Table 2). The computations follow the pseudocode provided in Table 1. As shown in Table 3, the first step is to sort the feature values in a descending order (1st column) and the corresponding labels are also adjusted (2nd column), accordingly. Further, the values of intermediate variables (i.e., p , n , tp and fp) are computed and listed in the respective column.

Table 4 provides the computation of the intermediate variable Y according to the formula $Y = (tp(2:n) + tp(1:n - 1))/2$. Table 4 shows the computed Y based on the values provided in Table 3.

Similarly, Table 5 provides the computation of the intermediate variable X according to the formula ' $X = fp(2:n) - fp(1:n - 1)$ '. Table 5 shows the computed X based on the values obtained in Table 3.

Finally, the AUC is computed based on the formula ' $auc = sum(Y.*X) - 0.5$ '. Table 6 shows the detailed values of intermediate variables Y and X and the AUC, respectively.

As shown in Table 6, the value obtained for AUC (z-value) is zero which means that the feature 'i' would not be a good option for further classification process and could be rejected. The process is repeated for all features in the EEG data matrix.

Table 5 Computation of $X = (fp(2:n) - fp(1:n - 1))$

fp(2:n)	0.1667	0.333	0.5	0.5	0.6667	0.6667	0.6667	0.8333	1
fp(1:n - 1)	0.1667	0.1667	0.333	0.5	0.5	0.6667	0.6667	0.667	0.8333
X	0	0.1667	0.1667	0	0.1667	0	0	0.1667	0.1667

Table 6 Computation of $AUC = \text{sum}(Y.*X) - 0.5$

$AUC = \text{sum}(Y.*X) - 0.5;$

$AUC = (0.125 \times 0 + 0.25 \times 0.1667 + 0.25 \times 0.1667 + 0.375 \times 0 + 0.5 \times 0.1667 + 0.625 \times 0 + 0.825 \times 0 + 0.1667 \times 1 + 0.1667 \times 1)$
 $- 0.5$

$AUC = 0.5 - 0.5 = 0$

2.4.4 Classification models

Classification has been an important procedure during automatic identification of patterns specific to a group or disease condition. In this study, three different classifiers were selected named as logistic regression (LR) classifier, support vector machine (SVM) and Naïve Bayesian (NB) classifier. The selection of LR and NB classifier was because they could easily be trained with small sample data. In addition, the problem at hand was to discriminate between two conditions: MDD versus Controls which is a dichotomous study and suited for classifier such as LR. On the other hand, SVM has been considered as a high-performance classifier. Since it has been considered as a benchmark, its sole purpose was to provide comparison with LR and NB. SVM is also good for low sample size data because it involves support vectors. Once the support vectors could be identified, SVM can perform better for small datasets as well.

In this study, the LR classifier was used to model the relationship between the reduced set of features and the corresponding treatment outcomes (MDD patients and controls) $y = [MDD, Controls]$, according to Eq. (5) [20]. For LR classifier, the coefficients estimation was based on maximum likelihood method. The LR classifier resulted into a likelihood value $l(x)$, where $0 < l(x) \leq 1$, which was an indication of subjects, associated either with MDD patients or controls. If $l(x)$ was greater than the threshold = 0.5, the subject was declared as MDD patients and otherwise associated with control group.

$$F(z) = E(Y/x) = \frac{1}{1 + e^{-z}} \quad (5)$$

where Y indicates the class labels and assigned a value either 'MDD' or 'Controls.' In addition, x represents a combination of different EEG features, i.e., the wavelet coefficients achieved after applying WT technique, power computations and interhemispheric asymmetry. To obtain the LR model from the logistic function, we used Eq. (6):

$$z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (6)$$

where z is a linear combination of α plus β_1 multiplied with X_1 , plus β_2 multiplied with X_2 and plus β_k multiplied with X_k , where the X_k are the independent variables and α and β_i are constant terms representing unknown parameters. Furthermore, by replacing the value of z from Eqs. (6) to (5), following Eq. (7) represents the logistic function:

$$F(z) = E(Y/x) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i X_i)}} \quad (7)$$

In terms of response and nonresponse, the risk of a person to be nonresponder or a responder is estimated and represented by Y or $l(x)$. The LR classifier resulted into a likelihood value $l(x)$, where $0 \leq l(x) \leq 1$, which was an indication of subjects, associated either with R or NR category. If $l(x)$ was greater than the *threshold* = 0.5, the subject was declared as R (responder) and otherwise as a NR (nonresponder). In summary, the LR classifier generated probability values to cater MDD patients as either R or NR to the treatment.

The second classification model employed was SVM classifier with linear kernel [53]. It can classify the feature space based on a 'hyperplane' that separated MDD patients and controls according to the class labels. The SVM is considered as a high-efficiency classifier model and used here for a comparison purposes. According to SVM, a linear decision boundary can be found based on this high-dimensional space. Use of linear kernel instead of a nonlinear kernel reduced the risk of over-fitting the data and improves the performance for our data and significantly reduces the overall model complexity. In summary, the LR classifier generated probability values to cater MDD patients as either MDD patients or controls and the SVM concluded a hyperplane to achieve the maximum classification accuracy.

The third classification model was the NB classification [35], based on generating the conditional posterior probabilities for each sample involving the target groups, i.e., MDD patients and healthy controls. The classifier was formed by assigning the sample to the class for which the sample has higher posterior probabilities.

2.4.5 Validation of classification models

After classifier design, a fair evaluation requires assessment of its performance over a range of selected features and classifier design (with suitable coefficient values until convergence) that corresponds to large number of subjects. To address this consideration, we evaluated classification performance based on tenfold cross-validation by dividing the data sample points (study participants) into ten equal segments. During each round, 9 of the segments were utilized as training subset and the remaining 1 as test subset. Tenfold cross-validation provides a fair test of validation in cases where the data points are limited while utilizing features for both test and train classifier models.

For each feature subset, 10 times runs of the simulations were performed involving tenfold cross-validation to achieve the accuracies, sensitivities and specificities. Since the individual iteration resulted into 10 different values of performance metrics (the accuracy, the sensitivity and the specificity), the final confusion matrix was computed by averaging over 10 times. The performance metrics computed from the confusion matrix were presented by Eqs. (8–10). The sensitivity of a classification model corresponds to the percentage of true cases (TP) which are correctly classified as cases defined by Eq. (8). The specificity of a classification model refers to the percentage of true noncases (TN) which are correctly classified as noncases as described by Eq. (9). The accuracy of a classification model illustrates the percentage of correctly classified cases and noncases among all the example points as depicted in Eq. (10). In the scenario of a confusion matrix, both the FP and FN are considered as type-I and type-II errors. More particularly, the FP stands for false positive, e.g., a depressed patient that is wrongly identified as a healthy control. On the contrary, a FN implies as a healthy person that is wrongly identified as depressed patient.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (9)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (10)$$

To compare the three classification models, *F*-score as defined in Eq. (11) was applied as a measure of choice for interested class. *F*-score (Rijsbergen, 1979) could be interpreted as a weighted harmonic average of precision and recall value where precision was probability that a (randomly selected) patient analyzed to be MDD was really MDDs and recall was the probability that an (randomly selected) MDD patient was correctly identified as

MDD. *F*-score was calculated using harmonic averaging because it preferred the balance between precision and recall so it would determine better for the optimal pair in highly unbalanced datasets. Since we did not have any prior information to either precision or recall, the beta value was set to 1 and the *F*-score was also named F1 score:

$$F = (1 + \beta^2) \times \frac{\text{precision} \times \text{recall}}{\beta \times \text{precision} + \text{recall}} \quad (11)$$

The comparison between the classification models is shown while constructing their receiver operating characteristics (ROC) plots [29]. The plot has shown the performance comparison in graphical form involving their sensitivity and 1-specificity.

3 Results

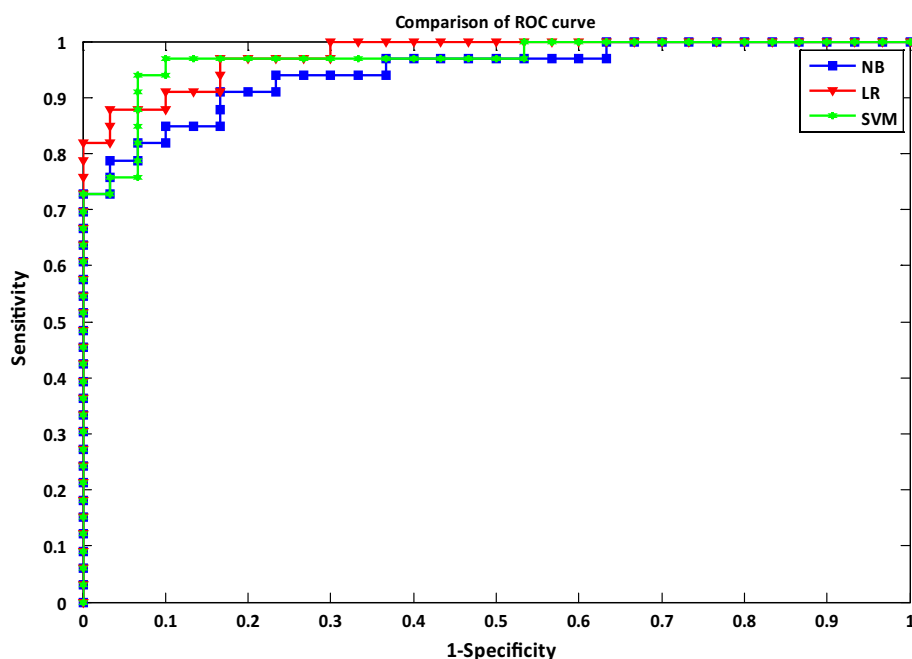
Table 7 shows specific values of the parameters assigned for each classifier during training and testing. Regarding the LR classifier, the link function showed the relationship between the EEG features and clinical outcomes. The value was set as ‘logit’ since the classifier used was logistic regression. Binary classification assumes binomial distribution; this was set as binomial because the data were supposed to originate from two classes, i.e., MDD patients and healthy controls. The offset value was set to 1, whereas the mathematical model of LR classifier has included a constant term. Regarding the SVM classifier, the *C* values were assigned as 0.787 for MDD patient’s class and 1.3684 for the healthy control’s class, accordingly. The parameter value ‘*C*’ indicates an optimum value adjusted because of automatic parameter estimation during training process. The values were computed with formula $(N/2 \times N_1)$ and $(N/2 \times N_2)$, respectively. The ‘*N*’ denoted total number of study participants; N_1 indicated the number of MDD patients and N_2 indicated the number of healthy controls. Other parameters such as ‘Degree of polynomial,’ ‘No. of classes’ and ‘kernel function’ were assigned as 1, 2 and ‘Linear,’ respectively. The parameters for Naïve Bayesian were assigned with normal uniform distribution for MDD patient’s class and the healthy control class. The parameter value ‘*C*’ indicates an optimum value adjusted because of automatic parameter estimation during training process.

Figure 2 shows the ROC plots for SVM, LR and NB classifiers implied in this study. According to the plots, the classification efficiencies are comparable. However, the SVM classifier performed better than others.

Table 8 describes the most significant features found during the feature selection stage. According to the

Table 7 Performance of extracted EEG (Infinity Ref. Data) features based on logistic regression (LR) classification while discriminating MDD patients and healthy controls

Sr.	Classifier	Parameters	Value
1	Logistic regression classification	Link function	Logit
		Distribution	Binomial
		Offset	1
		Constant term	A constant term is added in the model
2	Support vector machine (SVM)	C for class 1 ($N/2 \times N1$)	0.787
		C for class 2 ($N/2 \times N2$)	1.3684
		Degree of polynomial	1
		No. of classes	2
		Kernel function	Linear
3	Naïve Bayesian classification	Distribution	Normal
		Prior	Uniform distribution for all classes

Fig. 2 The ROC plots for SVM, LR and NB

selected features, the abnormal brain areas such as the frontal, temporal and occipital regions have shown the features which are most significant.

Since the features have tested one by one as mentioned earlier and the best accuracy achieved has been reported only. Tables 9, 10 and 11 include the same EEG features, and these features are the most adequate according to the feature selection method.

Table 9 shows results of classification performance based on the LR classification model. According to the table, the highest performance (*accuracy* = 91.7%,

sensitivity = 86.66%, *specificity* = 96.6% and *f-measure* = 0.90) was achieved with synchronization likelihood. The second highest accuracy was achieved with mutual information measure. The integration of these features provided the performance such as *accuracy* = 92%, *sensitivity* = 88.3%, *specificity* = 94.5% and *f-measure* = 0.90.

Table 10 shows results of classification performance based on the NB classification model. According to the table, the highest performance (*accuracy* = 93.6%, *sensitivity* = 100%, *specificity* = 87.9% and *f-measure* = 0.95) was achieved with synchronization likelihood. The second

Table 8 Most significant functional connectivity between scalp locations

Functional connectivity between electrode pairs	Frequency band	Absolute z -values	p value
Fp1-Fp2	Delta	0.5000	0.022
C3-F4	Delta	0.4990	0.008
F7-T3	Delta	0.4980	0.010
F3-T7	Delta	0.4970	0.016
F7-F8	Theta	0.4960	0.022
T4-T5	Theta	0.4949	0.022
F8-T3	Theta	0.4939	0.008
F3-T4	Delta	0.4929	0.010
F4-T7	Theta	0.4919	0.002
F8-Fp1	Delta	0.4909	0.045
T3-F4	Theta	0.4899	0.0021
C4-T3	Delta	0.4879	0.016
F8-T3	Theta	0.4869	0.022
T4-T3	Delta	0.4859	0.013
P3-Fp1	Theta	0.4848	0.022

highest accuracy was achieved with mutual information measure. The integration of these features provided the performance such as *accuracy* = 92.6%, *sensitivity* = 100%, *specificity* = 84.5% and *f-measure* = 0.93.

Table 11 shows results of classification performance based on the SVM classification model. According to the table, the highest performance (*accuracy* = 98%,

sensitivity = 99.9%, *specificity* = 95% and *f-measure* = 0.97) was achieved with synchronization likelihood. The second highest accuracy was achieved with mutual information measure. The integration of these features provided the performance such as *accuracy* = 94.7%, *sensitivity* = 98.3%, *specificity* = 91.4% and *f-measure* = 0.95.

Figure 3 shows the diagnosis accuracy as a function of number of features (while showing for 100 features). The figure reveals the fact that by increasing the number of features the classification accuracy increases and at a certain point shows constant behavior. The figure depicts a constant accuracy as the number of features keeps increasing because the addition of any new features could not add useful information to the classifier model.

In this study, all features were tested one by one. It has been observed that the SVM classifier accuracy remained constant after certain number of features. In results section, the classification tables have reported the best accuracies only. According to this scheme, the features that show maximum classification accuracy would be considered as a reduced set of features sufficient for classification.

In this study, a highest performance accuracy (98%) has been reported which is not shown previously, e.g., the study only concluded with 90% accuracy [21]. Knott et al. [26] have reported a classification of 91.3% involving 70 depressed patients and 23 normal subjects' based on linear features. However, due to imbalance between the numbers of samples in 2 groups the classification results may be biased toward the MDD patient's class. Moreover, in their

Table 9 LR classification while discriminating MDD patients and healthy controls

EEG features	Accuracy (%)	Sensitivity (%)	Specificity (%)	f -measure
Synchronization likelihood	91.7	86.6	96.6	0.90
Interhemispheric coherence	77.85	83.3	71.6	0.77
Mutual information (MI)	79.1	83.3	82.0	0.78
Integration of SL, coherence and MI	92	88.3	94.5	0.90

Table 10 NB classification while discriminating MDD patients and healthy controls

EEG features	Accuracy (%)	Sensitivity (%)	Specificity (%)	f -measure
Synchronization likelihood	93.6	100	87.9	0.95
Interhemispheric coherence	89	86.6	91.6	0.88
Mutual information (MI)	70.8	60	80.8	0.63
Integration of SL, coherence and MI	92.6	100	85.4	0.93

Table 11 SVM classification while discriminating MDD patients and healthy controls

EEG features	Accuracy (%)	Sensitivity (%)	Specificity (%)	f -measure
Synchronization likelihood	98	99.9	95.0	0.97
Interhemispheric coherence	87.3	81.6	91.8	0.87
Mutual information (MI)	69.5	67.8	74.3	0.62
Integration of SL, coherence and MI	94.7	98.3	91.4	0.95

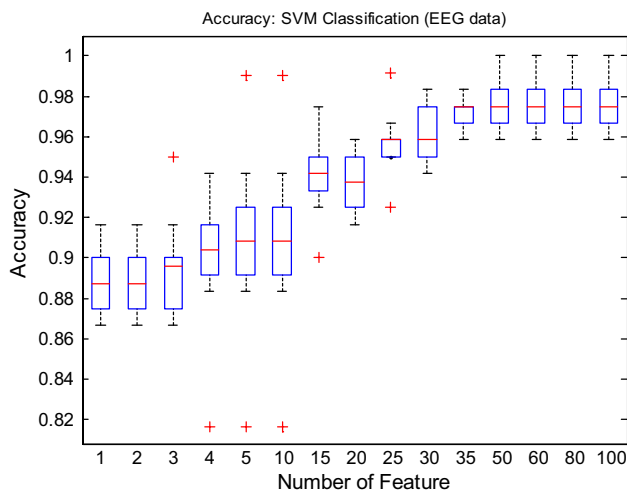


Fig. 3 SVM accuracy as a function of number of features

study Lee et al. [27] analyzed EEG data using detrended fluctuation analysis (DFA) for 11 depressed and 11 controls. The study concluded higher DFA values in MDD patients when compared with controls. Due to small sample sizes the generalization of result could not be possible. In another study, EEG analysis based on wavelet entropy analysis has achieved 80% accuracy while involving 26 MDD participants only [30].

Both studies have utilized different methods. For examples, the first study [48] has utilized fMRI data and built an SVM classifier, whereas the second study [17, 48] investigated the functional brain network metrics with SVM. On the contrary, the present study advocates the use of functional connectivity measure as a discriminating factor between the MDD patients and healthy controls.

4 Discussion

In this paper, a machine learning (ML) framework has been proposed that involved the EEG-derived SL features as input data. It was based on the hypothesis that the SL could discriminate the MDD patients and the healthy controls with an acceptable accuracy better than measures such as interhemispheric coherence and mutual information. In this work, the classification models such as SVM, LR and NB were employed to model the relationship between the MDD patient and their respective groups and ultimately achieved the discrimination of the study participants. Consequently, the model could predict a new participant as either a MDD patient or a healthy subject.

In this study, the SVM classifier has been used with linear kernel because its performance has shown better results than the nonlinear polynomial and Gaussian kernel. This

behavior might be since the sample size for properly training the nonlinear kernel might be insufficient for this study. However, a linear kernel is found suitable in this scenario. In this study, both linear and nonlinear kernels were tested. Consequently, the linear kernel performed better than the nonlinear kernel. Therefore, only the results from linear kernel have been reported.

In this study, an option of calculating the classifier efficiency while considering all the features has been included. In case of SVM, the accuracy started increasing with the increase in the number of features and remains constant at a stage. Therefore, we recommended to perform the selection of features before applying to data for classification purposes.

EEG as a standard modality offers high temporal resolution and lower cost than functional magnetic resonance imaging (fMRI) which makes it suitable for portable and remote clinical applications such as automated monitoring of epileptic patients [4, 7], quantifying sleep stages [3], indexing anesthesia dosage [43] and screening patients with alcohol addiction [38]. In addition, the EEG data acquisition is faster than fMRI and a trained nursing staff could handle the EEG-based CAD system. Furthermore, other applications may include detection of schizophrenia [22] and predicting treatment response for depression [28]. Based on these evidences, EEG has been considered as a modality of choice for this study.

In this paper, we have introduced a new ML scheme that provides automatic diagnosis for MDD patients involving SL as input data. The aim is to achieve higher efficiencies so that EEG utilization can be proved feasible as a clinical diagnostic tool. The advantage of employing the ML technique is to incorporate objectivity during the diagnosis of depression. In contrary, the conventional clinical methods involve pen-and-paper techniques, main depending on well-structured questionnaires and interviews from MDD patients. On the other hand, during the diagnosis process, questionnaire-based methods may introduce subjectivity and may be vulnerable to human errors. Moreover, the conventional methods are time-consuming and cannot be standardized because of heterogenetic behavior of MDD. Conventionally, MDD is diagnosed based on the DSM-IV [1].

The EEG features such as power computation from the frequency bands such as delta, theta, alpha and beta bands, alpha asymmetry and wavelet coefficients from the db4 window function have shown their promise. The translation ability of previous QEEG-based methods to clinical purposes is lacking due to multiple factors including less efficient methods, small sample sizes that poses a hindrance to the generalization of the methods and methodological difference incorporated by the heterogeneity of sample populations and the disease conditions. MDD is

also considered as a heterogeneous disease. In contrary, our proposed method has shown highest efficiencies indicating that diagnosis of MDD can be made automatic by utilizing EEG data.

In this study, we have achieved highest performance (97%) accuracies which are not shown previously, e.g., the study only concluded with 90% accuracy [21]. In addition to EEG, the ERP features such as the P600 component-based SVM classification model achieved more than 90% efficiency [24]. However, the method failed to provide high values of specificities. Knott et al. [26] have reported a classification of 91.3% involving 70 depressed patients and 23 normal subjects' based on linear features. However, due to imbalance between the numbers of samples in 2 groups the classification results may be biased toward the MDD patient's class. Moreover, in their study Lee et al. [27] analyzed EEG data using detrended fluctuation analysis (DFA) for 11 depressed and 11 controls. The study concluded higher DFA values in MDD patients when compared with controls. Due to small sample sizes, the generalization of result could not be possible. In another study, EEG analysis based on wavelet entropy analysis has achieved 80% accuracy while involving 26 MDD participants only [30]. In addition to EEG study, an fMRI study has achieved 90% accuracy [40].

In comparison with all these studies, our proposed methodology has achieved highest accuracy based on linear EEG features. This proves our hypothesis that a careful selection of ML techniques can achieve high performances even with linear features as input data. For example, it is evident from the results that not every frequency band is efficient enough for each classification method. In addition, a combination of the frequency bands resulted into lower classification efficiency than the individual frequency bands. This observation is in accordance with the research ideology that useful features may be discovered from EEG sub-bands as compared to EEG full-band analysis [8]. Moreover, the highest classification performances are provided by SVM by using comparatively more number of features as compared with LR and NB. This is because the SVM has more complex structure than the LR and NB classifiers. Hence, the SVM requires more data samples to train appropriately compared with LR and NB classifiers.

To remove the likelihood that the resulting classifier models are concluded due to noise present in the EEG data, we have adopted the following precautions. First, during preprocessing, the artifacts are carefully removed and tested by plotting their histogram plots to check the presence of any kind of outliers and found the data to be suitable for classification purposes. Second, we have selected equal sample sizes in both the groups. In addition to this

the gender distribution is equal among the groups as well. This is eliminating the gender biasness from the final results. Third, the logistic regression and Naive Bayesian classifier models are relatively simpler than the support vector machine classifier. Therefore, the SVM takes more number of features to produce the classification results. The incorporation of classifier with 3 different structures has proved the validity of our data also. Fourth, the over-fitting may happen; therefore, we have incorporated 100-time permutation test with tenfold cross-validation to improve the robustness of the underlying models.

The study is confounded with few limitations. First, the MDD patients are asked to be medication free at least 2 weeks before the first EEG recording. However, the effect of medication cannot be ruled out completely. Second, since the patients are outpatients, variable such as sleep patterns, appetite and lifestyle could not be controlled. The small sample size poses a constraint that the results should not be generalized to a wider population. As a future work, the creation of a complete dataset and the validation of the method over it will be provided. In addition, EEG as a single modality could be integrated with other modalities such as fMRI or functional near-infrared imaging (fNIRS) so that the multimodal data could be achieved and analyzed. The different types of data fusion could provide a wider perspective to study the brain. Hence, the fusion of different modalities could implicate into better classification models.

5 Conclusion

In conclusion, despite such limitations the classifier models have shown promising discrimination abilities. The high values of specificities may implicate that the proposed ML framework could be feasible for clinical applications. The ML method can learn the patterns specific to each group and can discriminate the two study groups. Hence, the proposed method can formulate the basis of an automated diagnosis system that can be utilized as an assistant in clinics. Since the study has based on relatively small sample size, caution must be adopted while translating these findings to clinical utilities. The selection of minimum number of features can simplify the models. In addition, a method based on less number of features would be simple in interpretation. The less computational burden encourages the system to be commercialized.

Acknowledgement This research work was supported by HiCoE research fund, Grant Number (0153CA-005), Ministry of Higher Education (MoHE), Malaysia.

Compliance with ethical standards

Conflict of interest There is no conflict of interest found for this study.

References

1. A. American Psychiatric Association and A. P. Association (1994) Diagnostic and statistical manual of mental disorders
2. A. P. Association (2013) Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub, Arlington
3. Acharya UR, Faust O, Kannathal N, Chua T, Laxminarayan S (2005) Non-linear analysis of EEG signals at various sleep stages. *Comput Methods Programs Biomed* 80:37–45
4. Acharya UR, Sree SV, Alvin APC, Yanti R, Suri JS (2012) Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals. *Int J Neural Syst* 22:1–14
5. Acharya UR, Sudarshan V, Adeli H, Santhosh J, Koh J, Adeli A (2015) Computer-aided diagnosis of depression using EEG signals. *Eur Neurol* 73:329–336
6. Acharya UR, Sudarshan VK, Adeli H, Santhosh J, Koh JE, Puthankatti SD et al (2015) A novel depression diagnosis index using nonlinear features in EEG signals. *Eur Neurol* 74:79–83
7. Adeli H, Ghosh-Dastidar S, Dadmehr N (2007) A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy. *IEEE Trans Biomed Eng* 54:205–211
8. Ahmadlou M, Adeli H, Adeli A (2012) Fractality analysis of frontal brain in major depressive disorder. *Int J Psychophysiol* 85:206–211
9. Alhaj H, Wisniewski G, McAllister-Williams RH (2011) The use of the EEG in measuring therapeutic drug action: focus on depression and antidepressants. *J Psychopharmacol* 25:1175–1191
10. Anand A, Li Y, Wang Y, Wu J, Gao S, Bukhari L et al (2005) Activity and connectivity of brain mood regulating circuit in depression: a functional magnetic resonance study. *Biol Psychiatry* 57:1079–1088
11. Bae JN, MacFall JR, Krishnan KRR, Payne ME, Steffens DC, Taylor WD (2006) Dorsolateral prefrontal cortex and anterior cingulate cortex white matter alterations in late-life depression. *Biol Psychiatry* 60:1356–1363
12. Berg P, Scherg M (1994) A multiple source approach to the correction of eye artifacts. *Electroencephalogr Clin Neurophysiol* 90:229–241
13. Culpepper L (2014) Misdiagnosis of bipolar depression in primary care practices. *J Clin Psychiatry* 75:e05–e05
14. Erguzel TT, Ozekes S, Tan O, Gultekin S (2015) Feature selection and classification of electroencephalographic signals an artificial neural network and genetic algorithm based approach. *Clin EEG Neurosci* 46:321–326
15. Fingelkurts AA, Fingelkurts AA, Rytälä H, Suominen K, Isometsä E, Kähkönen S (2007) Impaired functional connectivity at EEG alpha and theta frequency bands in major depression. *Hum Brain Mapp* 28:247–261
16. Greenberg PE, Fournier A-A, Sisitsky T, Pike CT, Kessler RC (2015) The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J Clin Psychiatry* 76:155–162
17. Guo H, Cao X, Liu Z, Li H, Chen J, Zhang K (2012) Machine learning classifier using abnormal brain network topological metrics in major depressive disorder. *NeuroReport* 23:1006–1011
18. Hamilton JP, Gotlib IH (2008) Neural substrates of increased memory sensitivity for negative stimuli in major depression. *Biol Psychiatry* 63:1155–1162
19. Hoechstetter K, Berg P, Scherg M (2010) BESA research tutorial 4: distributed source imaging. *BESA Research Tutorial*, pp 1–29
20. Hosmer DW Jr, Lemeshow S (2004) Applied logistic regression. Wiley, New York
21. Hosseinifard B, Moradi MH, Rostami R (2013) Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Comput Methods Programs Biomed* 109:339–345
22. Jalili M, Knyazeva MG (2011) EEG-based functional networks in schizophrenia. *Comput Biol Med* 41:1178–1186
23. Jasper HH (1958) The ten twenty electrode system of the international federation. *Electroencephalogr Clin Neurophysiol* 10:371–375
24. Kalatzis I, Piliouras N, Ventouras E, Papageorgiou CC, Rabavilas AD, Cavouras D (2004) Design and implementation of an SVM-based computer classification system for discriminating depressive patients from healthy controls using the P600 component of ERP signals. *Comput Methods Programs Biomed* 75:11–22
25. Klöppel S, Abdulkadir A, Jack CR, Koutsouleris N, Mourão-Miranda J, Vemuri P (2012) Diagnostic neuroimaging across diseases. *Neuroimage* 61:457–463
26. Knott V, Mahoney C, Kennedy S, Evans K (2001) EEG power, frequency, asymmetry and coherence in male depression. *Psychiatry Res Neuroimaging* 106:123–140
27. Lee J-S, Yang B-H, Lee J-H, Choi J-H, Choi I-G, Kim S-B (2007) Detrended fluctuation analysis of resting EEG in depressed outpatients and healthy controls. *Clin Neurophysiol* 118:2489–2496
28. Lee T-W, Wu Y-T, Yu YW-Y, Chen M-C, Chen T-J (2011) The implication of functional connectivity strength in predicting treatment response of major depressive disorder: a resting EEG study. *Psychiatry Res Neuroimaging* 194:372–377
29. Lehmann C, Koenig T, Jelic V, Prichep L, John RE, Wahlund L-O et al (2007) Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). *J Neurosci Methods* 161:342–350
30. Li Y, Li Y, Tong S, Tang Y, Zhu Y (2007) More normal EEGs of depression patients during mental arithmetic than rest. In: Noninvasive functional source imaging of the brain and heart and the international conference on functional biomedical imaging, 2007. Joint meeting of the 6th international symposium on NFSI-ICFBI 2007, pp 165–168
31. Liu H, Motoda H (2007) Computational methods of feature selection. CRC Press, Boca Raton
32. Magnin B, Mesrob L, Kinkingnéhun S, Pélérini-Issac M, Colliot O, Sarazin M et al (2009) Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology* 51:73–83
33. Mahmud WMRW, Awang A, Herman I, Mohamed MN (2004) Analysis of the psychometric properties of the Malay version of Beck Depression Inventory II (BDI-II) among postpartum women in Kedah, north west of peninsular Malaysia. *Malays J Med Sci MJMS* 11:19
34. Mamitsuka H (2006) Selecting features in microarray classification using ROC curves. *Pattern Recogn* 39:2393–2404
35. Mitchell TM (1997) Machine learning. WCB ed: McGraw-Hill, Boston
36. Mohammadi M, Al-Azab F, Raahemi B, Richards G, Jaworska N, Smith D et al (2015) Data mining EEG signals in depression for their diagnostic value. *BMC Med Inform Decis Mak* 15:1

37. Mumtaz W, Malik AS, Yasin MAM, Xia L (2015) Review on EEG and ERP predictive biomarkers for major depressive disorder. *Biomed Signal Process Control* 22:85–98
38. Mumtaz W, Vuong PL, Xia L, Malik AS, Rashid RBA (2016) Automatic diagnosis of alcohol use disorder using EEG features. *Knowl-Based Syst* 105:48–59
39. Mumtaz W, Xia L, Ali SSA, Yasin MAM, Hussain M, Malik AS (2017) Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD). *Biomed Signal Process Control* 31:108–115
40. Mwangi B, Ebmeier KP, Matthews K, Steele JD (2012) Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder. *Brain* 135:1508–1521
41. Olbrich S, Arns M (2013) EEG biomarkers in major depressive disorder: discriminative power and prediction of treatment response. *Int Rev Psychiatry* 25:604–618
42. Olbrich S, Tränkner A, Chittka T, Hegerl U, Schönknecht P (2014) Functional connectivity in major depression: increased phase synchronization between frontal cortical EEG-source estimates. *Psychiatry Res Neuroimaging* 222:91–99
43. Olofsen E, Sleight J, Dahan A (2008) Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect. *Br J Anaesth* 101:810–821
44. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A (2012) Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev* 36:1140–1152
45. Park C-A, Kwon R-J, Kim S, Jang H-R, Chae J-H, Kim T et al (2006) Decreased phase synchronization of the EEG in patients with major depressive disorder. *World Congress Med Phys Biomed Eng* 2007:1095–1098
46. Puthankattil SD, Joseph PK (2012) Classification of EEG signals in normal and depression conditions by ANN using RWE and signal entropy. *J Mech Med Biol* 12:1240019
47. Qin Y, Xu P, Yao D (2010) A comparative study of different references for EEG default mode network: the use of the infinity reference. *Clin Neurophysiol* 121:1981–1991
48. Ramasubbu R, Brown MR, Cortese F, Gaxiola I, Goodyear B, Greenshaw AJ et al (2016) Accuracy of automated classification of major depressive disorder as a function of symptom severity. *NeuroImage Clin* 12:320–331
49. Stam C, Van Dijk B (2002) Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets. *Phys D* 163:236–251
50. Stam C, Montez T, Jones B, Rombouts S, Van Der Made Y, Pijnenburg Y et al (2005) Disturbed fluctuations of resting state EEG synchronization in Alzheimer's disease. *Clin Neurophysiol* 116:708–715
51. Takens F (1981) Detecting strange attractors in turbulence. In: *Dynamical systems and turbulence*, Warwick 1980 (ed). Springer, pp. 366–381
52. Theiler J (1986) Spurious dimension from correlation algorithms applied to limited time-series data. *Phys Rev A* 34:2427
53. Vapnik VN, Vapnik V (1998) *Statistical learning theory*, vol 2. Wiley, New York
54. Volkert J, Schulz H, Härter M, Włodarczyk O, Andreas S (2013) The prevalence of mental disorders in older people in Western countries—a meta-analysis. *Ageing Res Rev* 12:339–353
55. Willner P, Scheel-Krüger J, Belzung C (2013) The neurobiology of depression and antidepressant action. *Neurosci Biobehav Rev* 37:2331–2371
56. Zeng LL, Shen H, Liu L, Hu D (2014) Unsupervised classification of major depression using functional connectivity MRI. *Hum Brain Mapp* 35:1630–1641



Dr. Wajid Mumtaz is affiliated with Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi PETRONAS, Malaysia. His research interest includes neuro-signal processing, machine learning for diagnosing neurological disorder, e.g., depression and alcohol use disorder.



Dr. Syed Saad Azhar Ali is attached with Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi PETRONAS, Malaysia. His research interests include signal and image processing for biomedical engineering.



Dr. Mohd Azhar Mohd Yasin is attached with Psychiatry Clinic, Hospital Universiti Sains Malaysia. His research interests include techniques based on electroencephalography to improve patient care in major depression, schizophrenia and substance abuse disorder.



Dr. Aamir Saeed Malik is from Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi PETRONAS, Malaysia. His research interests include signal and image processing for biomedical applications.