Journal of Medical Systems (2020) 44: 28
https://doi.org/10.1007/s10916-019-1486-z

SYSTEMS-LEVEL QUALITY IMPROVEMENT

Classification of Depression Patients and Normal Subjects Based on Electroencephalogram (EEG) Signal Using Alpha Power and Theta Asymmetry

Shalini Mahato[1] · Sanchita Paul[1]

Received: 11 November 2018 / Accepted: 15 October 2019 / Published online: 13 December 2019

Abstract

Depression or Major Depressive Disorder (MDD) is a mental illness which negatively affects how a person thinks, acts or feels. MDD has become a major disease affecting millions of people presently. The diagnosis of depression is questionnaire based and is not based on any objective criteria. In this paper, feature extracted from EEG signal are used for the diagnosis of depression. Alpha, alpha1, alpha2, beta, delta and theta power and theta asymmetry was used as feature. Alpha1, alpha2 along with theta asymmetry was also used as a feature. Multi-Cluster Feature Selection (MCFS) was used for feature selection when feature combination was used. The classifiers used were Support Vector Machine (SVM), Logistic Regression (LR), Naïve-Bayesian (NB) and Decision Tree (DT). Alpha2 showed higher classification accuracy than alpha1 and alpha power in all applied classifier. From t-test it was found that there was a significant difference in the theta power of left and right hemisphere of normal subjects, but there was no significant difference in depression patients. Average theta asymmetry in normal subjects is higher than MDD patients but the difference in theta asymmetry in normal subjects and MDD patients is not significant. The combination of alpha2 and theta asymmetry showed the highest classification accuracy of 88.33% in SVM.

Keywords Major depressive disorder (MDD) · Multi-cluster feature selection (MCFS) · Support vector machine (SVM) · Logistic regression (LR) · Naïve-Bayesian (NB) and decision tree (DT)

Introduction

The world is moving ahead fast and so is the aspiration of individuals. The desire for more has become a habit of the human society. Any form of failure seems to be unacceptable to them, thereby pushing them towards a state of depression.

Be it children, youth or adults, nobody seems to escape out of it. It affects the daily chores of life. The sleep gets affected, the working enthusiasm gets hampered, the dietary interest gets lost and the general health gets deteriorated. Psychologists have been able to analyze depression on the basis of standardized questionnaire i.e. Diagnostic and Statistical Manual of Mental Disorder (DSM-V) [1]. The diagnosis is subjective in nature. Physiological data associated with the EEG signals can be an effective tool for diagnosis of depression for providing a more scientific approach to the problem. The bio-electric signals captured by the EEG electrodes are the sum of post-synaptic potentials in the cerebral cortical neurons. EEG is affordable to the common mass with no risk involved. It can easily be administered and gives a real time analytics of data.

World Health Organization projects threats due to depression to take a toll of more than 300 million worldwide and it is expected to go further ahead at much faster rate [2].

Substantial amount of research has been done in order to improve the classification accuracy [3].

Prediction accuracy of 88.6% was achieved with SVM along with genetic algorithm for selection and band power was used as feature [4]. Classification accuracy of 79.27% was achieved with only three electrodes (Fp1, Fp2 and Fz) using K-Nearest Neighbor (KNN) as a classifier along with Min-Redundancy-Max-Relevance (mRMR) for feature

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

✉ Shalini Mahato
swarup.shalini@gmail.com

[1] Department of Computer Science and Engineering, Birla Institute of Technology, Ranchi, Mesra 835215, India

selection out of total 14 features derived from time domain, frequency domain and non-linear feature [5].

Alpha peak and high frontal asymmetry was found to be present in subjects suffering from depression or having history of depression [6]. It was found that alpha-waves were lower in depression patients as compared to normal subjects in Eyes Closed (EC) as well as Eyes Open (EO) condition [7]. MDD patients showed less frontal alpha activity as compared to normal subjects during withdrawal and approach based emotional challenge [8]. On contrary, it was found that alpha asymmetry was significantly higher in MDD patients as compared to normal subjects [9]. In this study, delta, theta, alpha, alpha1, alpha2, beta and theta asymmetry was used as features and classifier used are SVM, NB, LR and DT along with MCFS as feature selection technique as shown in Fig. 1.

Material and Methods

Subject

The data set used in the study was described in the paper by Mumtaz et al. and is publically available [9]. The complete data set consists of eyes closed (EC) and eyes open (EO) data of 34 MDD patients (17 females +17 males, mean age = 40.3 $\pm$ 12.9 yrs) and age matched 30 normal subjects (9 females +21 males, mean age = 38.3 $\pm$ 15.6 yrs). The subjects were the outpatients of Hospital Universiti Sains Malaysia (HUSM), Malaysia. The experimental setup of the study [9] was passed by ethical committee of HUSM. DSM IV criteria was used for detection of depression [10]. Written consent was taken from the subjects and were totally informed about the total experimental setup.

In this study only EC EEG data of 30 MDD patients and 30 normal subjects have been used.

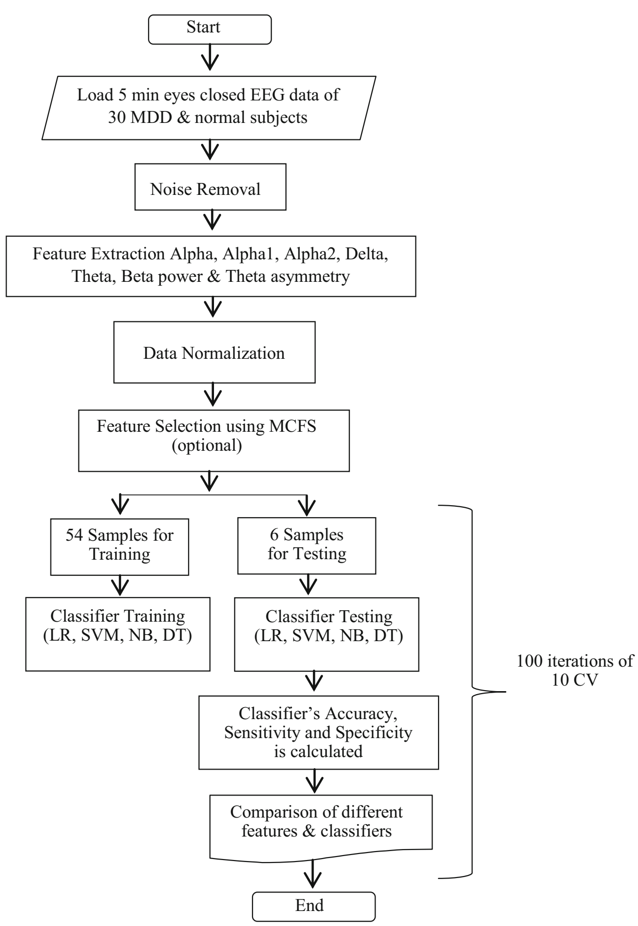

Fig. 1 Flow Chart of the proposed work

Data Acquisition

5 min EC resting state EEG data was recorded using 19 channels in accordance with international 10–20 system as shown in Fig. 2 [11]. EEG recording was done using Linked Ear reference [12]. The sampling rate of 256 samples per second was used. EEG data was band pass filtered from 0.1 to 70 Hz. Noise from power line was removed using 50 Hz notch filter.

EEG Data Preprocessing

In this study all the signal processing and analysis has been done using MATLAB software (version 2015a) along with EEGLAB.

Normally the EEG signal suffers with interference from the background activity such as electrical power line, eye blink, muscle activity, heart muscle activity. For proper signal analysis and classification it is very important to remove the noise from the signal.

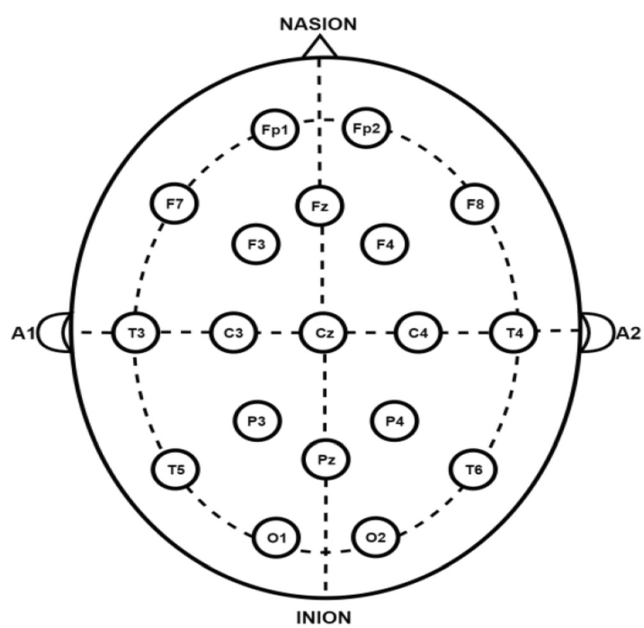

Fig. 2 Electrode placement for International 10–20 system

The linked ear reference is re-referenced to common average reference in order to remove background noise [13]. Since we are interested in the study of band 0.5 to 32 Hz, the signal is band passed from 0.5 to 32 Hz. Independent Component Analysis (ICA) was used for the removal of ocular artifacts [14].

Feature Extraction

In this study band power feature (delta, theta, alpha, beta, alpha1, alpha2) and theta asymmetry (average theta asymmetry and paired theta asymmetry) has been used. All feature sets are normalized using Min-Max normalization.

Band Power

Power of four frequency band i.e. delta (0.5 to 4 Hz), theta (4 to 8 Hz), alpha (8 to 13 Hz) and beta (13 to 32 Hz) was used as feature. Further alpha band was divided into alpha1 (8 to 10.5 Hz) and alpha2 (10.5 to 13 Hz) of which power was again calculated. Power spectrum of the bands was calculated using Welch method.

Inter-Hemispheric Theta Asymmetry

It is calculated as the average of difference of log theta power of left electrodes to the log theta power of right electrodes.

Paired theta asymmetry was calculated as

$$\mathrm{PairedThetaAsymm} = \log(\mathrm{TPowR}) - \log(\mathrm{TPowL}) \tag{1}$$

where, TPowR is theta power of right electrode which can be any of F4, Fp2, F8, T4, C4, O2 and P4.

TPowL is each pairs corresponding left electrode's theta power. Thus paired theta asymmetry is calculated as the difference of log power between the paired electrodes.

The paired electrodes includes: F4-F3, Fp2-Fp1, F8-F7, T4-T3, T6-T5, C4-C3, O2-O1 and P4-P3.

Thus, the total 8 features were added to the feature set for the paired theta asymmetry.

Feature Matrix

The feature matrix consists of n-rows and m-columns where n-rows represents n number of subjects EEG data and m-columns represents m number of features extracted from each subjects EEG data. The feature matrix consists of 60×103. Here 60 is the number of subjects and 103 is the total number of features ($19 \times 5 + 8$). In 19×5, 19 are number of channels and 5 are different band power. 8 are paired theta asymmetry. Basically two types of feature sets are used (i) 60×19 feature set for each alpha, alpha1, alpha2, beta, delta and theta power (ii) 60×27 feature set for combination of alpha+paired theta asymmetry, alpha1 + paired theta asymmetry and alpha2 + paired theta asymmetry.

For the first feature set, no selection method is used. For the second feature set Multi-Cluster Feature Selection (MCFS) is used to select the relevant features.

Feature Reduction & Selection

Data with high dimension reduces the efficiency of the model as well as increases the time and space complexity. Efficiency of the model decreases with large number of features because of overfitting, so dimensional reduction is important. Dimensional reduction can be mainly of two types: feature reduction and feature selection [15].

In feature reduction, high dimensional features are projected into new lower dimensional space. Thus the high dimensional features are converted into lower number of features. Example: Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), etc.

In feature selection, a small number of feature subset is selected which has maximum relevance on the basis of certain relevance evaluation criteria which in term increases the performance of the model. Example, Fisher Score, Lasso, etc.

Feature Selection is better than feature reduction in terms of interpretability and readability. It has also been seen that the feature selection like Variance Inflation Factor (VIF) and Minimum Redundancy and Maximum Relevance (mRMR) provided higher accuracy in the model as compared to feature reduction technique PCA [16].

Feature Selection can be divided into three models based on selection strategy: Filter, wrapper and embedded model [17].

Filter Model uses statistical criteria for evaluation of relevance of a feature. In Wrapper model classifier is used for feature selection. Embedded are combinations of filter and wrapper model. Initially it uses filter model for selecting a fixed number of features and then uses wrapper method to select a subset of feature which provides highest classification accuracy.

In this study we have used MCFS which is a filter based method.

Multi-Cluster Feature Selection (MCFS)

It was developed by Cai et al. [15, 18, 19]. It is a kind of filter based unsupervised feature selection method. It is inspired by spectral analysis of the data and L1 regularised model for subset selection [20, 21]. In this method, the features which best preserves all the clusters in the data is selected. Spectral analysis uses unsupervised learning to measure the correlation between different features. Spectral cluster technique is used to find the cluster in given data sample's using graph Laplacian's top eigen vector.

K-Nearest Neighbor (KNN) is applied for constructing the graph of data sample. The weight matrix W_{ij} on the graph is calculated commonly using either 0–1 weighing, heat kernel weighing or dot-product weighing. Here, W_{ij} implies the nodes of the graph.

Graph-Laplacian G is calculated as

$$G = D - W \tag{2}$$

where, D is the diagonal matrix which is the column sum of W.

The eigen problem is solved by using the following equation:

$$Gy = \lambda Dy \tag{3}$$

Suppose $Y = [y_i, \ldots\ldots y_k]$ where y_k values are calculated on the basis of eq. 3 and highest k eigen vector. Here k is generally the total number of clusters.

Next L1 regression problem in eq. 4 given below is solved using Least Angel Regression (LAR) by minimizing the following objective function [22].

$$\min_{bk} \left\| y_k - X^T a_k \right\|^2 \tag{4}$$

subject to $\|b_k\|_0 = l$.

where b_k is m dimensional vector and $\|b_k\|_0$ denotes the total number of non-zero elements in b_k.

Here X is a given set of data points i.e. $X = [X_1, X_2, \ldots, X_n]$.

MCFS score is calculated for each j feature using the eq. 5

$$MCFScore(j) = \max_k |b_{kj}| \tag{5}$$

where b_{kj} represents j^{th} element of b_k.

Finally top l features are selected according to the MCFS score.

Classifier

LR, SVM, NB and DT have been used as classifier. The performance of all the algorithms have been computed using 10 fold cross validation. Using confusion matrix, sensitivity, specificity and accuracy was calculated for classifiers [23].

Table 1 SVM performance for band power feature

Features Used	Accuracy	Specificity	Sensitivity
Delta	76.93%	90.13%	53.87%
Theta	78.02%	90.03%	78.02%
Alpha	84.50%	96.33%	84.50%
Beta	81.50%	81.67%	81.33%
Alpha 1	85.50%	87.33%	84.50%
Alpha 2	**86.96%**	**89.92%**	**86.00%**

Table 2 SVM performance for combination of alpha band power feature with paired theta asymmetry

Features Used	Accuracy	Specificity	Sensitivity
Alpha 1 + Paired Theta asymm	86.20%	95.00%	77.41%
Alpha 2 + Paired Theta asymm	**88.33%**	**89.41%**	**90.81%**

Support Vector Machine (SVM)

It was developed by Vapnik in 1995 [23, 24]. SVM has the ability to classify both linear and non-linear data. SVM finds the best separating Maximal Marginal Hyperplane (MMH) i.e. largest margin which separates both class separating hyperplane and is represented as

$$V.X + b = 0 \tag{6}$$

where weight vector $V = \{v_1, v_2, \ldots v_n\}$.n is the total number of attributes, b is the bais. X is the training tuple.

Sides of the margin is represented as:

$$S1 : v_0 + v_1 x_1 + v_2 x_2 \geq 1 \quad \text{for } y_i = +1 \tag{7}$$

and

$$S2 : v_0 + v_1 x_1 + v_2 x_2 \leq -1 \quad \text{for } y_i = -1 \tag{8}$$

Here y_i is class label associated with the training tuple.

y_i belongs to $\{+1, -1\}$.

Equation 7 and 8 can be combined to form the equation below:

$$y_i(v_0 + v_1 x_1 + v_2 x_2) \geq 1, \text{for all I} \tag{9}$$

Support vectors are the training tuples that fall on hyperplane.

On the basis of Lagranges theorem MMH can be represented as

$$d(X^T) = \sum_{i=1}^{v} y_i \alpha_i X_i X^T + b_0 \tag{10}$$

Here, X_i is the support vector and y_i is the associated class label. X^T is the test tuple. α_i is Lagrangian multiplier. v is the

Table 3 LR performance for band power feature

Features Used	Accuracy	Specificity	Sensitivity
Delta	66.63%	68.40%	64.87%
Theta	68.90%	67.17%	70.63%
Alpha	72.98%	66.73%	79.23%
Beta	71.58%	72.70%	74.47%
Alpha 1	73.37%	69.23%	77.50%
Alpha 2	**84.05%**	**80.30%**	**87.80%**

Table 4 LR performance for combination of alpha band power feature with paired theta asymmetry

Features Used	Accuracy	Specificity	Sensitivity
Alpha 1 + Paired Theta asymm	74.04%	65.59%	72.50%
Alpha 2 + Paired Theta asymm	**85.82%**	**73.57%**	**82.07%**

number of support vector. b_0 represents numeric parameter calculated by SVM.

When the value of the training tuple put in eq. 10 is positive then the tuple belongs to +1 class otherwise it belongs to −1 class. If the data set is non-linearly separable in lower dimension, it is transformed to the higher dimension space by using kernel function and then further classification is done. Some of the kernel function includes polynomial kernel, sigmoid kernel, Gaussian Radial basis kernel, etc. In this study, Gaussian Radial basis kernel was used.

Naïve Bayesian (NB)

It is based on Baye's theorem [23]. For simplicity independence of feature is assumed and is known as class conditional independence. Baye's theorem computes the probability of occurring an event given another events probability that has already happened. It is mathematically represented as

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)} \tag{11}$$

From P(B/A) we are calculating the probability of an event B, given the event P has already occurred. Here P(B) is the prior probability of event H. Here P(A) is the prior probability of event A. P(A/B) is the posterior probability of event A condition on H. To predict the class label of test tuple X $P(X/Y_i).P(Y_i)$ is calculated for each class Y_i. The predicted class label for X is Y_i for which $P(X/Y_i).P(Y_i)$ is maximum.

Considering class conditional independence is not always practically true which leads to error in prediction.

Table 5 NB performance for band power feature

Features Used	Accuracy	Specificity	Sensitivity
Delta	82.46%	82.94%	88.74%
Theta	83.83%	87.83%	87.30%
Alpha	84.21%	84.65%	89.03%
Beta	83.42%	88.59%	90.11%
Alpha 1	84.54%	85.18%	87.72%
Alpha 2	**84.78%**	**85.23%**	**85.30%**

Table 6 NB performance for features alpha band power feature with paired theta asymmetry

Features Used	Accuracy	Specificity	Sensitivity
Alpha 1 + Paired Theta asymm	85.21%	85.55%	89.33%
Alpha 2 + Paired Theta asymm	**86.63%**	**87.43%**	**86.63%**

Logistic Regression (LR)

It was developed by David Cox in 1958 [25]. LR is a classification technique where the response variable is binary. LR is named after the core function used which is logistic function and regression stands for classification. Logistic or sigmoid function is S-shaped curve of which real input (s belongs to R) and maps into value between 0 & 1. Logistic function is defined as

$$\sigma(s) = \frac{e^s}{e^s + 1} = \frac{1}{1 + e^{-s}} \tag{12}$$

where $s = \beta_0 + \beta_1 X_1 + \ldots.. + \beta_j X_j$.

here β_0 represent intercept.

$\beta_1, \beta_2, \ldots.., \beta_j$ represents coefficient.

$X_1, X_2, \ldots.., X_j$ represents variables.

The logistic function can be rewritten as

$$P(x) = \frac{1}{1 + e^{-\left(\beta_0 + \beta_1 X_1 + \ldots.. + \beta_j X_j\right)}} \tag{13}$$

where P(x) is represented as the probability of dependent variable.

The logit (log odds) transformation of σ(s) is defined as

$$logit(P(x)) = ln\left(\frac{P(x)}{1-P(x)}\right) = \beta_0 + \beta_1 X_1 + \ldots.. + \beta_j X_j \tag{14}$$

Generally maximum likelihood is used to find the value of the parameter.

Decision Tree (DT)

DT is a flowchart which resembles tree structure [23]. Each internal node corresponds to a test node on feature. Each leaf

Table 7 DT performance for band power feature

Features	Accuracy	Specificity	Sensitivity
Delta	76.50%	81.18%	93.33%
Theta	83.17%	92.50%	92.25%
Alpha	82.17%	82.93%	87.58%
Beta	83.33%	75.00%	86.67%
Alpha 1	83.33%	84.83%	83.33%
Alpha 2	**84.00%**	**98.56%**	**83.73%**

Table 8 DT performance for features alpha band power feature with paired theta asymmetry

Features Used	Accuracy	Specificity	Sensitivity
Alpha 1 + Paired Theta asymm	83.83%	84.02%	92.42%
Alpha 2 + Paired Theta asymm	**87.50%**	**89.28%**	**90.83%**

node represents a class label. Outcome of the test is represented by the branch of the tree. The path from the route node to the leaf node represents the classification rule. Attribute selection measures like gain ratio, gini index and information gain are used to select the most significant attribute which is most efficient in class distinction of the tuples. Some of the decision tree methods are ID3 (Iterative Dichotomiser 3), C4.5 and CART (Classification and Regression Trees). In this study CART was used.

Validation

Validation is a way of evaluating how well the model works for the given dataset. In this study, 100 iteration of 10-fold cross validation has been used. Final value of accuracy (ACC), sensitivity (SN) and specificity (SP) was computed based on confusion matrix.

TP Count of total number of MDD patients correctly identified

TN Count of total number of normal subjects correctly identified as normal

FN Count of total number of MDD patients wrongly identified as normal subjects

FP Count of total number of normal subjects wrongly identified as MDD patients

$$SN = \frac{TP}{(TP + FN)} \tag{15}$$

$$SP = \frac{TN}{(TN + FP)} \tag{16}$$

$$ACC = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{17}$$

Results

Statistical Analysis

Paired t-test was used to evaluate significant difference of theta asymmetry used with $p < 0.05$. It was found that there was a significant difference in the theta power of left and right hemisphere (Theta asymmetry) of normal subjects. But there was no significant difference in depression patients.

Average theta asymmetry in normal subjects is higher (0.936) than MDD patients (0.3750) but the difference in theta asymmetry in normal subjects and MDD patients is not significant.

Based on Band Power

From the Tables 1, 2, 3, 4, 5, 6 and 7, it can be observed that out of alpha, theta, delta and beta power, alpha band power showed high classification accuracy in all classifiers. From Table 1 to Table 8, highest performance provided by the classifier has been emphasised in bold. Highest classification accuracy of 84.50% was achieved in SVM using alpha power as shown in Table 1. Since alpha band showed the highest accuracy it was further divided into alpha1 and alpha2. It was found that alpha2 provided the accuracy higher than alpha band as well as alpha1 in all cases as shown in Table 1, 3, 5 and 7. Alpha2 provided highest classification accuracy of 88.33% in SVM as shown in Table 2.

Results based on combination of alpha-power and theta asymmetry: From the Tables 2, 4, 6, 8 it can be seen that combination of alpha2 power with theta asymmetry provided highest accuracy of 88.33% in SVM.

Discussion

Highest classification accuracy based on band power was achieved in alpha band power which is similar to results found in literature [26–29].

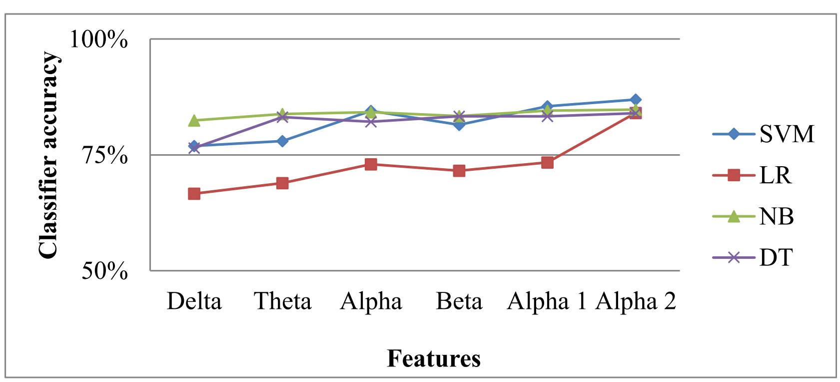

Fig. 3 Comparison of classification accuracy of different classifiers using band power feature

Alpha2 power showed higher classification accuracy than alpha1 power as well as alpha power in all classifiers. Theta asymmetry increases the accuracy significantly when it is used in combination with alpha, alpha1 and alpha2 power. Thus, the potential power of theta asymmetry has been explored along with alpha power. The highest classification 88.33% accuracy was achieved by alpha2 power in combination with theta asymmetry using SVM.

From Fig. 3 it can be observed that out of four classifiers i.e. LR, SVM, NB and DT used, SVM provides higher classification accuracy for most of the features. This is due to the fact that SVM is not prone to overfitting and has the capability to represent decision boundaries which are non-linear and complex.

The main drawback of the study i.e. the EEG data under consideration is collected from less number of individuals. More number of EEG data needs to be considered for the study for generalization of the results.

For future works, other subbands or combination of subbands could be studied. Specific areas of the brain could be considered for the study. Also asymmetry and paired asymmetry could also be studied in other bands and subbands i.e. delta, theta, beta and gamma.

Conclusion

The study explores the strength of alpha power along with theta symmetry in classifying MDD patients and normal subjects using NB, LR, SVM and DT. It was found that alpha2 power provides higher classification accuracy than alpha and alpha1 power. Classification accuracy of alpha2 power was further improved when used along with theta asymmetry. Highest classification accuracy of 88.33% was obtained using alpha2 power and theta asymmetry. Thus, use of alpha2 band as a feature rather than using Alpha band can improve the accuracy significantly. Also combination of alpha2 power with theta asymmetry as feature can further enhance accuracy. Accuracy of SVM classifier is found to be higher as compared to all the other applied classifiers i.e. SVM, LR, NB and DT.

Acknowledgments The authors would like to thank Mumtaz et al. [9] for the dataset contribution.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

The dataset used in this study is described in the paper by Mumtaz et al. and is publically available dataset (https://figshare.com/articles/EEG-based_Diagnosis_and_Treatment_Outcome_Prediction_for_Major_Depressive_Disorder/338516) [9]. According to the contributors of the dataset [9], all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent According to the contributors of the dataset [9], informed consent was obtained from all individual participants included in the study.

References

1. American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders. 5th edition. American Psychiatric Association Washington, DC.
2. World Health Organization, Depression and other common mental disorders Global Health estimates. Geneva: WHO Document Production Services, 2017.
3. Mahato, S., and Paul, S., Electroencephalogram (EEG) signal analysis for diagnosis of major depressive disorder (MDD): A review. In: Nath, V., Mandal, J. K. (Eds), Nanoelectronics, circuits and communication systems (NCCS 2017). Singapore: Springer, 2017, 323–336.
4. Hosseinifard, B., Moradi, M. H., Rostami, R., Classifying depression patients and Normal subjects using machine learning techniques. In: 19th Iranian conference on electrical engineering, Tehran, Iran, pp 339–345, (2011).
5. Cai, H., Han, J., Chen, Y., Sha, X., Wang, Z., et al. A pervasive approach to EEG-based depression detection. In: Hindawi, pp 1–13, (2018).
6. Niemiec, A. J., Lithgow, B. J. Alpha-band characteristics in EEG spectrum indicate reliability of frontal brain asymmetry measures in diagnosis of depression. In: IEEE engineering in medicine and biology 27th annual conference, Shanghai, China, pp 7517–7520, (2005).
7. Kan, D. P. X., Lee, P. F., Decrease alpha waves in depression: An electroencephalogram (EEG) study. In: International conference on BioSignal analysis, processing and systems (ICBAPS), pp 155-161, (2015).
8. Stewart, J. L., Coan, J. A., Towers, D. N., Allen, J. J., Resting and task-elicited prefrontal EEG alpha asymmetry in depression: Support for the capability model. In: Psychophysiology, pp 1–18, (2014).
9. Mumtaz, W., Xia, L., Ali, S. S. A., Yasin, M. A. M., Ali, S. S. A., Malik, A. S., A wavelet-based technique to predict treatment outcome for major depressive disorder. In: PLOS ONE, pp 1–30, (2017).
10. American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders. 4th edition. American Psychiatric Association Washington, DC, pp 339–345
11. Teplan, M., Fundamental of EEG measurement. In: Measurement Science Review, Vol 2, pp 1–11, (2002).
12. Dien, J., Issues in the application of the average reference: Review, critiques and recommendations. In: Behavior Research Methods Instruments & Computers, pp 34–43, (1998).
13. Gandhi, V., Brain computer interfacing for assistive robotics. Electroencephalograms, recurrent quantum neural networks, and user-centric graphical interfaces, 1st edition, academic press, pp 21-29, (2014).
14. Delorme, A., Makeig, S., EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent

component analysis. In: Journal of Neuroscience Methods, pp 9–21, (2004).
15. Alelyani, S., Tang, J., Liu, H., Feature selection for clustering: A review.In: Reddy CK, Agarwal CC, data clustering: Algorithms and applications pp 1-12, (2013).
16. Mahato, S., Roy, A., Verma, A., Paul, S., Analysis of feature selection and reduction techniques for detection of depression using EEG signals. International conference on computational intelligence and data analytics (ICCIDA-2018), Bhubaneswar, (2018).
17. Roffo, G., Melzi, S., Ranking to learn: Feature ranking and selection via eigenvector centrality. In: Lecture Notes in Computer Science, pp 19–35, (2017).
18. Cai, D., Zhang, C., He, X., Unsupervised feature selection for multi-cluster data. In KDD, pp 333-342, (2010).
19. Roffo, G., Melzi, S., Castellani, U., Infinite latent feature selection: A probabilistic latent graph-based ranking approach. IEEE international conference on computer vision (ICCV 2017), Venice, (2017).
20. Belkin, M., Niyogi, P., Laplacian Eigen maps and spectral techniques for embedding and clustering. In: Advances in Neural Information Processing Systems, pp 585–591, (2001).
21. Hastie, T., Tibshirani, R., Friedman, J., The elements of statistical learning. In: Data Mining, Inference, and Prediction, Springer-Verlag, New York, (2001).
22. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., Least angle regression. In: Annals of Statistics, pp 407–499, (2004).
23. Han, J., Kamber, K., Pei, J., Data mining: Concepts and techniques. Morgan Kaufmann, Elsevier, USA, 3rd edition, pp 327-413, (2012).
24. Vapnik, V., Cortes, C., Support-vector networks. In: Machine Learning, pp 273–297, (1995).
25. Kutner, M. H., Nachtsheim, C. J., Neter, J., Applied linear statistical models, 5th edition, McGraw-hill, New York.Pp 555-603, (2005).
26. Mumtaz, W., Xia, L., Ali, S. S. A., Yasin, M. M. A. M., Hussain, M., Malik, A. S., Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD). In: Biomedical Signal Processing and Control, Volume 31, January 2017, pp 108–115, (2017).
27. Grin-Yatsenko, V. A., Baas, I., Ponomarev, V. A., Kropotov, J. D., Independent component approach to the analysis of EEG recordings at early stages of depressive disorders. In: Clinical Neurophysiology, Vol. 121, 2010, pp 281–289, (2010).
28. Mahato, S., Paul, S., Detection of major depressive disorder using linear and non-linear features from EEG signals. In: Microsystem Technologies, Springer, 2018, pp 1–12, (2018).
29. Gollan, J. K., Hoxha, D., Chihade, D., Pflieger, M. E., Rosebrock, L., Cacioppo, J., Frontal alpha EEG asymmetry before and after behavioral activation treatment for depression. In: Biol Psychol, pp 1–28, (2014).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.