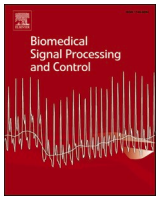
Biomedical Signal Processing and Control 85 (2023) 104923

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Improving EEG major depression disorder classification using FBSE coupled with domain adaptation method based machine learning algorithms

Hadeer Mohammed [a], Mohammed Diykh [a,b,c,*]

[a] University of Thi-Qar, College of Education for Pure Science, Iraq
[b] Information and Communication Technology Research Group, Scientific Research Centre, Al-Ayen University, Iraq
[c] UniSQ College, University of Southern Queensland, QLD, Australia

ARTICLE INFO

Keywords:
Fourier Bessel series expansion
Domain adaptation
Segmentation
LS-SVM
Student *t*-test
Wilcoxon test
Statistical features

ABSTRACT

Major depression disorder (MDD) has become the leading mental disorder worldwide. Medical reports have shown that people with depression exhibit abnormal wave patterns in EEG signals compared with the healthy subjects when they are exposed to positive and negative stimuli. In this paper, we proposed an intelligent MDD detection model based on Fourier-Bessel series expansion (FBSE) coupled with domain adaptation (DA). First, EEG signals are segmented into intervals and each segment is passed through FBSE. Two types of features, including statistical and nonlinear features are investigated and extracted from each FBSE coefficient to detect MDD. Student *t*-test and Wilcoxon test are employed to remove noisy and bad features that can compromise the performance of data-driven learners. Then, DA method named Independence Domain Adaptation was applied to reduce the difference of feature distributions among subjects. The selected features are sent to a least square support vector machine (LS-SVM), and other classifiers named SVM, k-nearest (KNN), ransom forest, Bagged ensemble, boosted ensemble, decision tree, gradient boosting and stacked ensemble for the comparison purpose. Our experiments are simulated by using publicly available dataset. The performance of the proposed model is evaluated in both subject dependence experiment by 10-fold cross validation, subject independence experiment by leave-one-subject-out cross-validation, and Confidence interval respectively. Results showed that the features reduction method can significantly improve the mean accuracy by 4.20. The proposed model is compared with previous studies and the results show that the proposed model outperforms the other methods.

1. Introduction

Major depressive disorder (MDD) is a serious mood disorder which is associated with anger, feelings of sadness, and anxiety [1-3]. It can lead to mental and physical issues combined with cognitive impairments which can increase the risk of developing Alzheimer's disease [64,65]. [4,5] According to the World Health Organization, more than 350 million individuals around the world aged between 20 and 70 are suffered from MDD and it is likely to be the leading cause of disease burden by 2030 [6,7].

Clinical experts mainly diagnose patients with MDD using standard approaches including physical exam, lab test, psychiatric evaluation, and statistical manual of mental disorders. Patients with mild depression disorder can be treated by medications, however, some patients with severe depression symptoms need to be hospitalized as they are at more risk of attempting suicide [8]. As a result, a total of 3.7 %, approximately 850,000 suicidal deaths in the world are due to major depression disorder [9].

The main cause of depression could be genetic or environmental factors [10]. In addition, many recent studies reported that during COVID-19 pandemic the rates of people having depression are dramatically increased due to home quarantine, unemployment, and social distancing [3,7,11,12]. Early diagnosis of depression is crucial to determine the effective treatment and reduce the escalation of the disorder. The traditional approaches of depression diagnosing such as clinical trials and questionnaires receive many criticisms as they are subjective in their nature [13]. However, there is a high demand to design an automatic method for diagnosing MDD to support healthcare professionals. This paper aims to design an intelligent model to diagnose MDD.

In clinical rehabilitation, and assessment units, MDD can be diagnosed when the clinical symptoms start to be developed [14]. Patients

* Corresponding author at: University of Southern Queensland, Australia.
E-mail addresses: Msc21co1@utq.edu.iq (H. Mohammed), Mohammed.diykh@utq.edu.iq, mohammed.diykh@usq.edu.au (M. Diykh).

https://doi.org/10.1016/j.bspc.2023.104923
Received 23 December 2022; Received in revised form 20 March 2023; Accepted 5 April 2023
Available online 14 April 2023

could be referred to an occupational therapist or physical therapist that focuses on motor recovery. This procedure takes a long time, therefore the effects of depression on recovery could be greater. However, there is a critical need for developing an efficient, and intelligent depression detection method to provide a fast and accurate depression diagnosis in the early stage [14–18].

Many studies based on machine learning techniques have been designed to diagnose depression through facial expressions, speech, emotions, electroencephalogram EEG signals and subject's behaviors in social media [3,11–20]. Among them, EEG signals are considered a vital tool to diagnose depression as they reflect any decrease or abnormal patterns in brain activity during depression episodes [6].

For decades, researchers have been working on testing the impact of several transformation techniques such as wavelet and Fourier techniques on the depression classification. For example, Saeedi et al. [18] suggested a model-based wavelet technique coupled with machine leering models to detect depressed patients. A wavelet transform was applied to decompose EEG signals into five frequency bands named delta, theta, alpha, beta, and gamma. A set of features was extracted from each wavelet band and, then a genetic algorithm was employed to select the most significant features. Akbari et al. [19] proposed a reconstructed phase space (RPS) to detect depression detection from EEG signals. Based on RPS, a total of 34 geometrical features were pulled out from each EEG recording. They showed that EEG patterns from the right hemisphere significantly reflected depression compared with the left hemisphere. Bachmann et al [20] investigated various types of EEG features to identify depressed subjects. EEG characteristics including spectral asymmetry index Higuchi's fractal dimension, detrended fluctuation analysis and Lempel-Ziv complexity, alpha power variability and relative gamma power and nonlinear features were extracted. A logistic regression technique was applied to categorise EEG features into control and depressed subjects. Sharma et al., [21] designed a computer aided depression diagnosis model based on a wavelet filter bank. EEG signals were decomposed into seven sub-bands. A least square support vector machine (LS-SVM) was employed to classify EEG features. Mahato et al., [22] used a multi cluster feature selection model to identify the most powerful features.

In the past few years, much attention has been devoted to deep learning models-based EEG for depression diagnostic. Saeedi et al [23] studied the relationships between EEG channels and depression. A brain effective connectivity approach was adopted to convert EEG signal into an image form. A conventional neural network (CNN) model integrated with a Long short-term memory (LSTM) was applied to identify the depressed individuals. Jiang et al., [24] used deep neural networks based facial expressions model. An ImageNet pre-trained based CNN was used to extract emotion features from video recordings. Song et al., [25] integrated a convolution neural network a with long-short term memory to recognize depressed subjects. Movahed et al., [26] studied several EEG features including statistical, spectral, wavelet, functional connectivity, and nonlinear analysis features. A sequential backward feature selection was utilized to select the most relevant features. Harati et al., [27] used speech signals to identify depression severity during Deep Brain Stimulation treatment. A deep neural network was applied to extract emotion features from speech signal.

Many researchers Inspired by text mining techniques and speech signals to diagnose depression. Nikravan et al., [28] studied the effects of Transcranial Magnetic Stimulation on the complexity of EEG signals recorded from depressed subjects. Dessai et al., [29] applied a text mining and natural language processing model to detect depressed users in twitter from their tweets. Alghowinem et al., [30] proposed a framework for depression detection based on head movements, eye activities, and speech signals.

Although existing studies have achieved good performances in MDD detection, most of them only focus on training models for subject dependence classification tasks. Training models in a such way can lead to poor accuracy when models are used to predict new subjects. The main cause is that EEG patterns vary among different subjects. In addition, selection of the optimal channels and features to diagnose MDD is not yet addressed in depth. To cope with these issues, this paper proposes an intelligent model based on the Fourier-Bessel series expansion (FBSE) to diagnose MDD. The FBSE method has been utilised in previous studies for various applications, such as detection of epileptic seizures, detection of alcoholism, elimination artifacts from EEG signals. Chaudhary et al., [31] diagnosed breast cancer using an iterative FBSE. Their results demonstrated the efficiency of the iterative FBSE in the detection of breast cancer from images. Khan et al., [32] classified six eye movements using the FBSE based empirical wavelet transform. Their study demonstrated the effectiveness of the FBSE in analysing EMG and EOM signals. Das et al., [33] studied the effect of chanting Hare Krishna Mantra on EEG rhythms. In that study, the FBSE was utilised to analyse EEG signals. Nalwaya et al., [34] applied FBSE to classify emotion EEG signals. Their results showed that the FBSE was effective tool to reveal the hidden patterns in EEG signals. Pachori et al., [35] combined Fourier Bessel expansion with Wigner–Ville distribution to obtain the time–frequency representation. Pachori et al., [36] used the coefficients of FBSE to segment EEG signals. In that study, a second-order linear time-varying autoregressive was employed for parametric representation of EEG signals.

The main novelty of this paper is that the FBSE is integrated with domain adaptation (DA) model to find common feature representations that are invariant across subjects. In this paper, a combination of nonlinear and statistical features is extracted from the coefficients of FBSE. The most optimal features are selected using Student *t*-test and Wilcoxon test. Then, we applied the DA method to reduce the dispersion of feature distributions of multi-channel MDD EEG signals in subjects' independence classification. The selected features are sent to veracious classification models. In this paper, the most powerful EEG channels to classify the EEG samples into the MDD and healthy subjects are also investigated.

The remainder of this paper is structured as follows: in section 2, introduces the methodology, including dataset description, signal preprocessing, feature extraction, feature selection, domain adaptation method and classifiers. The experiment and discussion are presented in Section 3 and 4, respectively. Section 5 concludes this work.

2. Experimental EEG dataset

The proposed model is evaluated using publicly available dataset which is contributed by Mumtaz et al. [37]. The EEG recordings were collected from 34 MDD subjects (male/female = 17/17, aged 40.3 ± 12.9 years) and 30 health subjects (male/female = 21/9, aged 38.3 ± 15.6 years). DSMIV criteria were applied to diagnose the MDD subjects (American Psychiatric Association 1994). Ethical approval was obtained by Human Ethics committee of Hospital Universitii Sains Malaysia (HUSM), Malaysia [63][38]. The EEG recordings were recorded for 5 min eyes closed condition in resting position. All EEG recording were acquired using 19 electrodes that were deployed on patient's scalp according to the international standard 10/20 system. The 19 electrodes were placed on different regions including: the frontal included 7 electrodes named (Fp1, Fp2, F3, F4, F7, F8, Fz), the temporal included four electrodes (T3, T4, T5, T6), the parietal included three electrodes (P3, P4, Pz), the occipital included two electrodes (O1, O2), and the central included electrodes (C3, C4, Cz). A linked ear (LE) was utilised to record the EEG data. All EEG recording were sampled at frequency of 250 Hz. Fig. 1 Plots nineteen EEG channels with magnitude of FBSE.

3. Research methodology

The objective of this paper is to develop an effective depression diagnose approach based on EEG signals by employing Fourier-Bessel technique coupled with domain adaptaion. Fig. 2 shows the block diagram of the proposed model which mainly consists of the pre-processing

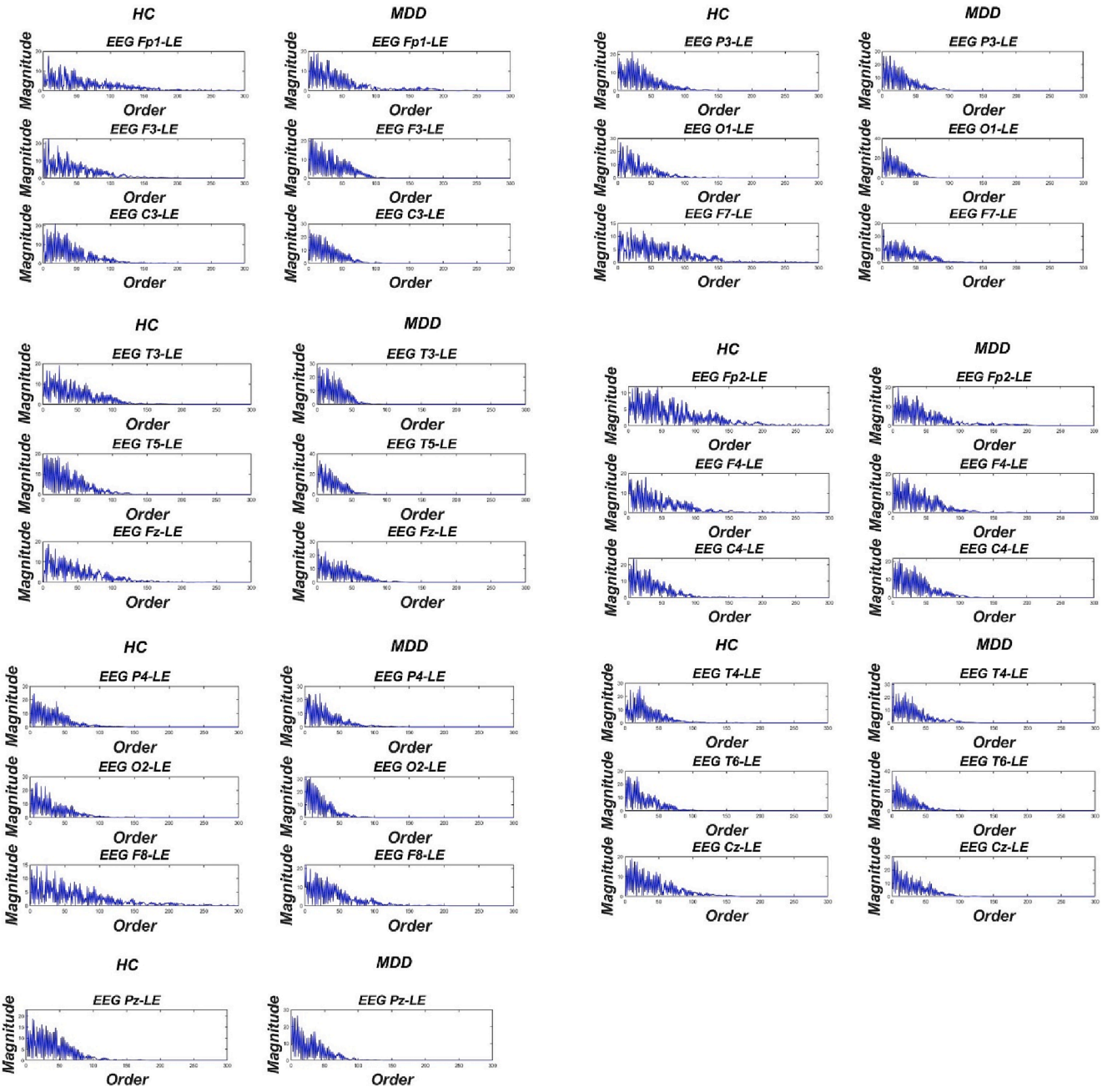


Fig. 1. Plot of nineteen eeg channels with the magnitude of fbse coefficients for healthy and mdd.

phase, feature extraction, feature selection, domain adaptation (DA) and classification.

3.1. EEG pre-processing and segmentation

EEG signals were passed through a high pass filter and low pass filter to remove noises and improve EEG signals quality. In this study, a high pass filter of 0.5 Hz cut off frequency, and low pass filter of 70 Hz cut off frequency is applied to filter EEG signals.

To partition EEG signals, we adopted our segmentation technique in [39–42] to divide each EEG channel into intervals. Fig. 3 illustrates an example of a single EEG channel being segmented. A fixed window of 1 s is applied to partition EEG signals. Suppose S is a single EEG channel signal of n datapoints, $S = \{s_1, s_2, s_3, \cdots\cdots, s_n\}$. Each single EEG channel S was divided into K windows where $K = \{k_1, k_2, \cdots.., k_M\}$, where each window contains d datapoints $k_M = \{x_1, x_2, \cdots., x_d\}$. As a result, each single channel EEG signal is partitioned into $k = 295$ segments, with each interval containing 256 datapoints.

3.2. Feature extraction using Fourier-Bessel technique

Feature extraction aims to pull out meaningful information from signals and remove any redundant data. Effective features can improve the performance of classification models. In medical applications, an appropriate feature extraction technique is vital to transfer raw signals into more informative data. The FBSE has been demonstrated to be an efficient technique in analysing nonstationary signals [34–36]. Many studies have shown that the FBSE has many advantages over FFT [43–48]. First the FBSE technique does not employ any window function for spectral analysis compared with Fourier transform (FT). Second, the length of coefficients produced by the FBSE is equal to length of the signal and the FBSE provides double frequency resolution compared to

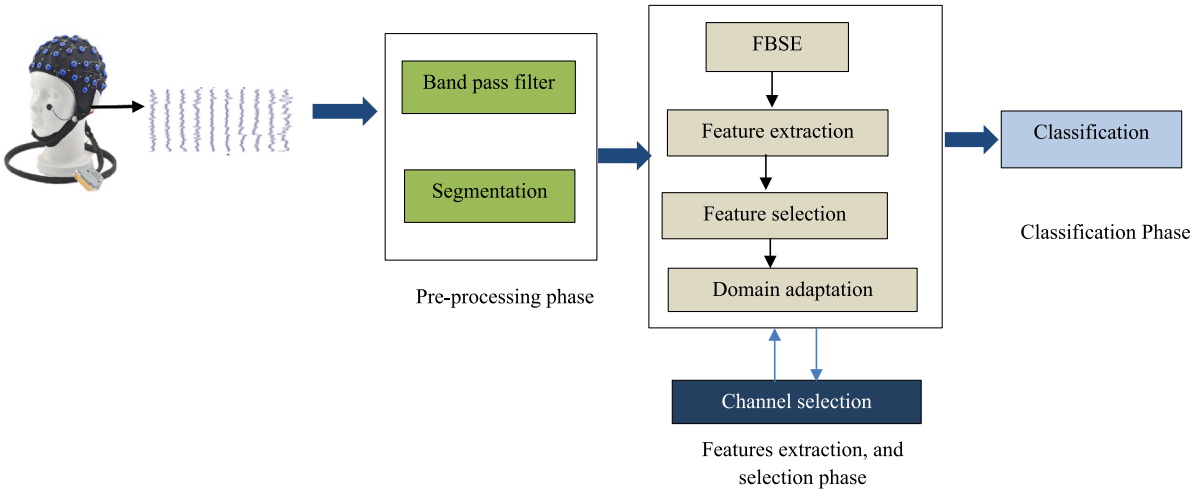


Fig. 2. Framework of the proposed model.

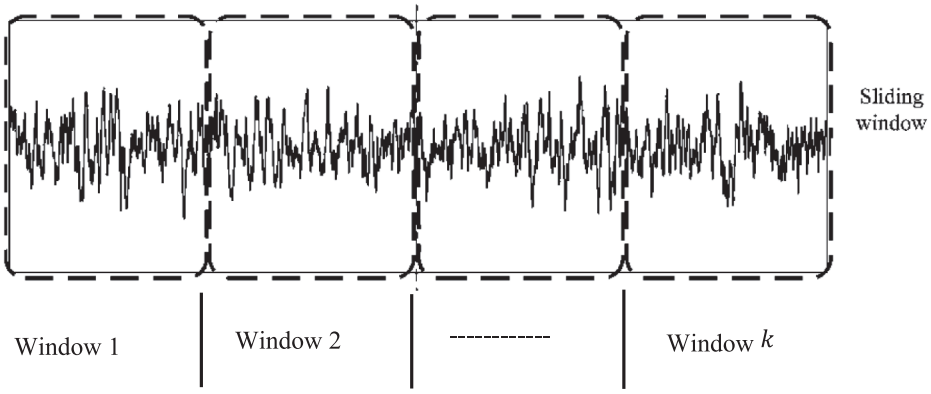


Fig. 3. An example of EEG signal is being segmented into intervals.

DFT. Fig. 4 shows the Bessel function (BF). We can notice that the BF has amplitude modulation, and is non-stationary in nature compared with FT. In addition, The FBSE employs non-stationary BF as a basis function, while FT utilizes stationary a basis function. Fig. 5 depicts the difference between FBSE and FT. Recent work showed that the FBSE is able to provide a compact representation compared to FT. In addition, the FBSE technique produces a good frequency resolution compared to FT. Gupta et al. [43] showed that the FBSE was more efficient to analyse EEG rhythms compared to FT. Tripathy et al., [44] reported that the FBSE coefficients are unique and provide a better spectral resolution compared with FT in analysis non-stationary signals. Gajbhiye et al., [45] indicated that the number of unique FBSE coefficients required for spectral representation was equivalent to the length of the discrete time signal while the FFT requires twice the length of the discrete time signal [43]. Motivated from the above studies, in this paper, the FBSE combined with domain adaptation was employed to diagnose major

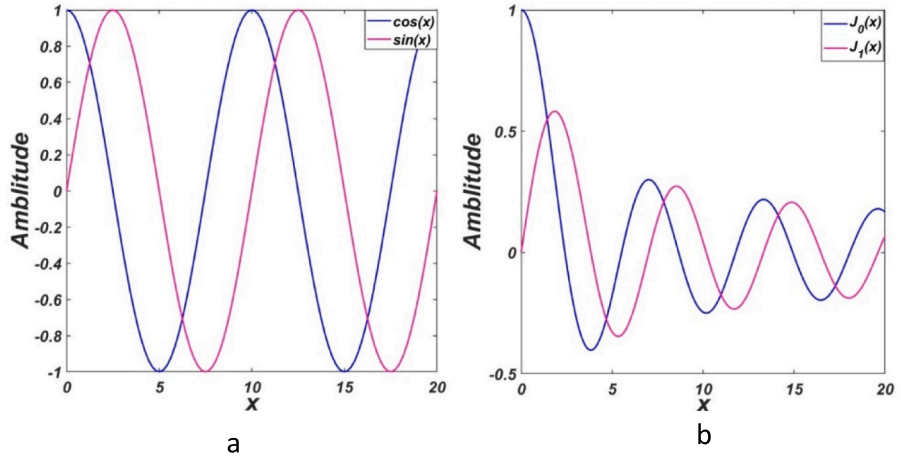


Fig. 4. Plot of basis function (a) bf, (b) fft coefficients.

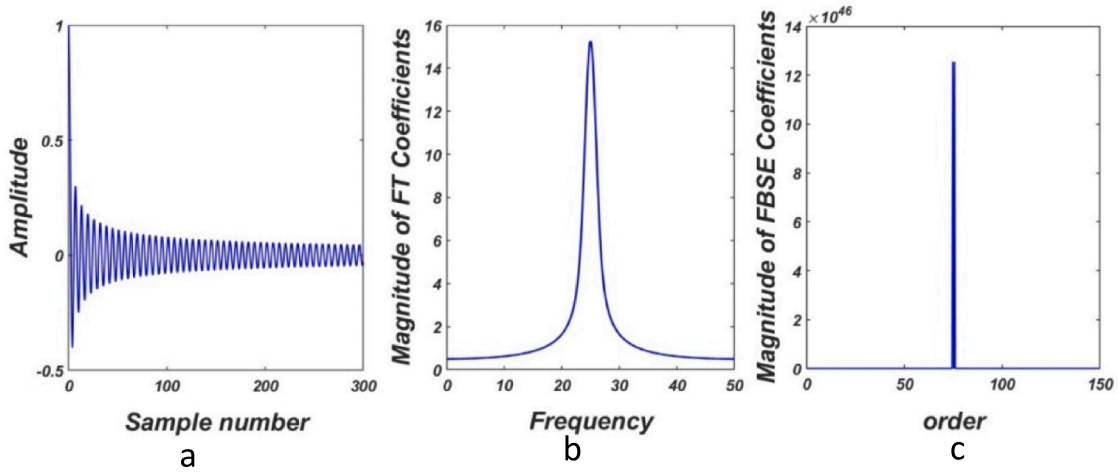


Fig. 5. Plot (a) bf, (b) fft coefficients, (c) bf coefficients.

depression disorder from EEG signals.

In this paper, EEG signals are passed through Fourier-Bessel series expansion. To obtain frequency spectrum for the i^{th} channel of the EEG signal [32,33] and are shown as follows:

$$x(m) = \sum_{i=1}^{N} C_i J_0(\lambda_{i,m}/N), m = 0, 1, \cdots, N-1 \tag{1}$$

where C_i represent the FBSE coefficients of signal x, and it is computed as follows:

$$C_i = \frac{2\sum_{j=0}^{N-1} nx\,(j) J_0(\lambda_i j/N)}{N^2 [J_1(\lambda_i)]^2} \tag{2}$$

Where J_0 and J_1 are the zero order and first order of Bessel functions, λ_i is the positive root of the zero order Bessel function. λ_i are defined in term of corresponding frequencies as follow:

$$\lambda_i \approx \frac{2\pi f_i N}{F_i} \tag{3}$$

Where $\lambda_i \approx \lambda_{i-1} + \pi \approx i\pi$, and F_i is the sampling frequency. Therefore, Eq. (3) can be expressed as

$$i \approx \frac{2 f_i N}{F_i} \tag{4}$$

Based on Eq. (4), i is varied from 1 to N (length of a discrete-time signal) which can cover all the frequency contents of the signal. Fig. 6 shows the FBSE coefficients of healthy and depressed subjects.

The coefficients C_i of FBSE are corresponding to different EEG frequency band of rhythms including $\delta(0.5-4\ Hz)$, $\theta(4-8\ Hz)$, $\alpha(8-13\ Hz)$, $\beta(13-30\ Hz)$, $\gamma(13-30\ Hz)$. These bands are used in Eq. (1) to obtain EEG rhythms, so, Eq. (1) can be written as

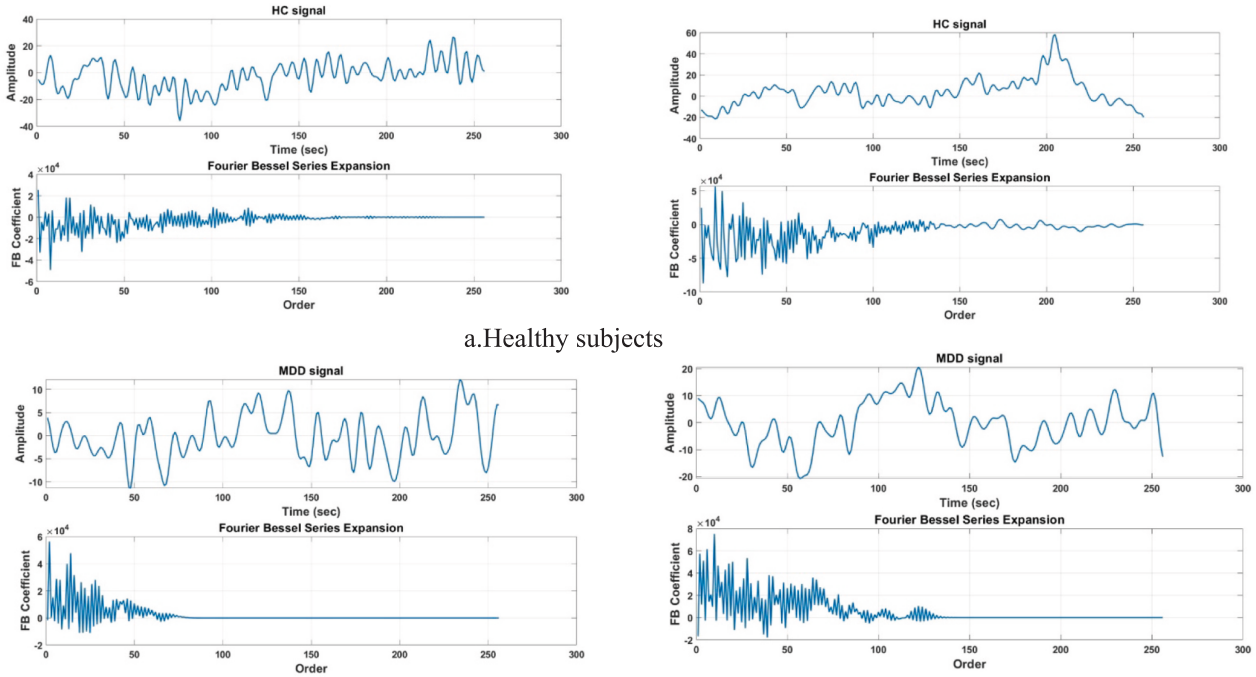


Fig. 6. FBSE coefficients plots for different rhythms for healthy and MDD subjects.

$$x(m)=\sum_{i=\delta_1}^{\delta_2}C_iJ_0\left(\frac{\lambda_{i,m}}{N}\right)+\sum_{i=\theta_1}^{\theta_2}C_iJ_0\left(\frac{\lambda_{i,m}}{N}\right)+\sum_{i=\alpha_1}^{\delta_2}C_iJ_0\left(\frac{\lambda_{i,m}}{N}\right)+\sum_{i=\beta_1}^{\beta_2}C_iJ_0\left(\frac{\lambda_{i,m}}{N}\right)+\sum_{i=\delta_1}^{\delta_2}C_iJ_0\left(\frac{\lambda_{i,m}}{N}\right) \quad (5)$$

In this paper, statistical features and nonlinear features are extracted from the FBSE coefficients.

3.2.1. Statistical features

A total of 12 In this study, different types of features are extracted including statistical, and nonlinear features [18,49]. Each EEG recording is represented by a feature matrix consisting of n rows and m columns. In this section, each category of the proposed features, which is used for classifying EEG signals into MDD and control, is described in detail. Statistical features are extracted from each segment. The statistical features are used to represent EEG signals including *{maximum, range, standard deviation, minimum, mean, mode, median, first quartile, third quartile, Variance, skewness, and kurtosis}*. Table 1 gives a short explanation for each statistical feature.

3.2.2. Nonlinear features

EEG data are stochastic and non-stationary containing some nonlinear patterns. These features lower the ability of linear methods to reveal and describe the hidden EEG signal's characteristics. As a result, many research based EEG signals have employed nonlinear features to study the complexity and dynamics of biomedical signals. In this paper, we employed nonlinear features such as Shannon entropy, dispersion entropy and approximate entropy to extract nonlinear features from EEG signals.

1. Approximate entropy: is a technique used to assess the amount of regularity and unpredictability a signal's based on its current and past amplitude values. It has been proven to be less influenced by noises [40,50]. Suppose a signal X of n datapoint, a new vector is $X(i)$ is obtained using the Eq. (17) on the assumption $\tau=1$. The distance between the pair $X(i)$ and $X(j)(D[X(i),X(j)]\}$ is computed as follows:

$$D[X(i),X(j)\}=max_{r=1,1,\cdots,m}[|X(i+r-1)-X(j+r-1)|]$$

Then, $y_i^m(k)$ is calculated for each $i,j=1,2,\cdots n-m$ using Eq. (6)

$$y_i^m(k)=\frac{No.ofD[X(i),X(j)]\le k}{n-m-1} \quad (6)$$

Where k is a threshold for each $X(i),andX(j)$. Finally, the *approximate entropy is calculated as follow*

$$ApEn=\varphi^m(k)-\varphi^{m+1}(k) \quad (7)$$

Where $\varphi^m(k)$ is defined as

$$\varphi^m(k)=\frac{1}{n-m-1}\sum_{i=1}^{n-m-1}\ln(y_i^m) \quad (8)$$

2. *Shannon entropy*: is used to determine the degree of uncertainty of a random time series [49]. The larger value of Shannon entropy means more randomness and uncertainty of the timeseries Shannon entropy can be defined as:

$$H=-\sum_{i=1}^{n}S_i\ln(S_i) \quad (9)$$

Where S_i is the probability of the i sample in the timeseries value and H is the Shannon entropy.

3. Dispersion entropy: is utilized to measure the complexity or irregularity of time series [50]. A filtered time series u_f ranged from 0 to 1 is produced by filtering analytic time series (u) using a normal cumulative distribution function (NCDF) with mean and standard deviation. The goal of this approach is to better handle outliers. Then, using the function $z^c(j)=round(c^*u_f(j)+0.5)$, which linearly transfers the range [0, 1] to [1,c], u_f is translated into c classes (1 to c). Let $y_m(i)$ be the collection of sequential z^c samples from i to $i+m-1$, where $i=1,2,\ldots,N-m+1$ then $y_m(i)=[z^c(i),z^c(i+1),z^c(i+2),\ldots,z^c(i+m-1)$.

A dispersion pattern is represented by each vector $y_m(i)$. There will be m possible dispersion patterns since each value in y_m can represent one of the c classes that are conceivable. Dispersion entropy can be defined as follows:

$$DispEn(m,c)=-\sum_{i=1}^{c^m}p[y_m(i)]\log(p[y_m(i)])) \quad (10)$$

4. *Zero crossing rate*: It is the number of a signal crosses zero during a certain interval for example, from positive to negative and vice versa divided by the length of the interval. It can be expressed as [37].

$$zcr=\frac{1}{N-1}\sum_{i=1}^{N-1}1_{R<0}(s_is_{i-1}) \quad (11)$$

where $1_{R<0}$ is an indicator function and s is a timeseries with a length N.

3.3. Feature selection methods

Feature selection is a technique by which most relevant characteristics of time series can be chosen from the available features [51,52]. The most influential features can be chosen from the set of all features. As a result, the performance of classification model can be improved as well as the complexity time can be reduced. In this paper, Student's *t*-

Table 1
Mathematical formula of the statistical features.

No.	Feature name	Formula	No.	Feature name	Formula
1	Max	$X_{Max}=\mathrm{Max}[x_n]$	7	Min	$X_{Min}=\min[x_n]$
2	Mean	$X_{Mean}=\frac{1}{n}\sum_1^n x_i$	8	Mode	$X_{Mod}=L+\left(\frac{f_1-f_0}{2f_1-f_2}\right)Xh$
3	Median	$X_{Me}=(\frac{N+1}{2})^{th}$	9	Range	$X_{\mathrm{Range}}=X_{\mathrm{Max}}$-$X_{\mathrm{Min}}$
4	1st Quartile	$X_{Q1}=\frac{1}{4(N+1)}$	10	Standard deviation	$X_{SD}=\sqrt{\sum_{n=1}^N(x_n-AM)}\frac{2}{n-1}$
5	Variation	$X_{Var}=\sum_{n=1}^N(x_n-AM)\frac{2}{N-1}$	11	Skewness	$X_{Ske}=\sum_{n=1}^N(x_n-AM)\frac{3}{(N-1)SD^3}$
6	Kurtosis	$X_{Ku}=\sum_{n=1}^N(x_n-AM)\frac{4}{(N-1)SD^4}$	12	2nd Quartile	$X_{Q2}=\frac{4}{4(N+1)}$

where $X_n=1, 2, 3, \ldots, n$, is a set of data, N is the number of data points in X, *and* AM is the mean of the sample.

test, and Wilcoxon test are used to select the most influential feature selection.

- **Student *t*-test**

The student's *t*-test approach is based on the meaning of two samples. The features with a higher t-value indicate that large differences between two classes. The *t*-test for two class *A* and *B* of size *m* is defined as:

$$t = \frac{m_A - m_B}{\sqrt{\frac{R^2}{n_D} + \frac{R^2}{n_E}}} \tag{12}$$

where, m_A and m_B are the mean values of classes *A* and *B* respectively, *and* R^2 is an estimator of the variance of groups *A* and *B*. R^2 *is* calculated as follow:

$$s^2 = \frac{\sum (x - m_A)^2 + (x - m_B)^2}{n_D + n_E - 2} \tag{13}$$

- **Wilcoxon test**

The second metric used to select the features is Wilcoxon. It is a non-parametric test, and it is used to determine whether the mean of two classes varies or not when they are not normally distributed.

3.4. Domain adaptation

After extraction, the feature matrix is vectorized as feature vector with size *MxM* to fed into a classifier, where *M* refers to channels number, *M* denotes to the features. We suppose the vector features of training and test dataset are stored as $Ds \in RMxM$ and $Dt \in R_{MxM}$. All features can be represented as a matrix $DD = [Ds, Dt] \in RMxM$. In subject independence classification, we hypothesize that each subject's recording contains multi-channels EEG signals that distribute differently across subjects. To deal with such issue, domain adaptation (DA) method is widely used to reduce the dispersion between *Ds* and *Dt* by finding invariant subspace features in the dataset *DD*. In this paper we investigated Maximum Independence Domain Adaptation (MIDA) DA method to construct the invariant subspace features DD' [53,54]. The MIDA find the maximize the independence between *DDtt* and *DDss* using the Hilbert-Schmidt independence criterion. Fig. 7 shows the domain adaptation process.

3.5. Classifiers

The extracted features from EEG signals were sent into various classification models with an aim of obtaining a robust classification rate. In this paper several distinct classification models named LS-SVM, SVM, KNN, k-means and ensembles classifiers were employed to classify EEG features. This section provides a short description of these classifiers [55–59].

1. SVM is an efficient algorithm for binary classification problems. The algorithm establishes a hyperplane between two classes to predict labels using one or more feature samples [55]. The nearest datapoints from each of these classes are placed away from the hyperplane. The nearest points are denoted as support vectors. Suppose a training set with labels: $(x_1, y_1), ..., (x_n, y_n), x_i \in R^d$ *and* $y_i \in (-1, +1)$

where y_i refers to the class label and x_i denotes to the feature vector. The best hyperplane is given as

$$wx^T + b = 0$$

Where *w* is the weight vector, *x* refers to the feature vector, and *b* is the bias. The parameters *w* and *b* should meet the following criteria:

$$wx_i^T + b \geq +1 \ \ if \ \ y_i = 1$$

$$wx_i^T + b \leq -1 \ \ if \ \ y_i = -1$$

Finding the best values of *w* and *b* is the main goal of the training phase. This allows the hyperplane to separate the input feature vector as well as maximize the margin $1/||w||^2$.

2. LS-SVM also used as a classifier in this paper. It is an enhanced version of SVM. The LS-SVM creates a hyperplane based on input vectors to distinguish the data classes. The input feature vector is mapped into hyperplane under the concept of kernels. Choosing the right kernel function is important to obtain the desired classification results. In this paper, the kernel function with parameters $\sigma, and \beta$ are selected carefully.
3. KNN is a semi-supervised learning algorithm that clusters datapoints based on the majority of their neighbours. The algorithm accomplishes the classification process based on two steps: find the number of nearest neighbours based on a distance metric such as Euclidean distance and classify the datapoints based on the results of the first step. It selects nearest *k* datapoints from the training set, then it takes the majority vote of their class.
4. Bagging is used to obtain the training subsets for training a base learner. At each iteration, *S* samples are chosen randomly with replacement from the original training set to learn each individual classifier. A uniform majority voting technique is used to aggregate the output of classifiers to predict the test sample of an ensemble. Bagging algorithm is shown in Algorithm 1.

Algorithm 1: Bagging ensemble pseudocode

Input: dataset X= $\{(x_1, y_1), (x_2, y_2), \cdots..(x_n, y_n)\}$
Base learning algorithm ξ
Number of base learning *k*
Process
1. For I = 1…, *k*;
2. $h_I = \xi(X, X_{bc})\%X_{bc}$ *is the bootstrap distribtution*
3. *end*
4. Output: H(x)=$arg_{y \in Y} max \sum_{I=1}^{k} ||(X_I(x) = y)$

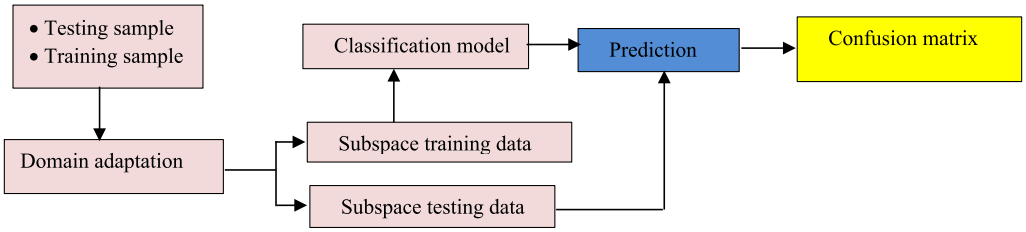


Fig. 7. DA method.

5. Boosting ensemble multiple models that complete each other. The boosting is similar to bagging ensemble as they both use the voting strategy in cases of classification and averaging to obtain the correct prediction. However, the boosting ensemble uses iteration to create new models which are influenced by those that are previously built. The Boosting algorithm is shown in Algorithm 2.

Algorithm 2: Boosting ensemble pseudocode

Input: dataset X: sample distribution
Base learning algorithm ξ
Number of base learning k
Process
a. $X_1 = X$
5. For I = 1, …..,k
6. $h_I = \xi(X_I)$; Train a weak leaner from sample X_I
7. $E_I = P_{n-x_I} h_I$ evaluate the error
8. X_{1+1} = Adjust distribution (X_I, E_I)
9. *output* $X(n)$ = *combine output* $(\{h_1\ (n), \ldots\ldots,\ h_k\ (n)\}\ )$

6. Stacking

Stacking is an ensemble technique which employs a Meta classifier to combine various classification algorithms in one model. Each classifier passes its predictions results to the model in the layer above, and the model in the topmost layer takes decisions depending on the models below. Several layers are stacked one on top of the other. The initial dataset's input features are fed into the bottom layer models. The prediction is made by the top layer model using the bottom layer's output. The original data is fed into various separate models during stacking. The weights of each model are then computed, and the Meta classifier is then used to estimate the input and output of each model. The models that perform the best are chosen, while the rest are eliminated.

7. Decision tree

The decision tree (DT) is a supervised classification model. It divided a complex problem into several sub-problems. This process is repeated to generate a tree. Each leaf node in DT gets a class label. Nonterminal node refers to the root node and other internal nodes. It contain features testing conditions to distinguish records with different characteristics [60].

8. **Gradient Boosting**: model has been used widely in time series classification. It has three main components named loss function, weak leaner, and additive model. The loss function is used to estimate the model performance by making predictions with the given data. The additive model refers to the iterative and sequential method of adding weak learners one at a time. After each iteration, we should be closer to our final model [61].

3.6. Performance evaluation

10-fold cross-validation and leave one subject out were used to fairly evaluate the performance of the proposed model. For the 10-fold cross-validation procedure, the dataset was divided into 10 folds, then, 9 folds were randomly used as the training set, and the remaining fold was used as the testing set. This procedure was repeated 10 times until each fold was used as the testing set. For further investigating the performance of the proposed model, the leave-one-subject-out (LOSO) cross-validation strategy was also adopted to assess the proposed model in which the data of one subject was used for the testing phase, while the remaining subject's data was utilized for the training purposes. The other metrics used in the evaluation phase were Accuracy, Sensitivity, Specificity, Precision, F-score are employed in the evaluation phase. The values of these metrics were calculated based on confusion matrix parameters including TP, FP, TN, and FN [62,63][48,49].

- $\text{Accuracy(ACC)} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$
- $\text{Sensitivity(SEN)} = \frac{\text{TP}}{\text{TP+FN}}$
- $\text{Specificity(SPE)} = \frac{\text{TN}}{\text{TN+FP}}$
- $\text{Precision(Prec)} = \frac{\text{TP}}{\text{TP+FP}}$
- $\text{F}-\text{score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision+Sensitivity}}$

TP: refers to the total number of MDD patients who have been correctly recognized. TN: denotes to the total number of participants who were appropriately classified as normal subjects. FN: is the total number of MDD patients who were mistakenly classified as normal subjects. FP: refers to the number of all normal participants who were mistakenly classified as MDD patients [54].

- **Confidence interval** is used to evaluate the proposed model. The confidence interval formula is expressed as:

$$ACCtest \pm z \times SE \tag{16}$$

where z is sample size, and the SE is the standard error which is calculated using the following formula

$$SE = \sqrt{\frac{1}{b-1}} \sum_{i=1}^{b} (Acc_i - Acc_{averg}) \tag{17}$$

4. Experimental results

In this paper, the EEG dataset was contributed by Mumtaz et al. [22]. After that the EEG signals were segmented into intervals, then the FBSE was applied to each segment to acquire the FBSE coefficients. Two types of EEG features were extracted including linear, and nonlinear feature. The extracted features were tested using paramedic and non-parametric tests to select the most influential feature set. The domain adaptation method was employed to improve the classification accuracy. The selected features were sent to several machine learning algorithms. The scikit-learn software was used to partition the collected data into the training set, the validation, and the testing set. All experiments were designed and conducted on a computer with 8G Ram, Intel (R) Core (TM) i7-6600U CPU @ 2.60 GHz. MATLAB R2021a software was used to implement the proposed model..

4.1. Feature selection

To enhance the classification rate of the proposed model, the redundant features were removed and the most influential once were selected to classify MDD from healthy subjects. The results of feature analysis were reported in Tables 2-4. We found that that some statistical and nonlinear features were rejected by a t-test, and not significance at $\alpha = 0.05$. Based on the t-test results the statistical features name max, *range, std,* min, *mean, median, skewness, kurtosis* and 3 linear features including approximation entropy, Shannon entropy, and zero were accepted by the t-test. Tables 2, 3 and 3, the results of t-test and Wilcoxon test for the statistical features, and nonlinear features.

In this paper, the extracted features were also tested based on the Wilcoxon test to select the most appropriate features that represented the EEG dataset test. Based on Wilcoxon test, the features were accepted when their statistic values were less than or equal to 0.05. However, the features with statistical values exceeding 0.05 were rejected and they were considered not significant. The analysis results were reported in Tables 3, and 4. The findings in Table 3, and 4 show that statistical features including {max, *range, std,* min, *mean, median, skewness, kurtosis}* and 3 linear features named *{approximation entropy, Shannon*

Table 2
Statistical feature assessment using *t*-test.

Channel	Max	Min	Range	Std	Mean	Median	skewness	Kurtosis	First Q	Second Q	mode	Third Q
1	0.0211	0.0377	0.0176	0.0332	0.0188	0.0379	0.0421	0.0157	0.5996	0.6219	0.6190	0.5892
2	0.0520	0.0487	0.1020	0.0931	0.1420	0.0932	0.0828	0.0709	0.6210	0.6912	0.5911	0.7271
3	0.0230	0.0054	0.0072	0.0391	0.0061	0.0270	0.0316	0.00091	0.6290	0.7043	0.5929	0.6210
4	0.0130	0.0121	0.0405	0.0581	0.0539	0.1301	0.0753	0.0824	0.4982	0.4920	0.586	0.5871
5	0.0621	0.0572	0.0580	0.0437	0.1842	0.0107	0.0761	0.0176	0.5621	0.6948	0.5742	0.7291
6	0.2085	0.1908	0.2690	0.1841	0.0095	0.0621	0.1090	0.1705	0.5921	0.6312	0.6001	0.6761
7	0.0103	0.0209	0.0111	0.0102	0.1290	0.0329	0.0130	0.0190	0.6291	0.7001	0.7212	0.6988
8	0.1302	0.1590	0.1040	0.1901	0.0931	0.0826	0.1430	0.0991	0.5929	0.8200	0.7866	0.6690
9	0.2301	0.2082	0.2100	0.1941	0.1927	0.1051	0.1927	0.1791	0.7761	0.7981	0.7221	0.6200
10	0.0111	0.0219	0.0311	0.0932	0.0830	0.0591	0.0071	0.0091	0.6442	0.5221	0.7432	0.6211
11	0.0123	0.0902	0.0038	0.0019	0.0059	0.0214	0.0140	0.0521	0.5721	0.5520	0.5700	0.6701
12	0.2721	0.2971	0.1934	0.1841	0.2302	0.2001	0.1801	0.1351	0.7009	0.7208	0.7991	0.7700
13	0.1350	0.0961	0.1720	0.1933	0.2911	0.0965	0.1981	0.1191	0.6844	0.6666	0.6939	0.6111
14	0.0753	0.0861	0.0810	0.0779	0.0601	0.0828	0.0681	0.0892	0.5900	0.6777	0.7910	0.6400
15	0.0210	0.0200	0.0393	0.0400	0.0109	0.0921	0.0391	0.0202	0.5744	0.6531	0.8300	0.8005
16	0.0312	0.0495	0.0389	0.0391	0.0910	0.0581	0.0471	0.0619	0.6790	0.7870	0.6410	0.6431
17	0.0256	0.0367	0.0329	0.0459	0.0447	0.0531	0.0712	0.0390	0.7651	0.7033	0.6193	0.7310
18	0.0237	0.0197	0.0201	0.0620	0.0309	0.0410	0.0267	0.0495	0.6591	0.6500	0.7492	0.8331
19	0.0035	0.0021	0.0101	0.0051	0.0049	0.0157	0.0189	0.0310	0.7780	0.6195	0.7401	0.9301

Table 3
Statistical feature assessment using Wilcoxon.

Channel	Max	Min	Range	Std	Mean	Median	skewness	Kurtosis	First Q	Second Q	mode	Third Q
1	0.0064	0.0010	0.0030	0.0012	0.0059	0.0078	0.0034	0.0051	0.7130	0.7721	0.7943	0.6940
2	0.0084	0.0078	0.0083	0.0058	0.0092	0.0093	0.0037	0.0074	0.6744	0.6808	0.6541	0.6661
3	0.0038	0.0011	0.0063	0.0023	0.0087	0.0049	0.0035	0.0082	0.6964	0.8120	0.7871	0.6219
4	0.1002	0.1039	0.1209	0.1912	0.1781	0.1365	0.1491	0.1650	0.5940	0.6319	0.6001	0.6233
5	0.0050	0.0019	0.1309	0.0191	0.1907	0.0982	0.0850	0.0841	0.6300	0.6711	0.6848	0.7111
6	0.0199	0.0091	0.0709	0.0690	0.0510	0.0330	0.0405	0.0506	0.8309	0.7942	0.6001	0.6761
7	0.0065	00.019	0.0201	0.0138	0.0211	0.0052	0.0710	0.0103	0.7091	0.8309	0.6186	0.7430
8	0.0301	0.0251	0.0329	0.0343	0.0394	0.0194	0.0821	0.0371	0.5940	0.6481	0.7391	0.6924
9	0.0209	0.0199	0.0490	0.0327	0.0241	0.0299	0.0384	0.0470	0.6652	0.7009	0.6999	0.7944
10	0.0302	0.0261	0.0353	0.0539	0.0524	0.0500	0,0385	0.0545	0.5987	0.6440	0.6164	0.8038
11	0.0403	0.0401	0.0431	0.0465	0.0472	0.0440	0.0432	0.3996	0.7491	0.7050	0.7009	0.6091
12	0.0621	0.0660	0.0692	0.0629	0.0623	0.0607	0.0509	0.0666	0.5911	0.6581	0.6710	0.6541
13	0.1030	0.1052	0.1590	0.1401	0.1901	0.1351	0.1272	0.1405	0.6757	0.7430	0.7002	0.7033
14	0.0609	0.0503	0.0291	0.0767	0.0549	0.0275	0.0494	0.0840	0.7917	0.7644	0.6908	0.6967
15	0.1409	0.1341	0.1701	0.2003	0.2041	0.1029	0.2105	0.2010	0.7461	0.6540	0.6416	0.6954
16	0.1650	0.2081	0.2491	0.2750	0.2810	0.2691	0.2603	0.1709	0.6835	0.6900	0.6700	0.7299
17	0.0951	0.0895	0.0826	0.0830	0.0991	0.0305	0.0876	0.0209	0.6020	0.6009	0.6446	0.6500
18	0.2071	0.1950	0.1521	0.1961	0.1749	0.1834	0.1791	0.1841	0.6811	0.6591	0.6008	0.6666
19	0.0396	0.0329	0.1040	0.1532	0.1311	0.1755	0.1067	0.0999	0.7271	0.7200	0.7000	0.8001

Table 4
Nonlinear assessment using wilcoxon, and*t*-test.

	t-test				Wilcoxon test			
Channel	Appr. entropy	Dep.entropy	Shan.entropy	Zero crossing	Appr, entropy	Dep.entropy	Shan.entropy	Zero crossing
1	0.1087	0.6301	0.0780	0.1502	0.1042	0.5831	0.0401	0.1050
2	0.1921	0.7038	0.1009	0.1100	0.1536	0.8402	0.1991	0.1785
3	0.0941	0.8503	0.0785	0.0886	0.0391	0.7901	0.1364	0.1489
4	0.0396	0.8331	0.0434	0.0392	0.0412	0.5982	0.0924	0.0421
5	0.0871	0.7020	0.3329	0.0288	0.0814	0.6088	0.1129	0.0319
6	0.0924	0.6120	0.0088	0.1129	0.0319	0.5329	0.0209	0.0581
7	0.0510	0.7261	0.3329	0.0209	0.0581	0.6099	0.0980	0.0702
8	0.0303	0.7210	0.0099	0.0550	0.0702	0.6050	0.0559	0.0621
9	0.0610	0.6295	0.0050	0.0039	0.0621	0.7039	0.0713	0.0597
10	0.0201	0.6021	0.1039	0.0773	0.0597	0.5973	0.0621	0.0319
11	0.0681	0.7565	0.4773	0.0321	0.0319	0.7129	0.0764	0.0312
12	0.0588	0.8031	0.3329	0.0438	0.0312	0.6978	0.0129	0.0213
13	0.0524	0.8028	0.0088	0.0129	0.0213	0.7109	0.0289	0.0809
14	0.0439	0.8036	0.3329	0.0289	0.0809	0.6169	0.0542	0.0602
15	0.0376	0.8494	0.0099	0.0950	0.0602	0.6021	0.0431	0.0832
16	0.0481	0.7073	0.0050	0.1031	0.0832	0.6943	0.3211	0.0321
17	0.0392	0.8003	0.1039	0.4100	0.0321	0.6871	0.0451	0.0430
18	0.0437	0.6007	0.4773	0.0129	0.0430	0.6429	0.0333	0.0821
19	0.0352	0.7020	0.3329	0.0101	0.0821	0.7891	0.0431	0.0319

entropy, and zero crossing} passed the test which support our finding in Table 2.

Further investigation was made based on box plot to examine the effectiveness of the extracted features on the identification of depression. Each box plot is divided into three parts, the upper part which represents 75% percentile of the box, and the lowest part which represents 25% percentile of the box and the middle line which refers to the median 50% percentile. The lines starting from the top of box to the bottom of box refer to the highest and lowest values. It was noticed that that the statistical features {*mode, first, second and third quartile*}, and nonlinear feature *{depression}* entropy, cannot be used as the key features to classify MD subjects from healthy participants. However, the statistical features {max*, range, std,* min*, mean, median, skewness, kurtosis}*, and nonlinear features *{approximation entropy, Shannon entropy, and zero crossing}* showed a high influence in the differentiating between MDD and healthy subjects. The obtained results by boxplots support our results in Tables 2,3,and 4. Fig. 8 shows an example of boxplots of some linear and no-linear features.

4.2. Results based on statistical features

The selected feature set was evaluated using domain adaptation method. The Maximum Independence Domain Adaptation (MIDA) was employed to construct the invariant subspace features set. Then, the final features set was sent to several classification methods. In this section, we discussed the efficiency of the statistical features and nonlinear features as well as the combination of statistical and nonlinear features set. First, the classification results based on statistical features were presented in Table 5. The eight statistical features {max*, range, std,* min*, mean, median, skewness, kurtosis*} were fed to different SVM, LS-SVM, KNN, decision tree, Gradient boosting, ransom forest, Bagged ensemble, boosted ensemble, and stacked ensemble. Stack ensemble, Gradient boosting and Boosted ensemble showed the high classification rates with accuracies of 0.93%, 0.925% and 0.92% respectively. However, KNN scored the lowest classification accuracy among the seven classification models.

Table 5
Classification results based on statistical features.

Classifier	Accuracy	Sensitivity	Specificity	Precision	f-score
SVM	0.8977	0.8891	0.9068	0.9132	0.8997
LS-SVM	0.8566	0.8565	0.8568	0.8642	0.8605
KNN	0.8173	0.8030	0.8050	0.8102	0.8047
Decision tree	0.8643	0.8543	0.8532	0.8532	0.8502
Random forest	0.8394	0.8237	0.8300	0.8103	0.8239
Gradient boosting	0.9254	0.9212	0.9200	0.9210	0.9201
Bagged Ensemble	0.9134	0.9097	0.9182	0.9012	0.9028
Boosted ensemble	0.9245	0.9120	0.9148	0.9203	0.9183
Stacked ensemble	**0.9310**	**0.9274**	**0.9227**	**0.9274**	**0.9273**

4.3. Results based on nonlinear features

In this experiment, 3 linear features including {*approximation entropy, Shannon entropy, and zero}* were sent to SVM, LS-SVM, KNN,

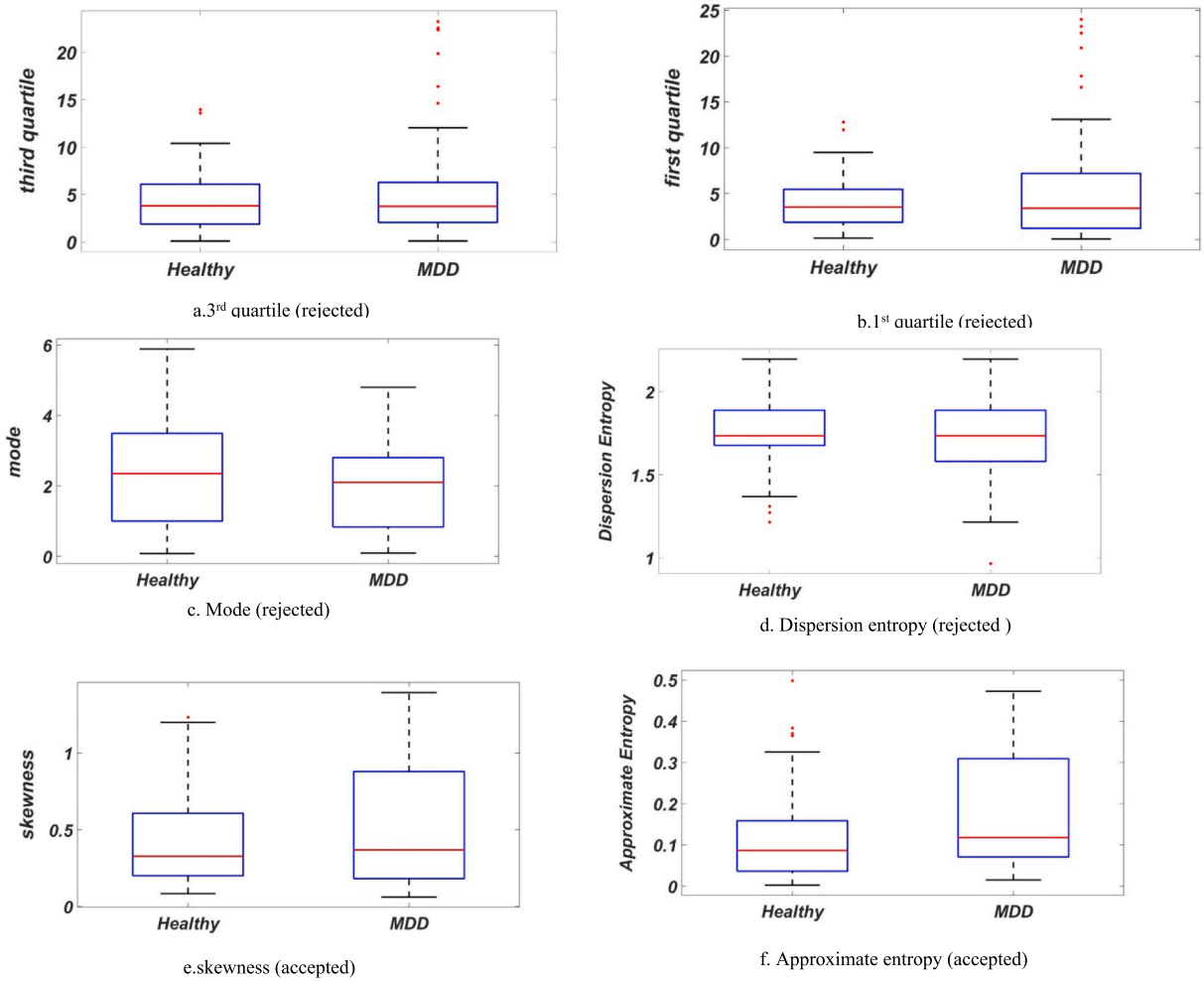


Fig. 8. An example of boxplots of some statistical and nonlinear features.

Table 6
Classification results based on statistical features.

Classifier	Accuracy	Sensitivity	Specificity	Precision	f-score
SVM	0.8735	0.8391	0.9030	0.8808	0.8595
LS-SVM	0.8517	0.8297	0.8706	0.8457	0.8376
KNN	0.8445	0.8580	0.8329	0.8144	0.8356
Decision tree	0.84321	0.8412	0.8362	0.8421	0.8496
Random forest	0.7951	0.7886	0.8005	0.7716	0.7800
Gradient boosting	0.8420	0.8400	0.8392	0.8420	0.8410
Bagged Ensemble	0.8488	0.8139	0.8787	0.8515	0.8323
Boosted ensemble	0.8576	0.8265	0.8841	0.8590	0.8424
Stacked ensemble	**0.8832**	**0.8865**	**0.8860**	**0.8806**	**0.8984**

ransom forest, decision tree, Gradient boosting, Bagged ensemble, boosted ensemble, and stacked ensemble. Table 6 reports the classification accuracy based on nonlinear features. The stacked ensemble was obtained the highest classification accuracy of 0.8832% while the SVM scored the second highest accuracy of 0.8735. Our finding showed that the combination of non-linear features did not perform well as the statistical features results in Table 5. The classification results with nonlinear features did not exceed 0.88% with all classifiers. However, the obtained classification results compared with the number of nonlinear features can be considered acceptable where only 3 features were used in for the classification.

4.4. Results based on a combination of statistical and nonlinear features

In this experiment, the statistical features and the nonlinear features were integrated and used to classify EEG signals into MD and healthy subjects. Table 7 reports the classification results based on different classification models. It can be noticed that the proposed technique provides the highest accuracy with stacked ensemble, while the lowest accuracy was achieved by KNN. It can be observed that the combination of statistical features and nonlinear features improved the classification compared by 0.6 compared with the statistical features and by 0.11 compared with the nonlinear features. In addition, the proposed model obtained high accuracies across all classifies compared with the results in Tables 5 and 6.

5. Discussion

In this section, the main findings are summarised as follows:

1. The proposed model was designed to demonstrate the superiorly of the FBSE coupled with domain adaptation technique in the depression detection from EEG signals. Our finding showed that FBSE gave good results in the detection of depression. It was noticed that integrating the statistical features with nonlinear features can improve the classification accuracy. The promising classification results shown by the proposed model demonstrated that the proposed model can be implemented and extended to monitor other brain disorder such as Alzheimer and sleep disorder. Although the obtained results showed that the proposed model can obtain satisfactory performance, it has certain limitations. Firstly, one public dataset with limited number of subjects were recruited for this study. A larger number of subjects suffering from depression disorder are required in our future work to evaluate the generalization of the proposed method. Secondly in this study, we tested one domain adaption method named investigated maximum independence domain adaptation, different domain adaptation technique including supervise and unsupervised are important to be tested in our future work to improve the accuracy of depression detection. Thirdly, the proposed model did not test in case of poor signal quality to expand depression detection outside of labs and into further in real-world settings.
2. To measure the efficiency and superiority of the proposed model, it is compared with the previous approaches for diagnosing MD from EEG signals. Table 8 presents the comparison results. For fair comparisons, most the previous studies in Table 8 were implemented on the Mumtaz et al. (Mumtaz et al., 2017) EEG dataset. Based on the results in Table 8, we can noticed that the proposed model scored the highest classification accuracy among state-of-the-art and it outperformed the other state-of-the-art approaches. The main contribution of the proposed model compared to the previous studies is that the features of EEG extracted by FBSE were studied using domain adaptation and features selection methods while the previous studies did not utilize the domain adaptation method MDD classification. In addition, EEG channels were investigated, and it was found that not all channels were efficient to classify EEG signals. As a result, in this study, channel P4-LE and F&-LE were excluded in our experiments as they provided low accuracies.

The features were evaluated using two statistical metrics named student *t*-test, and Wilcoxon test. Our finding showed that that integrating all feature sets provides a high classification performance compared with previous studies of MDD classification based on EEG signals. It is noteworthy that this paper utilized a set of classification models including individual classifiers and ensemble classifiers, and the obtained results showed that the ensemble classifiers performed better than individual classifiers.

3. Subject-independent evaluation: Subject independent evaluation employs the leave-one-subject-out (LOSO) cross validation method to examine MDD classification efficiency. In the LOSO cross-validation strategy, the EEG data of one subject was utilized for testing, while the remaining subjects EEG data were used for training. EEG subjects were separated into three groups for training, validation, and testing while ensuring that the same subject did not appear in two groups. During the division, at least one subject from each MDD and healthy control category must be included in each group, so that the subjects are divided into groups in a systematic manner. Fig. 9 report the classification results based on LOSO strategy. It can notice that the proposed model obtained a

Table 7
Classification results based on combination of statistical and nonlinear features.

Classifier	Accuracy	Sensitivity	Specificity	Precision	f-score
SVM	0.9556	0.9130	0.9243	0.9213	0.9345
LS-SVM	0.9744	0.9730	0.9727	0.7677	0.9571
KNN	0.8667	0.8696	0.8636	0.8696	0.8696
Decision tree	0.9021	0.8524	0.9010	0.9001	0.9021
Random forest	0.9111	0.9130	0.9091	0.9130	0.9130
Gradient boosting	0.9632	0.9632	0.9685	0.9645	0.9632
Bagged Ensemble	0.9509	0.9530	0.9582	0.9500	0.9550
Boosted ensemble	0.9533	0.94696	0.9186	0.9321	0.9302
Stacked ensemble	**0.9911**	**0.9800**	**0.9891**	**0.9830**	**0.9830**

Table 8
Comparisons among the proposed model and previous studies.

Authors	Technique	Results
Mumtaz et al., [60]	Spectral features based on SVM, LR, NB	Acc = 85.5%
Mumtaz et al., [37	Wavelet based on LR	Acc = 76.9%
Mahato et al., [64]	Spectral features coupled with MILPNN, RBFN, QDA, and LDA	Acc = 82%
Acharya et al., [61]	CNN	Acc = 94%
Mahato et al., [65]	Multi-Cluster Feature Selection, SVM	ACC = 88.33%
Ca et al., [60]	Nonlinear feature, KNN, CT, SVM	ACC = 86.98%
Current study	FBSE	Acc = 99%

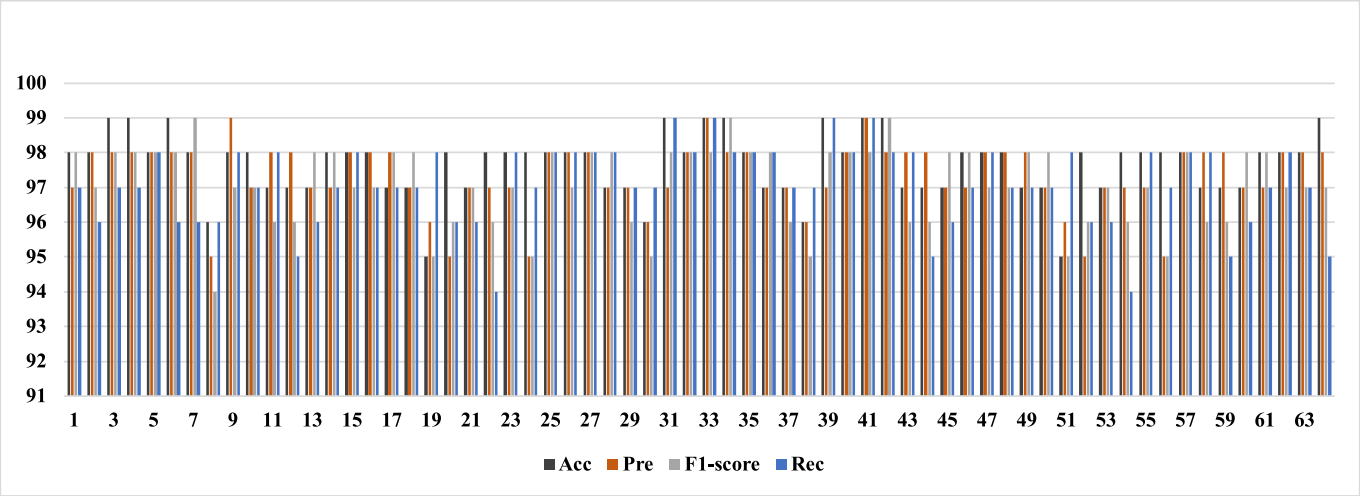


Fig. 9. Classification results based on LOSO.

consistence accuracy and there was not a high fluctuation in the results cross subjects.

4. The effect of DA on classification results was also evaluated. Our findings showed that DA technique significantly improved the classification results, compared to the baseline accuracy 89.26%. Table 9 reports the classification results with and without applying DA method. From the obtained results we can notice that the classification accuracy was significantly improved cross all classifiers after applying DA method.
5. Confidence intervals were also employed to evaluate the proposed model. Confidence intervals for each classifier were reported in Tables 10. The success of each confidence interval in obtaining the true accuracy was titled as 'Success'in the column header.
6. The accuracy of each EEG channel; was also investigated. The features were extracted from each channel and forwarded to the stacked ensemble. Fig. 10 illustrated the classification results based on EEG channel. The results showed that channel P4-LE and F&-LE provide low rate accuracies. In this paper, these two channel were excluded and not considered in the MDD detection.

6. Conclusions

In this paper, we proposed an intelligent model for MDD detection. Fourier-Bessel series expansion (FBSE) was utilised to study EEG signals during depression. Each EEG segment was passed through FBSE and two types of features were extracted including statistical and nonlinear features. These features were investigated using statistical metrics and domain adaptation method to find out the most powerful features. Both subject dependent experiments using 10-fold cross-validation and subject independent experiment using leave-one-subject cross-validation were conducted on the public MDD dataset. On the one hand, the experimental results demonstrated that the using a combination of statistical and nonlinear features gained provided a high classification rate. The results proved that the proposed model could improve the performance of MDD classification making small step toward application in actual situation.

Table 10
Confidence intervals for each classification model.

Classification model	95% Confidence Interval	Success
		✓
SVM	0.95254	✓
LS-SVM	0.90121	✓
KNN	0.89351	✓
Decision tree	0.90212	✓
Random forest	0.9410	✓
Gradient boosting	0.9540	✓
Bagged ensemble	0.9801	✓
Boosted ensemble	0.9621	✓

CRediT authorship contribution statement

Hadeer Mohammed: Conceptualization, Methodology, Software, Validation, Data curation, Formal analysis. **Mohammed Diykh:** Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table 9
Classification results based on statistical features.

	With DA			Without DA		
Classifier	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
SVM	0.9556	0.9130	0.9243	0.8513	0.8545	0.8666
LS-SVM	0.9744	0.9730	0.9727	0.8477	0.8571	0.8421
KNN	0.8667	0.8696	0.8636	0.8196	0.8196	0.8075
Decision tree	0.9021	0.9024	0.9010	0.8121	0.8021	0.8061
Random forest	0.9111	0.9130	0.9091	0.8630	0.8630	0.8543
Gradient boosting	0.9632	0.9632	0.9685	0.8421	0.8430	0.8491
Bagged Ensemble	0.9509	0.9530	0.9582	0.8700	0.8750	0.8543
Boosted ensemble	0.9533	0.94696	0.9186	0.8721	0.8702	0.8650
Stacked ensemble	**0.9911**	**0.9800**	**0.9891**	0.8930	0.8830	0.8843

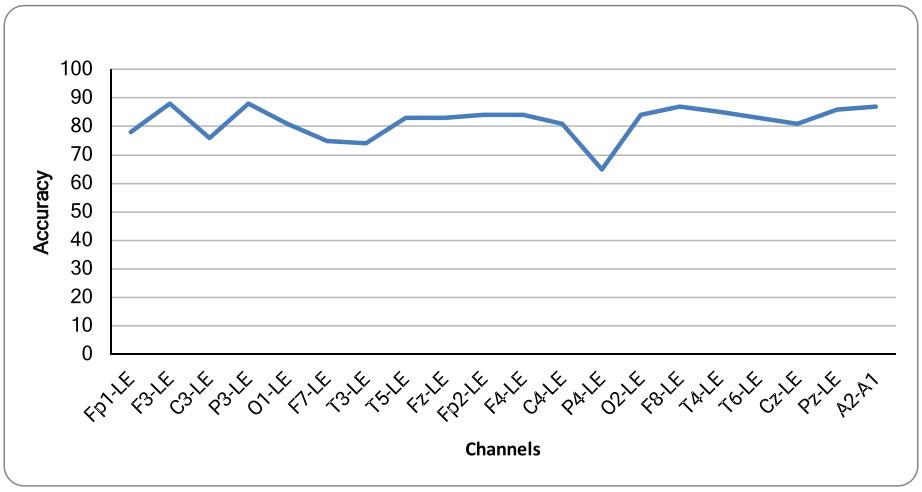


Fig. 10. Classification results based on EEG channels.

Data availability

The authors do not have permission to share data.

References

[1] N. Jaworska, et al., Leveraging machine learning approaches for predicting antidepressant treatment response using electroencephalography (EEG) and clinical data, Front. Psych. 9 (2019) 768.

[2] W. Mumtaz, et al., A wavelet-based technique to predict treatment outcome for major depressive disorder, PLoS One 12 (2) (2017) e0171409.

[3] B. Zhang, et al., Brain functional networks based on resting-state EEG data for major depressive disorder analysis and classification, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2020) 215–229.

[4] H. Cai, Z. Qu, Z. Li, Y. Zhang, X. Hu, B. Hu, Feature-level fusion approaches based on multimodal EEG data for depression recognition, Information Fusion 59 (2020) 127–138.

[5] S. Mahato, S. Paul, Classification of depression patients and normal subjects based on electroencephalogram (EEG) signal using alpha power and theta asymmetry, J. Med. Syst. 44 (2020) 1–8.

[6] Y. Zhu, et al., Classifying major depressive disorder using fNIRS during motor rehabilitation, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (4) (2020) 961–969.

[7] J. Li, et al., Altered Brain Dynamics and Their Ability for Major Depression Detection using EEG Microstates Analysis, IEEE Trans. Affect. Comput. (2021).

[8] C. Jiang, et al., Enhancing EEG-based classification of depression patients using spatial information, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 566–575.

[9] A. Dev, et al., Exploration of EEG-based depression biomarkers identification techniques and their applications: A systematic review, IEEE Access (2022).

[10] J. Elwood, E. Murray, A. Bell, M. Sinclair, W.G. Kernohan, J. Stockdale, A systematic review investigating if genetic or epigenetic markers are associated with postnatal depression, J. Affect. Disord. 253 (2019) 51–62.

[11] Z. Jiang, S. Harati, A. Crowell, H.S. Mayberg, S. Nemati, G.D. Clifford, Classifying major depressive disorder and response to deep brain stimulation over time by analyzing facial expressions, IEEE Trans. Biomed. Eng. 68 (2) (2020) 664–672.

[12] Y. Zhu, J.K. Jayagopal, R.K. Mehta, M. Erraguntla, J. Nuamah, A.D. McDonald, H. Taylor, S.H. Chang, Classifying major depressive disorder using fNIRS during motor rehabilitation, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (4) (2020) 961–969.

[13] S. Dessai, S.S. Usgaonkar, May. Depression Detection on Social Media Using Text Mining, in: 2022 3rd International Conference for Emerging Technology (INCET), IEEE, 2022, pp. 1–4.

[14] S. Harati, A. Crowell, H. Mayberg, S. Nemati, Depression severity classification from speech emotion, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018, pp. 5763–5766.

[15] C. Jiang, Y. Li, Y. Tang, C. Guan, Enhancing EEG-based classification of depression patients using spatial information, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 566–575.

[16] A. Dev, N. Roy, M.K. Islam, C. Biswas, H.U. Ahmed, M.A. Amin, F. Sarker, R. Vaidyanathan, K.A. Mamun, Exploration of EEG-based depression biomarkers identification techniques and their applications: A systematic review, IEEE Access, 2022.

[17] M. Nikravan, E. Ebrahimzadeh, Time-frequency analysis in EEG for the Treatment of Major Depressive Disorder Using rTMS, in: *2021 Asia-Pacific International Symposium on Electromagnetic Compatibility (APEMC)*, IEEE, 2021, September., pp. 1–3.

[18] M. Saeedi, A. Saeedi, A. Maghsoudi, Major depressive disorder assessment via enhanced k-nearest neighbor method and EEG signals, Phys. Eng. Sci. Med. 43 (3) (2020) 1007–1018.

[19] H. Akbari, et al., Depression recognition based on the reconstruction of phase space of EEG signals and geometrical features, Appl. Acoust. 179 (2021), 108078.

[20] M. Bachmann, et al., Methods for classifying depression in single channel EEG using linear and nonlinear signal analysis, Comput. Methods Programs Biomed. 155 (2018) 11–17.

[21] M. Sharma, et al., An automated diagnosis of depression using three-channel bandwidth-duration localized wavelet filter bank with EEG signals, Cogn. Syst. Res. 52 (2018) 508–520.

[22] S. Mahato, S. Paul, Classification of depression patients and normal subjects based on electroencephalogram (EEG) signal using alpha power and theta asymmetry, J. Med. Syst. 44 (1) (2020) 1–8.

[23] A. Saeedi, et al., Major depressive disorder diagnosis based on effective connectivity in EEG signals: A convolutional neural network and long short-term memory approach, Cogn. Neurodyn. 15 (2) (2021) 239–252.

[24] Z. Jiang, et al., Classifying major depressive disorder and response to deep brain stimulation over time by analyzing facial expressions, IEEE Trans. Biomed. Eng. 68 (2) (2020) 664–672.

[25] XinWang Song, et al., LSDD-EEGNet: An efficient end-to-end framework for EEG-based depression detection, Biomed. Signal Process. Control 75 (2022), 103612.

[26] R.A. Movahed, et al., A major depressive disorder classification framework based on EEG signals using statistical, spectral, wavelet, functional connectivity, and nonlinear analysis, J. Neurosci. Methods 358 (2021), 109209.

[27] Sahar Harati, et al., Depression severity classification from speech emotion. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018.

[28] Mehran Nikravan, Elias Ebrahimzadeh, Time-frequency analysis in EEG for the Treatment of Major Depressive Disorder Using rTMS. 2021 Asia-Pacific International Symposium on Electromagnetic Compatibility (APEMC), IEEE, 2021.

[29] Sukanya Dessai, Soniya Shakil Usgaonkar, Depression Detection on Social Media Using Text Mining. 2022 3rd International Conference for Emerging Technology (INCET), IEEE, 2022.

[30] S.M. Alghowinem, et al., Interpretation of depression detection models via feature selection methods, IEEE Trans. Affect. Comput. (2020).

[31] P.K. Chaudhary, K. Das, R.B. Pachori, Breast Cancer Diagnosis Using Iterative Fourier-Bessel Decomposition Method Based CNN-kernel Features, 2022.

[32] S.I. Khan, R.B. Pachori, Automated Eye Movement Classification Based on EMG of EOM Signals Using FBSE-EWT Technique, IEEE Transactions on Human-Machine Systems, 2023.

[33] K. Das, P. Verma, R.B. Pachori, Assessment of chanting effects using EEG signals, in: 2022 24th International Conference on Digital Signal Processing and its Applications (DSPA), IEEE, 2022, March., pp. 1–5.

[34] A. Nalwaya, K. Das, R.B. Pachori, Automated Emotion Identification Using Fourier-Bessel Domain-Based Entropies, Entropy 24 (10) (2022) 1322.

[35] R.B. Pachori, P. Sircar, A new technique to reduce cross terms in the Wigner distribution, Digital Signal Process. 17 (2) (2007) 466–474.

[36] R.B. Pachori, P. Sircar, EEG signal analysis using FB expansion and second-order linear TVAR process, Signal Process. 88 (2) (2008) 415–420.

[37] W. Mumtaz, L. Xia, M.A. Mohd Yasin, S.S. Azhar Ali, A.S. Malik, A wavelet-based technique to predict treatment outcome for major depressive disorder, PLoS One 12 (2) (2017) e0171409.

[38] W. Mumtaz, et al., Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD), Biomed. Signal Process. Control (2017, 31,) 108–115.

[39] M. Diykh, F.S. Miften, S. Abdulla, R.C. Deo, S. Siuly, J.H. Green, A.Y. Oudahb, Texture analysis based graph approach for automatic detection of neonatal seizure from multi-channel EEG signals, Measurement 190 (2022), 110731.

[40] I. Alsafy, M. Diykh, Developing a robust model to predict depth of anesthesia from single channel EEG signal, Phys Eng Sci Med 45 (3) (2022) 793–808.

[41] M. Diykh, S. Abdulla, R.C. Deo, S. Siuly, M. Ali, Developing a Novel Hybrid Method Based on Dispersion Entropy and Adaptive Boosting Algorithm for Human Activity Recognition, Computer Methods Prog Biomed (2022) 107305.
[42] R. Lafta, J. Zhang, X. Tao, Y. Li, M. Diykh, J.C.W. Lin, A structural graph-coupled advanced machine learning ensemble model for disease risk prediction in a telehealthcare environment, in: Big Data in Engineering Applications, Springer, Singapore, 2018, pp. 363–384.
[43] V. Gupta, R.B. Pachori, Epileptic seizure identification using entropy of FBSE based EEG rhythms, Biomed. Signal Process. Control 53 (2019), 101569.
[44] R.K. Tripathy, A. Bhattacharyya, R.B. Pachori, Localization of myocardial infarction from multi-lead ECG signals using multiscale analysis and convolutional neural network, IEEE Sens. J. 19 (23) (2019) 11437–11448.
[45] P. Gajbhiye, R.K. Tripathy, R.B. Pachori, Elimination of ocular artifacts from single channel EEG signals using FBSE-EWT based rhythms, IEEE Sens. J. 20 (7) (2019) 3687–3696.
[46] P.K. Chaudhary, V. Gupta, R.B. Pachori, Fourier-Bessel representation for signal processing: A review, Digital Signal Process. (2023), 103938.
[47] R.B. Pachori. Time-Frequency Analysis Techniques and their Applications.
[48] A. Anuragi, D.S. Sisodia, R.B. Pachori, EEG-based cross-subject emotion recognition using Fourier-Bessel series expansion based empirical wavelet transform and NCA feature selection method, Inf. Sci. 610 (2022) 508–524.
[49] M. Diykh, F.S. Miften, S. Abdulla, K. Saleh, J.H. Green, Robust approach to depth of anaesthesia assessment based on hybrid transform and statistical features<? show [AQ ID= Q1]?>, IET Sci. Meas. Technol. 14 (1) (2020) 128–136.
[50] U.R. Acharya, F. Molinari, S.V. Sree, S. Chattopadhyay, K.H. Ng, J.S. Suri, Automated diagnosis of epileptic EEG using entropies, Biomed. Signal Process. Control 7 (4) (2012) 401–408.
[51] H.B. Mann, D.R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. (1947) 50–60.
[52] H.W. Lilliefors, On the Kolmogorov-Smirnov test for normality with mean and variance unknown, J. Am. Stat. Assoc. 62 (318) (1967) 399–402.
[53] K. Yan, L. Kou, D. Zhang, Learning domain-invariant subspace using domain features and independence maximization, IEEE Trans. Cybern. 48 (1) (2017) 288–299.
[54] R. Zhao, Y. Xia, Y. Zhang, Unsupervised sleep staging system based on domain adaptation, Biomed. Signal Process. Control 69 (2021), 102937.
[55] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, in; *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144-152.
[56] T.W. Liao, Clustering of time series data—a survey, Pattern Recogn. 38 (11) (2005) 1857–1874.
[57] H. Rajadurai, U.D. Gandhi, A stacked ensemble learning model for intrusion detection in wireless network, Neural Comput. & Applic. (2020) 1–9.
[58] S. Abdulla, M. Diykh, R.L. Laft, K. Saleh, R.C. Deo, Sleep EEG signal analysis based on correlation graph similarity coupled with an ensemble extreme machine learning algorithm, Expert Syst. Appl. 138 (2019), 112790.
[59] H. Al-Hadeethi, S. Abdulla, M. Diykh, J.H. Green, Determinant of covariance matrix model coupled with AdaBoost classification algorithm for EEG seizure detection, Diagnostics 12 (1) (2021) 74.
[60] S. Mahato, S. Paul, Detection of major depressive disorder using linear and non-linear features from EEG signals, Microsyst. Technol. 25 (3) (2019) 1065–1076.
[61] S.Y. Kuzu, Evaluation of gradient boosting and deep learning algorithms in dimuon production, J. Mol. Struct. 1277 (2023), 134834.
[62] M. Diykh, S. Abdulla, R.C. Deo, S. Siuly, M. Ali, Developing a novel hybrid method based on dispersion entropy and adaptive boosting algorithm for human activity recognition, Comput. Methods Programs Biomed. 229 (2023), 107305.
[63] S. Abdulla, M. Diykh, S. Siuly, M. Ali, An Intelligent Model Involving Multi-Channels Spectrum Patterns Based Features for Automatic Sleep Stage Classification, Int. J. Med. Inf. (2023), 105001.
[64] P.N. Tan, M. Steinbach, V. Kumar, Introduction to data mining, Pearson Education India, 2016.
[65] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, H. Adeli, D.P. Subha, Automated EEG-based screening of depression using deep convolutional neural network, Comput. Methods Programs Biomed. 161 (2018) 103–113.