Volume 8, Issue 4, 082419 (1-11)

Signal Processing and Renewable Energy (SPRE)

https://dx.doi.org/10.57647/j.spre.2024.0804.19

Comparing linear, nonlinear and time frequency-based features in the EEMD domain for depression detection application using EEG signals

Masoud Rafiei[1,2], Mahsa Rastegar Pour[3], Hesam Akbari[4,*]

[1] *Department of Mechanical Engineering, University of North Texas, Texas, USA.*
[2] *Department of Computer Science and Engineering, University of North Texas, Texas, USA.*
[3] *Department of Electrical Engineering, University of North Texas, Texas, USA.*
[4] *Department of Information Science, University of North Texas, Texas, USA.*

[*] **Corresponding author:** hesam.akbari@unt.edu

Original Research

Received:
23 August 2024
Revised:
27 September 2024
Accepted:
2 October 2024
Published online:
30 December 2024

Abstract:

Depression is a disease that, if left untreated, can lead to risks such as self-harm and suicide. It can be treated with exercise and antidepressant drugs. Timely and accurate diagnosis of depressed subjects can help prevent such dangerous behaviors. An electroencephalogram (EEG) is an available tool in clinics and hospitals for measuring brain activity. In this article, the ability of three kinds of parameters, including linear, nonlinear, and time-frequency-based features, was evaluated for the accurate detection of normal and depressed EEG signals. The EEG signals of 22 normal and 22 depressed subjects were used to evaluate the proposed framework. In the first step, the input EEG signals were decomposed into their intrinsic mode functions (IMFs) using the ensemble empirical mode decomposition (EEMD) method, and the best mode with the highest effect was selected by mutual information (MI). After that, these three kinds of features were extracted from the best IMF and fed to a k-nearest neighbors (KNN) classifier. We found that time-frequency features are better than nonlinear and linear features in detecting depressed EEG signals. Additionally, EEG signals from the right hemisphere of the brain were better than those from the left side for depression detection. In the final step of the proposed method, significant features were selected by the particle swarm optimization (PSO) algorithm and fed to the KNN classifier, resulting in average classification accuracies (ACC) of 91% and 92% for the left and right hemispheres, respectively. The proposed method can be used in clinics and hospitals for accurate, fast, and accessible detection of depressed subjects.

Keywords: EEG; Depression; Feature extraction; Classification

1. Introduction

Depression is a medical disorder that can affect a patient's personal and social life. One in ten people struggles with depression, making it the most common cause of disability worldwide [1]. Unfortunately, because it is a mental illness, it is less understood by people than physical illnesses. Depression can present with a variety of symptoms, including a low mood, loss of interest in things typically enjoyed, changes in appetite, feelings of worthlessness or excessive guilt, sleeping too much or too little, poor concentration, restlessness or slowness, loss of energy, or recurrent thoughts of suicide [2].

In developing countries, many people do not have access to a psychiatrist. The diagnosis of depressed individuals typically requires the experience of a psychiatrist, which is prone to human error. For this reason, a method based on computer-aided diagnosis (CAD) systems is desirable for depression detection.

Magnetic resonance imaging (MRI) and positron emission tomography (PET) scans have previously been used in diagnosing brain disorders such as depression [3, 4], as shown in Fig. 1.

However, the main drawback of these methods is that this equipment is inaccessible to a large portion of the population. On the other hand, Electroencephalogram (EEG) signals have been widely used in clinics and hospitals to measure brain neuron activity [5, 6]. Many methods have been proposed based on EEG signals for detecting brain disorders such as epilepsy [7–10], schizophrenia [11], sleep apnea [12], autism [13], focal seizures [14], attention deficit hyperactivity disorder (ADHD) [15, 16], and alcoholism [17].

Most developed methods for the detection of depression in EEG signals are based on time-frequency decomposition methods, such as traditional wavelet transforms and nonlinear feature extraction. In [18], a method based on power, frequency, asymmetry, and coherence was developed for the detection of depression in EEG signals. A wavelet-chaos methodology using discrete wavelet transforms (DWT) and two fractal dimensions (Katz and Higuchi) was proposed in [19], where the input EEG signals were decomposed by DWT and fractal dimensions were computed as discrimination features. These features were then fed to an enhanced probabilistic neural network (EPNN) classifier, which reported a maximum classification accuracy (ACC) of 91.3%. In a similar study [20], DWT was used to decompose EEG signals, and relative wavelet energy (RWE) was extracted from the coefficients as features, which were then fed to an artificial neural network (ANN) classifier, achieving a classification ACC of 98.11%. In [21], a method based on extracting nonlinear features was proposed, including detrended fluctuation analysis, fractal dimension, correlation dimension, and the largest Lyapunov exponent. These features were fed to various classifiers such as k-nearest neighbor, linear discriminant analysis, and logistic regression, with a maximum classification ACC of 90% obtained by the logistic regression classifier.

A novel technique, namely spatiotemporal analysis of relative convergence (STARC), was proposed in [22] for the diagnosis of male and female MDD patients. In a comparison study [23], the ability of the spectral asymmetry index (SASI) and the Higuchi fractal dimension as discrimination features was tested for depression detection, with a maximum classification ACC of 94% obtained using the Higuchi fractal dimension feature, indicating its effectiveness in depression EEG detection. In [24], an entropy-based method was used where EEG signals were decomposed using wavelet packet transform into sub-bands. The approximate, sample, Renyi, and bispectral phase entropies were then computed from the sub-bands as features and fed to a probabilistic neural network (PNN) classifier, achieving a perfect classification ACC of 99.50% in depression diagnosis.

In [25], a new index for detecting depression in EEG signals was proposed based on nonlinear features, achieving a classification ACC of 98%. Liao et al. [26] proposed a feature selector called kernel eigen-filter-bank common spatial pattern (KEFB-CSP) and then computed the dimensionality reduction vector using kernel principal component analysis (kernel PCA), with a reported accuracy of about 81.23% from the SVM classifier. In [27], synchronization likelihood was employed as a feature for depression EEG signal detection, with classification sensitivity (SEN) and specificity (SPE) of 96.9% and 91%, respectively, using two classifiers.

Bairy et al. [28] applied the predictive coding (LPC) method and extracted higher-order statistical features. Features were ranked using ROC and fed to a bag tree classifier, resulting in an ACC of 94.30%. In [29], linear and nonlinear features were fed to a logistic regression classifier, which achieved a maximum classification ACC of 92% in depression EEG signal detection. In [30], a deep learning method based on a convolutional neural network (CNN) classifier attained maximum classification ACCs of 93.5% and 96% for the left and right hemispheres, respectively. In [31], a new wavelet called the bandwidth-duration localized (BDL) three-channel orthogonal wavelet filter bank (TCOWFB) decomposed the EEG signal into three levels. The L2 norm entropy was then computed from the sub-bands as a discrimination feature and fed to the least squares SVM classifier, reporting a classification ACC of 99.58% in normal and

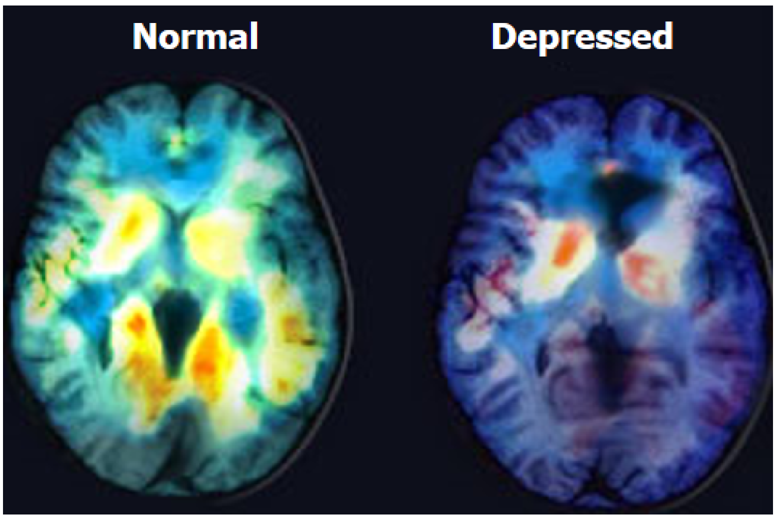

Figure 1. Sample of positron emission tomography (PET) scan in normal and depressed brain.

depression EEG signal classification.

It is clear that most of the proposed methods for depression detection are based on wavelet-based decomposition methods. By reviewing past works, it can be said that most of them rely on traditional wavelet transforms. The major drawback of DWT is that it is not adaptive to the frequency of the input signal. The empirical mode decomposition (EMD) technique has been proposed for analyzing nonlinear and non-stationary signals like EEG. EMD extracts the intrinsic mode functions (IMFs) of the input signal. Although EMD has shown acceptable performance in biomedical signal processing applications, its IMFs are very sensitive to noise and suffer from mode mixing problems. The ensemble EMD (EEMD) was proposed to address this issue. In this work, for the first time, the ability of EEMD as a processing tool is evaluated in the application of depression EEG signal detection.

Additionally, by reviewing feature extraction methods for depression detection in the literature, it is found that most of them are based on linear features like statistical features, and nonlinear features like chaotic and fractal dimensions. Recently, time-frequency-based features such as mean instantaneous amplitude (IA), mean instantaneous frequency (IF) and mean instantaneous phase (IP) of IMFs have shown acceptable performance in some EEG signal processing problems, such as brain-computer interfaces (BCI) [32, 33] and focal EEG signal detection [34, 35]. In this work, for the first time, the ability of these three time-frequency-based features is evaluated in depression detection. Moreover, no previous work has compared the ability of linear and nonlinear features in depression detection. For this reason, we compare the ability of linear, non-linear, and time-frequency-based features (i.e. IA, IF, and IP) for accurate detection of depressed patients.

In this study, for comparison, the EEG signals are first filtered with a low-pass filter. Then, the IMFs are extracted in the EEMD domain. The best IMF is selected using the mutual information (MI) method. After that, linear, nonlinear, and time-frequency-based features are extracted. Significant features are selected using the particle swarm optimization (PSO) algorithm and finally fed to a k-nearest neighbors (KNN) classifier. The tenfold cross-validation strategy is used during the training and testing phases of the classifier. The paper is organized as follows:In Section 2, the database used is defined. In Section 3, the proposed framework is described, covering EEMD, MI, feature extraction, PSO, and KNN. The experimental steps are provided in Section 4. The experimental results and discussion are presented in Section 5. The conclusion of our paper is provided in Section 6.

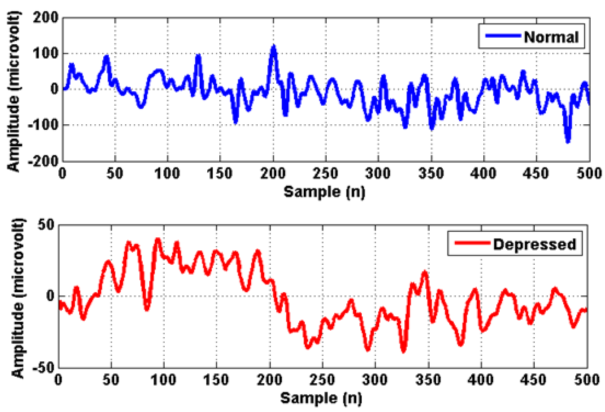

Figure 2. Shows a sample of normal and depressed EEG signals.

2. Used database

In this study, we used a dataset consisting of 44 subjects (male and female). The dataset was obtained from 22 healthy individuals with no history of psychiatric disease and 22 individuals with depression. Participants in both the depressed and normal groups were selected from the age group of 23 to 58 years. EEG signals were recorded in a resting condition from the participants with open and closed eyes for ten minutes [36, 37]. The recordings were performed using bipolar surface electrodes according to the International $10-20$ system. EEG signals were recorded from the FP1-T3 channel on the left hemisphere and the FP2-T4 channel on the right hemisphere of the brain during the experiment. The sampling frequency rate was set to 256 Hz, and a notch filter was used to eliminate power line interference noise. Additionally, an experienced expert was consulted to manually eliminate and discard other factors such as muscle and eye movement artifacts from the signals. Given the non-stationary nature of EEG signals [10], they were divided into segments of 500 samples. In this work, MATLAB 2016 software was used for signal processing and analysis. A sample of normal and depressed EEG signals is shown in Fig. 2. This experiment was approved by the Research Ethics Committee of AJA University of Medical Sciences, Tehran, Iran [36, 37].

3. Proposed framework

This subset describes the steps of the proposed framework. The block diagram of the proposed framework for depression detection is shown in Fig. 3.

3.1 EEMD

EMD of the ensemble average of the Gaussian noise-assisted data is provided by EEMD. In fact, lower levels of

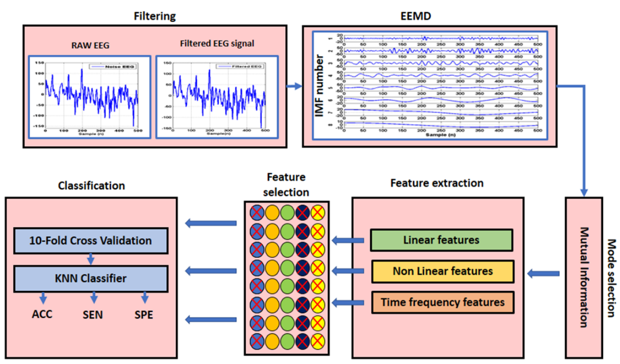

Figure 3. Block diagram of the proposed method.

Gaussian noise will be mixed with the original signal and then EMD will decompose it [38]. Here, we present EEMD steps.

1. Obtaining $x^{\mathrm{i}}(n) = x(n) + w^{\mathrm{i}}(n)$

where $x(n)$ represents the real signal and $w^{\mathrm{i}}(n)$ stands for a distinct realization (i =1,2,...,R) of the white Gaussian noise with the unit variance and a zero mean.

2. Decomposing $x^{\mathrm{i}}(n)$ by averaging, called relative IMFs as $\mathrm{IMF}_{\mathrm{k}}^{\mathrm{i}}(n)$, where k =1,2, ..., L, and L represents various modes.

3. Determining the modes average as:

$$\mathrm{IMF}_{\mathrm{k}}(n) = \frac{1}{\mathrm{L}} \sum_{\mathrm{i}=1}^{\mathrm{R}} \mathrm{IMF}_{\mathrm{k}}^{\mathrm{i}}(n) \tag{1}$$

Fig. 4 is a representation of the extracted IMFs for normal and depressed EEG signals.

3.2 Mode selection

The similarity or interdependency of the random variables can be quantified through mutual information (MI). This study deals with using such an approach as the mode selection method to determine the best IMF which is known as the most significant data of the input EEG signal. Then, such data has been used in multi-resolution analysis of the EEG signal, through EEMD. The MI can be mathematically formulated for individual variables of U and V, as follows [27]:

$$\mathrm{MI(U,V)} = \mathrm{H(U)} + \mathrm{H(V)} - \mathrm{H(U,V)} \tag{2}$$

in which, the function H shows the computed entropy value of each variable, and the joint entropy of the variables is shown as H(U,V). It should be noted that the available length of the records has been considered through the IMF selection process.

3.3 Feature extraction

Linear features: In this work, to evaluate the linear behavior of EEG signals, five simple statistical parameters-mean (Men), median (Med), standard deviation (Std), maximum (Max), and minimum (Min) of IMFs-are extracted as linear features.

Nonlinear features: To measure the nonlinear behavior of EEG signals, five entropies-Shannon entropy (ShEn), log energy entropy (LeEn), threshold entropy (ThEn), sure entropy (SureEn), and norm entropy (NrEn) of IMFs-are extracted as nonlinear features. Table 1 describes the extracted linear and nonlinear features.

Time-frequency based features: It has been suggested that the Hilbert transform should be applied to all IMFs obtained through the EMD procedure [24]. Any real IMF's analytic signal is determined by:

$$z(t) = c(t) + jc_H(t) \tag{3}$$

which the $c(t)$ Hilbert transform is defined as $c_H(t) = c(t) * \frac{1}{\pi t}$. Eq. (4) could be defined as:

$$z(t) = A(t)e^{j\varphi(t)} \tag{4}$$

The amplitude of the analytic signal $(A(t))$ and instant phase$\varphi(t)$ could be written as:

$$A(t) = \sqrt{c^2(t) + c_H^2(t)} \tag{5}$$

$$\varphi(t) = \arctan\left[\frac{c_H(t)}{c(t)}\right] \tag{6}$$

The analytic IMF instant frequency $z(t)$ is defined as:

$$w(t) = \frac{d\varphi(t)}{dt} \tag{7}$$

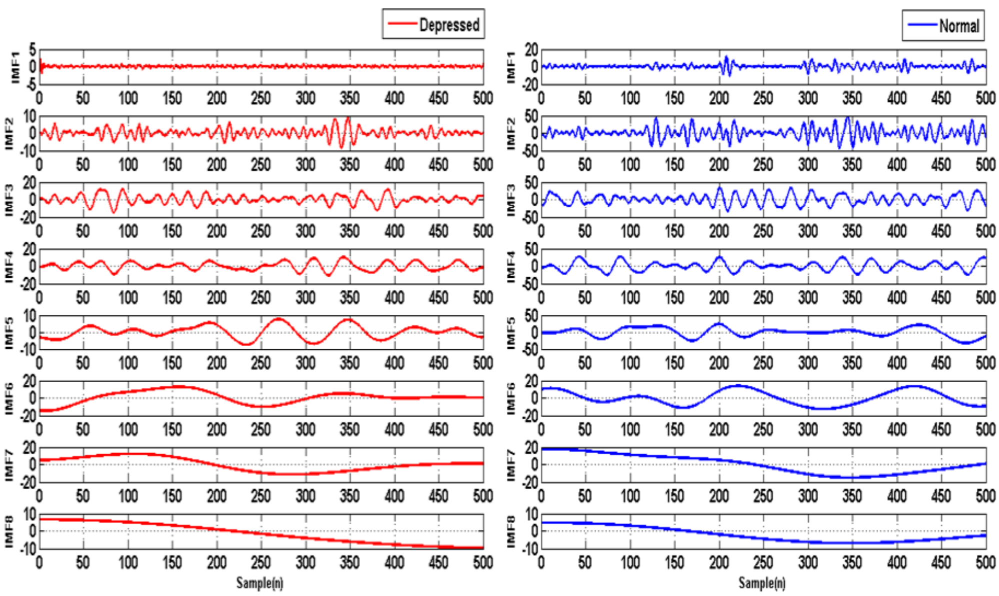

Figure 4. IMFs of EEG signals of normal and depressed EEG signals.

Table 1. Mathematical description of linear and nonlinear features.

No.	Feature	Expression	Remarks
1	Men	Mean= $\frac{1}{n}\sum_{i=1}^{n} x_i$	n and x_i is the lengths and ith sample of input data
2	Med	Median = $\begin{cases} x_{n/2} & if\ n\ be\ even \\ x_{[(n/2)+1]} + x_{[n/2]} & if\ n\ be\ even \end{cases}$	n and x_i is the lengths and ith sample of input data after sorting
3	Std	STD= $\sqrt{(1/n)\sum_{i=1}^{n}(x_i - mean)^2}$	n and x_i is the lengths and ith sample of input data
4	Max	MAX = $max(x_i)$	Maximum value of input data
5	Min	MIN = $min(x_i)$	Minimum value of input data
6	ShEn	ShEn = $\sum_{i=1}^{n} x_i \log(1/x_i)$	n and x_i is the lengths and ith sample of input data
7	LeEn	LeEn = $\sum_{i=1}^{n} \log(x_i^2)$	n and x_i is the lengths and ith sample of input data
8	ThEn	So, $\begin{cases} ThEn(x_i) = \begin{cases} 1 & \lvert x_i \rvert > \varepsilon \\ 0 & elsewhere \end{cases} \\ ThEn = \#\{\mathrm{i}\ such\ that \lvert x_i \rvert > \varepsilon\} \end{cases}$	x_i is the lengths of input data and ε is threshold
9	SureEn	SureEn= $n - \#\{\mathrm{i}\ such\ that \lvert x_i \rvert \le \varepsilon\} + \sum_i min(x_i^2, \varepsilon^2)$	x_i is the lengths of input data and ε is threshold
10	NrEn	NrEn = $\sum_{i=1}^{n} \log \lvert x_i \rvert^2$	n and x_i is the lengths and ith sample of input data

The signalx(t) presented in Eq. (3) could be stated in a Fourier-like expression as:

$$x(t) \approx \Re\left\{\sum_{m=1}^{M} A_m(t) e^{j\varphi_m(t)}\right\} \tag{8}$$

which the index m relates to mth IMF and $\Re\{.\}$ refers to the real component of a complex quantity.

In this work, three parameters of IMFs such as mean instantaneous amplitude (IA), mean instantaneous frequency (IF), and mean instantaneous phase (IP) of IMFs are extracted as discriminating features.

3.4 PSO

The binary PSO algorithm is one of the meta-heuristic algorithms that solve the problems with the least information. And also in the other categorization, belongs into the swarm intelligence algorithms category that was first introduced by Kennedy and Eberhart in 1997 and used as a solution to solve binary optimization problems [39]. In the initial PSO model, there are populations of N particles and each has a position and an objective value and it moves in the D dimension search space. Particles flow according to their own best previous experience and the best experience of the entire population to seek out the best possible optimum solution [40]. Note that the objective value is the suggestion of each particle to solve the problem and also in general, the vector representing the motion direction of the particle is known as the velocity vector since the velocity vector is always tangent upon the motion direction vector of a moving particle.

As mentioned each i particle of the population has some information in its own current position which comprises: the particle velocity, the best experienced position, and the best overall experienced position (Fig. 5(a)). In a d-dimension space, the position of the particle is expressed as $X = (x_{i1}, x_{i2}, \ldots, x_{iD})$ and the velocity of the particle is

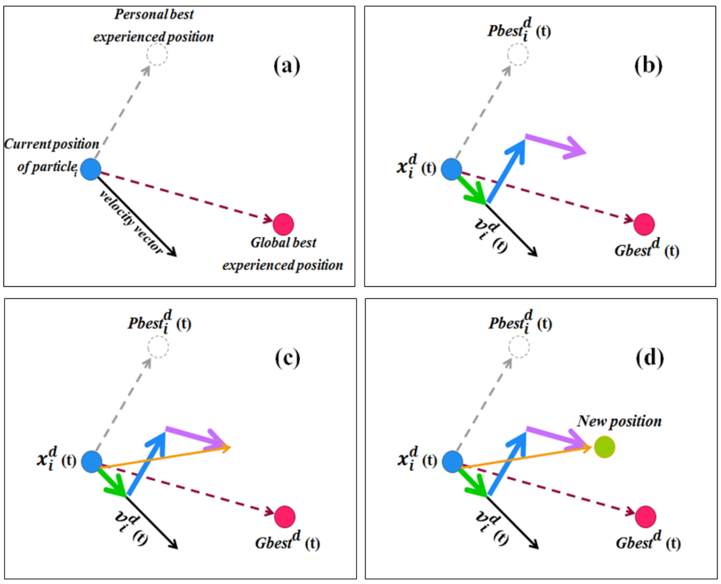

Figure 5. The steps of the performance of the BPSO algorithm to seek out the best solution to solve binary optimization problems.

denoted as $V = (v_{i1}, v_{i2}, ..., v_{iD})$; where i is the order of particles. The personal best experienced position of each particle is called Pbest and the global best experienced position is known as Gbest (Fig. 5(a)); for each iteration t, the Pbest of particles is compared with the Gbest; if the Pbest of a particle was better than Gbest then particle does not shift otherwise it has to move in the Gbest direction to find the best solution. The change of location occurs as follow: if each particle moves based on the weighted coefficient from its own previous velocity, Pbest and Gbest (Fig. 5(b)), then the resulted displacement vector of three vectors produces a vector that reveals the new position of the particle (Fig. 5(c) and (d)).
According to the above, the new velocity of each particle is updated as follow:

$$\begin{aligned} v_i^d(t+1) = & w(t) \times v_i^d(t) + \\ & c_1 \times r_1 \times \left(Pbest_i^d(t) - x_i^d(t)\right) + \\ & c_2 \times r_2 \times \left(Gbest^d(t) - x_i^d(t)\right) \end{aligned} \tag{9}$$

where x, v and i are the positions, the velocity, and the order of the particle, respectively; d is the dimension of search space; w is inertia weight; r_1 and r_2 are two random numbers with a uniform distribution between 1 and 0; c_1 and c_2 are the learning coefficients or the acceleration coefficients. And also the new position of each particle is updated as (Fig. 6):

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{10}$$

The better performance is influenced by the value of w parameters. The choice of the high values of inertia weight improves the public search capability of the algorithm and more spaces are explored, on the contrary, the lower values bound the search space, and the exploration is cautiously done, and generated responses are used for more exploitation. In this algorithm, at first, w is usually set to a higher value then it is gradually decreased to a lower value in order to balance between exploration and exploitation strategies

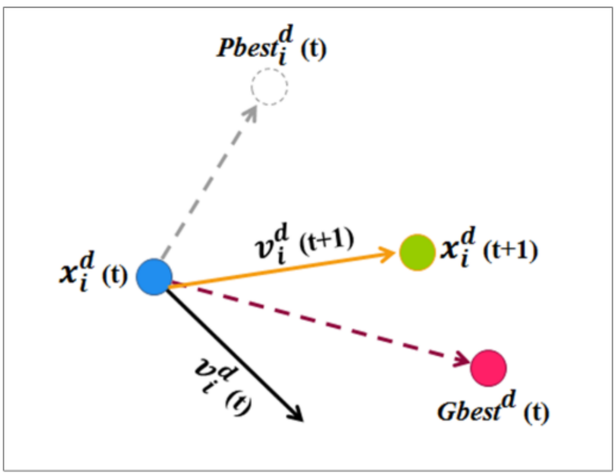

Figure 6. shows the updated velocity and new position of each particle.

[41]. The inertia weight is computed as follows:

$$w(t) = w_{max} - (w_{max} - w_{min}) \times \frac{t}{T_{max}} \tag{11}$$

where w_{min} and w_{max} are the bounds on the inertia weight and were fixed at 0.4 and 0.9, respectively; t is the current iteration, and T_{max} is the maximum number of iterations.
The PSO algorithm for solving the binary optimization problems operates as follows:
At first, the updated velocity is fed to the sigmoid function as follows:

$$S\left(v_i^d(t+1)\right) = \frac{1}{1+e^{-v_i^d(t+1)}} \tag{12}$$

The output of the sigmoid function is a number between 0 and 1; afterward, the new position of each particle is updated based on the probability value of the sigmoid function as follow:

$$x_i^d(t+1) = \begin{cases} 1, & if\,\delta < S\left(v_i^d(t+1)\right) \\ 0, & Otherwise \end{cases} \tag{13}$$

where δ is random number numbers with a uniform distribution between 1 and 0.

3.5 KNN

k-nearest neighbors (KNN) is a non-parametric and simple method with easy implementation that first, uses methods to the selecting of the optimal value K or manually determine the value of the K. The KNN algorithm computes the distance between every test data with all samples in the training dataset. Then K samples in the training dataset that have the nearest distance to that test data are chosen and saved as a new dataset. The class that has the most frequent in the new set is assigned as the class of that test data. In this algorithm, the value of parameter K will be crucial, because directly affects the produced output. According to Fig. 7,

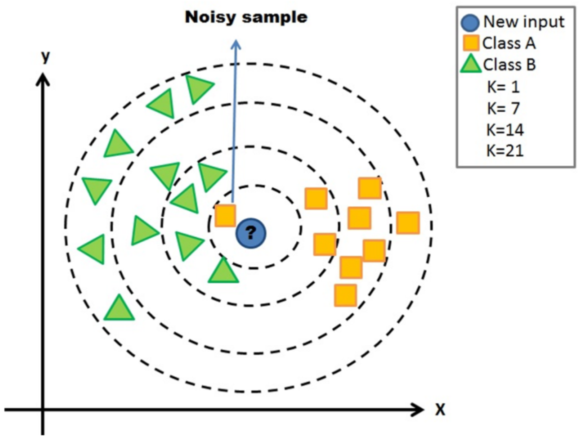

Figure 7. The schematic of the KNN classifier in the new data classification for different values of K. Visually, the new input belongs to class B but for K= 1, 7, 14, 21, KNN outputs will be A, B, A, and B, respectively. Hence, in the KNN method, the correct classification depends to the K value.

the choice of small K values makes the algorithm sensitive to the noisy sample.
The different metrics able to compute the distance between n vectors such as Euclidean distance, city block, and Minkowski metrics that is explained in the following (Assuming that x_i and y_i be two vectors):
• Euclidean distance is defined as:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} |x_\mathrm{i} - y_\mathrm{i}|^2} \tag{14}$$

• City block metric is defined as:

$$d(x,y) = \sum_{i=1}^{n} |x_\mathrm{i} - y_\mathrm{i}| \tag{15}$$

• Minkowski metric is defined as:

$$d(x,y) = \sqrt[p]{\sum_{i=1}^{n} |x_\mathrm{i} - y_\mathrm{i}|^p} \tag{16}$$

where n is the number of training data and p can be variable. If p=1 then the Minkowski and city block metric will be equivalent, and if p=2 then the Minkowski metric and the Euclidean distance will be equivalent. In this work, the data classification is implemented using the KNN classifier with Euclidean metric for k values of 1 to 9.

4. Experimental step

In the EEG database used, there are 400 files: 200 files for the normal group (100 files for the left hemisphere and 100 files for the right hemisphere) and 200 files for the depressed group (100 files for the left hemisphere and 100 files for the right hemisphere). In this work, the classification results of normal EEG signals versus depressed EEG signals were obtained separately for the right and left hemispheres.
First step: The EEG signals were filtered using a third-order Butterworth low-pass filter with a cutoff frequency of 80 Hz.
Second step: The input EEG signals were decomposed using the EEMD method to extract their IMFs. In this work, EEMD MATLAB functions written by Patrick Flandrin were used, which are freely available at http://perso.ens-lyon.fr/patrick.flandrin/emd.html. The EEMD function has three inputs: noise standard deviation (Nstd), number of realizations (NR), and maximum number of sifting iterations allowed (MaxIter). For this work, these parameters were set to 0.2, 500, and 5000, respectively. After decomposing the EEG signals, the IMFs were used in the next step to select the best IMF.
Third step: The mutual information (MI) between the input EEG signal and its IMFs was computed to obtain the common information between the input EEG signal and its IMFs. MI was used as a mode selection technique, with the IMF that had the highest MI value selected as the best mode. This IMF was considered to contain the most information about the complex behavior of the input EEG signal and the most important frequency components by having a narrower frequency bandwidth.
Fourth step: Five statistical characteristics (i.e. mean (Men), median (Med), standard deviation (Std), maximum (Max), and minimum (Min)) as linear features, five entropies (i.e. Shannon entropy (ShEn), log energy entropy (LeEn), threshold entropy (ThEn), sure entropy (SureEn), and norm entropy (NrEn)) as nonlinear features, and three time-frequency based characteristics (i.e. mean instantaneous amplitude (IA), mean instantaneous phase (IP), and mean instantaneous frequency (IF)) were extracted from the selected IMF. In other words, the feature matrix for the left and right hemispheres in the normal and depressed groups, after extracting the linear, nonlinear, and time-frequency based features, had sizes of 100×5, 100×5, and 100×3, respectively.
Fifth step: The Kruskal-Wallis statistical test was used to evaluate the capability of linear, nonlinear, and time-frequency based features in discriminating between normal and depressed subjects. The output of the Kruskal-Wallis test is the p-value. In statistics, a significant difference between normal and depressed classes is indicated if the p-value for a feature is less than 0.05. In this work, the "Kruskal-Wallis" MATLAB function was used to calculate the p-value.
Sixth step: The ten-fold cross-validation technique was used to mitigate the overfitting problem during the training and testing phases of the KNN classifier. In ten-fold cross-validation, the dataset is randomly divided into ten equal subsets. For each iteration, one subset is used as the testing data, and the remaining subsets are used as training data. Let γ be the number of selected features. The size of the feature vector matrix for testing and training data was γ×180 and γ×20, respectively. The average value of the objective parameters was reported as the performance of the KNN classifier.
In binary classification of normal and depressed EEG signals, after classifying by any classifier, there are four possible outcomes:
True Positive (TP): When a classifier correctly identifies an input data point as a depressed EEG signal.
True Negative (TN): When a classifier correctly identifies an input data point as a normal EEG signal.
False Positive (FP): When a classifier incorrectly identifies an input data point as a depressed EEG signal.
False Negative (FN): When a classifier incorrectly identifies an input data point as a normal EEG signal.
For evaluating the performance of the classifier, ACC, SEN, and SPE of the classifier defines as below:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \tag{17}$$

$$SEN = \frac{TP}{TP+FN} \times 100 \tag{18}$$

$$SPE = \frac{TN}{TN+FP} \times 100 \tag{19}$$

Seventh step: Features with a p-value less than 0.05 were selected and fed into the KNN classifier with Euclidean distance. The values of accuracy (ACC), sensitivity (SEN), and specificity (SPE) for linear, nonlinear, and time-frequency based features were reported to compare the capabilities of these three types of features in detecting depressed EEG

signals.
Seventh step: All of the extracted features were fed into the PSO algorithm to obtain the best feature vector with fewer dimensions, which resulted in the highest classification accuracy (ACC).

5. Results and Discussion

In the proposed framework, the EEG signals are decomposed into their IMFs using the EEMD method. Subsequently, linear, nonlinear, and time-frequency based features are extracted to discriminate between normal and depressed EEG signals. Fig. 4 shows samples of normal and depressed EEG signals, and Fig. 7 shows their decomposed IMFs. The significant mode is then selected using the MI technique. At the first decomposition levels of EEG signals (e.g., IMF1 and IMF2), the IMFs represent higher frequency components. In contrast, at the last decomposition levels (e.g., IMF7 and IMF8), the IMFs represent lower frequency components. We found that in the normal and depressed groups, the significant modes identified by MI were predominantly from the first and lower decomposition levels, respectively. The mean, standard deviation, and p-values of the linear, nonlinear, and time-frequency features, after normalization, are presented in Table 2.
We can observe that the mean values of the extracted features in the normal group are higher than those in the depressed group for both hemispheres. In a similar study [25], nonlinear features were used for classifying normal and depressed EEG signals, and it was found that the mean values of the normal features were higher than those of the depressed features. This may be attributed to the decreased complexity of depressed EEG signals. Recent research [42] demonstrated that depressed EEG signals exhibit simpler shapes compared to normal ones when plotted in a 2D space using the second-order difference plot, likely due to reduced brain activity in depressed states. Consequently, the lower standard deviation values in Table 1 could be due to this reduced complexity.
Another notable point in Table 2 is the p-values. It can be seen that all extracted linear, nonlinear, and time-frequency based features significantly impact the normal versus depressed classification task for both hemispheres (p-value < 0.05). This indicates that all these features can be effectively used for detecting depressed EEG signals.
To compare the abilities of linear, nonlinear, and time-frequency based features, we trained and tested the KNN classifier using these feature groups with a tenfold cross-validation strategy. The resulting classification parameters for linear, nonlinear, and time-frequency based features are presented in Table 3.
It is evident that time-frequency based features are more effective in detecting depressed EEG signals compared to nonlinear and linear features, which could be attributed to the use of mode selection. Additionally, from Table 3, it can be observed that the performance of features extracted from the right hemisphere is superior to those from the left hemisphere. This conclusion is supported by visual analysis of Fig. 1 and aligns with previous studies indicating that the left side of the brain is generally better for detecting depression [14, 36, 42, 43].
In the final step, to achieve the highest classification accuracy (ACC), the PSO algorithm was employed to obtain the optimal feature vector subsets. The features selected by the PSO algorithm include Max, LeEn, ThEn, SureEn, IA, IF,

Table 2. Statistical point of view of features.

Left Halve		**Men**	**Med**	**Std**	**Max**	**Min**	**ShEn**	**LeEn**	**ThEn**	**SureEn**	**NrEn**	**IA**	**IF**	**IP**
Normal	**Mean**	0.65	0.68	0.80	0.79	0.29	0.45	0.84	0.75	0.85	0.86	0.79	0.89	0.54
	std	0.30	0.60	0.02	0.25	0.21	0.26	0.24	0.24	0.24	0.32	0.31	0.26	0.12
Depressed	**Mean**	0.54	0.63	0.45	0.67	0.23	0.39	0.79	0.64	0.68	0.58	0.56	0.64	0.49
	std	0.19	0.39	0.02	0.15	0.18	0.16	0.20	0.19	0.23	0.29	0.21	0.27	0.13
p-value		< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05
Right Halve		**Men**	**Med**	**Std**	**Max**	**Min**	**ShEn**	**LeEn**	**ThEn**	**SureEn**	**NrEn**	**IA**	**IF**	**IP**
Normal	**Mean**	0.68	0.71	0.76	0.76	0.25	0.49	0.79	0.69	0.79	0.85	0.76	0.86	0.63
	std	0.29	0.59	0.02	0.23	0.19	0.23	0.23	0.22	0.26	0.26	0.26	0.24	0.09
Depressed	**Mean**	0.53	0.65	0.55	0.66	0.24	0.37	0.76	0.54	0.73	0.72	0.47	0.56	0.53
	std	0.16	0.29	0.03	0.13	0.16	0.17	0.19	0.16	0.18	0.25	0.23	0.23	0.12
p-value		< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05

Table 3. Classification results for left and right EEG signals by linear, nonlinear and time- frequency based features.

Left Halve				
Feature	ACC	SEN	SPE	Number of neighbors in KNN classifier
Linear	81	78	79	5
Nonlinear	84	79	80	5
Time frequency	86	81	82	8
Right Halve				
Feature	ACC	SEN	SPE	Number of neighbors in KNN classifier
Linear	82	78	80	6
Nonlinear	85	83	84	7
Time-frequency	88	85	86	9

and IP from the left hemisphere, and Max, ThEn, SureEn, IA, IF, and IP from the right hemisphere. The classification parameters for these selected features are presented in Table 4.

The highest classification accuracies of 91% and 92% were achieved using EEG signals from the left and right hemispheres, respectively, with 7 and 6 features.

The main founding and contribution of the proposed framework can be listed as follows:

1. This is the first study to evaluate the effectiveness of three types of features for detecting depressed EEG signals in the EEMD domain.
2. We employed a simple low-pass filter to remove noise and artifacts, whereas [24] used the total variation method for noise cancellation. The proposed method is computationally less intensive compared to theirs.
3. We utilized a mode selection method based on MI to select the best mode and reduce the number of IMFs before feature extraction.
4. The proposed method requires approximately 0.3 seconds for IMF extraction, 0.02 seconds for mode detection, 0.01 seconds for feature extraction, and 0.003 seconds for classification, indicating its efficiency.
5. The proposed method was evaluated using bipolar EEG signals with two electrodes on each hemisphere, whereas [19] used seven electrodes, [26] used eight electrodes, and [44] used nineteen electrodes.
6. The results in this work were reported using a 10-fold cross-validation technique, while [18, 20, 24, 25] reported results without cross-validation.
7. The proposed framework uses only seven and six features for classifying EEG signals from the left and right hemispheres, respectively, whereas [20] used twenty features, [21] used thirty features, and [25] used fifteen features for classifying normal and depressed EEG signals.

6. Conclusion

In this work, the effectiveness of three types of features in the EEMD domain was evaluated for classifying depressed and normal EEG signals. To select the optimal mode, mutual information (MI) was applied to the input EEG signals and their intrinsic mode functions (IMFs), with the mode having the highest MI being chosen as the best mode. The features extracted from the selected IMF include five linear parameters (mean, median, standard deviation, maximum, minimum), five nonlinear parameters (Shannon entropy, log energy entropy, threshold entropy, sure entropy, norm entropy), and three time-frequency based parameters (instantaneous amplitude, instantaneous frequency, instantaneous phase). The mean and standard deviation values of these features in the normal group were higher than in the depressed group for both the right and left hemispheres, indicating decreased brain activity in depression compared to normal states. All extracted features significantly impacted the classification task (p-value < 0.05). The PSO algorithm selected the best features, which were then fed into the KNN classifier. The proposed method achieved average classification accuracies (ACC) of 91% and 92%, sensitivities (SEN) of 88% and 91%, and specificities (SPE) of 88% and 90% for EEG signals from the left and right hemispheres, respectively. Features extracted from the right hemisphere had a better impact on depression detection compared to those from the left hemisphere. Future work will explore the method's capability in detecting other disorders such as sleep apnea and predicting seizures.

Table 4. Classification results for left and right EEG signals by selected features.

Halve	ACC	SEN	SPE	KNN*	NF**
Left	91	88	88	8	7
Right	92	91	90	9	6

*Number of neighbors in KNN classifier.
**Number of selected features.

Authors contributions
All authors contributed equally to prepare the paper.

Availability of data and materials
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interests
The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] W. H. O. *Depression*, 2020. URL https://www.who.int/news-room/fact-sheets/detail/depression.

[2] J. B. Henriques and R. J. Davidson. "Left frontal

hypoactivation in depression.". *Journal of abnormal psychology*, **100**(4):535, 1991.

[3] R. Elliott et al. "Prefrontal dysfunction in depressed patients performing a complex planning task: a study using positron emission tomography.". *Psychological medicine*, **27**:931–942, 1997.

[4] L. L. Zeng et al. "Unsupervised classification of major depression using functional connectivity MRI.". *Human brain mapping*, **35**(4):1630–1641, 2014.

[5] N. Sairamya et al. "A new approach for automatic detection of focal EEG signals using wavelet packet decomposition and quad binary pattern method.". *Biomedical Signal Processing and Control*, **63**: 102096, 2021.

[6] H. Akbari and S. Esmaili. "A Novel Geometrical Method for Discrimination of Normal, Interictal and Ictal EEG Signals.". *Traitement du Signal*, **37**(1):59–68, 2020.

[7] H. Akbari, S. Saraf Esmaili, and S. Farzollah Zadeh. "Detection of Seizure EEG Signals Based on Reconstructed Phase Space of Rhythms in EWT Domain and Genetic Algorithm.". *Signal Processing and Renewable Energy*, **4**(2):23–36, 2020.

[8] S. Ghofrani and H. Akbari. "Comparing nonlinear features extracted in EEMD for discriminating focal and non-focal EEG signals.". *International Society for Optics and Photonics*, 2019.

[9] H. Akbari and S. Ghofrani. "Fast and accurate classification f and nf EEG by using sodp and EWT.". *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, **11**(11):29–35, 2019.

[10] H. Akbari and M. T. Sadiq. "Detection of focal and non-focal EEG signals using non-linear features derived from empirical wavelet transform rhythms.". *Physical and Engineering Sciences in Medicine*, pages 1–15, .

[11] M. Sharma and U. R. Acharya. "Automated detection of schizophrenia using optimal wavelet-based $$ |1$$| 1 norm features extracted from single-channel EEG.". *Cognitive Neurodynamics*, pages 1–14, 2021.

[12] M. Alimardani and G. de Moor. "Automatic Classification of Sleep Apnea Type and Severity using EEG Signals.". 2021.

[13] Q. Mohi-Ud-Din and A. Jayanthy. "EEG feature extraction using wavelet transform for classifying autism spectrum disorder.". *Materials Today: Proceedings*, 2021.

[14] H. Akbari and M. T. Sadiq. "Detection of focal and non-focal EEG signals using non-linear features derived from empirical wavelet transform rhythms.". *Physical and Engineering Sciences in Medicine*, **44**, .

[15] R. Y. Karimui, S. Azadi, and P. Keshavarzi. "The ADHD effect on the high-dimensional phase space trajectories of EEG signals.". *Chaos, Solitons & Fractals*, **121**:39–49, 2019.

[16] R. Y. Karimui, S. Azadi, and P. Keshavarzi. "The ADHD effect on the actions obtained from the EEG signals.". *Biocybernetics and Biomedical Engineering*, **38**(2):425–437, 2018.

[17] S. Bavkar, B. Iyer, and S. Deosarkar. "Optimal EEG channels selection for alcoholism screening using EMD domain statistical features and harmony search algorithm.". *Biocybernetics and Biomedical Engineering*, **41**(1):83–96, 2021.

[18] V. Knott et al. "EEG power, frequency, asymmetry and coherence in male depression.". *Psychiatry Research: Neuroimaging*, **106**(2):123–140, 2001.

[19] M. Ahmadlou, H. Adeli, and A. Adeli. "Fractality analysis of frontal brain in major depressive disorder.". *International Journal of Psychophysiology*, **85**(2):206–211, 2012.

[20] S. D. Puthankattil and P. K. Joseph. "Classification of EEG signals in normal and depression conditions by ANN using RWE and signal entropy.". *Journal of Mechanics in Medicine and biology*, **12**(04):1240019, 2012.

[21] B. Hosseinifard, M. H. Moradi, and R. Rostami. "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal.". *Computer methods and programs in biomedicine*, **109**(3):339–345, 2013.

[22] M. Ahmadlou, H. Adeli, and A. Adeli. "Spatiotemporal analysis of relative convergence of EEGs reveals differences between brain dynamics of depressive women and men.". *Clinical EEG and neuroscience*, **44**(3):175–181, 2013.

[23] M. Bachmann et al. "Spectral asymmetry and Higuchi's fractal dimension measures of depression electroencephalogram.". *Computational and mathematical methods in medicine*, **2013**, 2013.

[24] O. Faust et al. "Depression diagnosis support system based on EEG signal entropies.". *Journal of mechanics in medicine and biology*, **14**(03):1450035, 2014.

[25] U. R. Acharya et al. "A novel depression diagnosis index using nonlinear features in EEG signals.". *European Neurology*, **74**(1-2):79–83, 2015.

[26] S. C. Liao et al. "Major depression detection from EEG signals using kernel eigen-filter-bank common spatial patterns.". *Sensors*, **17**(6):1385, 2017.

[27] W. Mumtaz et al. "Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD).". *Biomedical Signal Processing and Control*, **31**:108–115, 2017.

[28] G. M. Bairy et al. "Automated diagnosis of depression electroencephalograph signals using linear prediction coding and higher order spectra features.". *Journal of Medical Imaging and Health Informatics*, **7**(8):1857–1862, 2017.

[29] M. Bachmann et al. "Methods for classifying depression in single channel EEG using linear and nonlinear signal analysis.". *Computer methods and programs in biomedicine*, **155**:11–17, 2018.

[30] U. R. Acharya et al. "Automated EEG-based screening of depression using deep convolutional neural network.". *Computer methods and programs in biomedicine*, **161**:103–113, 2018.

[31] M. Sharma et al. "An automated diagnosis of depression using three-channel bandwidth-duration localized wavelet filter bank with EEG signals.". *Cognitive Systems Research*, **52**:508–520, 2018.

[32] M. T. Sadiq. "Motor imagery BCI classification based on novel two-dimensional modelling in empirical wavelet transform.". *Electronics Letters*, **56**(25): 1367–1369, 2020.

[33] M. T. Sadiq et al. "Motor imagery EEG signals decoding by multivariate empirical wavelet transform-based framework for robust brain–computer interfaces.". *IEEE Access*, **7**:171431–171451, 2019.

[34] P. Singh and R. B. Pachori. "Classification of focal and nonfocal EEG signals using features derived from Fourier-based rhythms.". *Journal of Mechanics in Medicine and Biology*, **17**(07):1740002, 2017.

[35] R. Sharma, R. B. Pachori, and S. Gautam. "Empirical mode decomposition based classification of focal and non-focal EEG signals.". *IEEE*, , 2014.

[36] H. Akbari, M. T. Sadiq, and A. U. Rehman. "Classification of normal and depressed EEG signals based on centered correntropy of rhythms in empirical wavelet transform domain.". *Health Information Science and Systems*, **9**(1):1–15, 2021.

[37] H. Akbari et al. "Depression Detection Based on Geometrical Features Extracted from SODP Shape of EEG Signals and Binary PSO Depression Detection Based on Geometrical Features Extracted from SODP Shape of EEG Signals and Binary PSO.". .

[38] S. Bhat. "Autism: cause factors, early diagnosis and therapies.". *Reviews in the Neurosciences*, **25**(6):841–850, 2014.

[39] J. Kennedy and R. C. Eberhart. "A discrete binary version of the particle swarm algorithm. in 1997 IEEE International Conference on systems, Man, and Cybernetics. Computational cybernetics and simulation.". *IEEE*, , 1997.

[40] H. S. Behera, P. K. Dash, and B. Biswal. "Power quality time series data mining using S-transform and fuzzy expert system.". *Applied Soft Computing*, **10**(3): 945–955, 2010.

[41] A. Unler and A. Murat. "A discrete particle swarm optimization method for feature selection in binary classification problems.". *European Journal of Operational Research*, **206**(3):528–539, 2010.

[42] H. Akbari et al. "Depression Detection Based on Geometrical Features Extracted from SODP Shape of EEG Signals and Binary PSO.". *Traitement du Signal*, **38**(1), 2021.

[43] H. Akbari et al. "Depression recognition based on the reconstruction of phase space of EEG signals and geometrical features.". *Applied Acoustics*, **179**:108078, 2021.

[44] W. Mumtaz et al. "A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD).". *Medical & biological engineering & computing*, **56**(2):233–246, 2018.