This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2019.2934412, IEEE Transactions on Affective Computing

An Improved Empirical Mode Decomposition of Electroencephalogram Signals for Depression Detection

Jian Shen, Xiaowei Zhang*, Bin Hu, Gang Wang, Zhijie Ding, Bin Hu*, *Senior Member, IEEE*

Abstract—Depression is a mental disorder characterized by persistent low mood that affects a person's thoughts, behavior, feelings, and sense of well-being. According to the World Health Organization (WHO), depression will become the second major life-threatening illness in 2020. Electroencephalogram (EEG) signals, which reflect the working status of human brain, are regarded as the best physiological tool for depression detection. Previous studies used the Empirical Mode Decomposition (EMD) method, which can deal with the highly complex, nonlinear and non-stationary nature of EEG, to extract features from EEG signals. However, for some special data, the neighboring components extracted through EMD could certainly have sections of data carrying the same frequency at different time durations. Thus, the Intrinsic Mode Functions (IMFs) of the data could be linearly dependent and the features coefficients of expansion based on IMFs could not be extracted, which can make the pre-proposed EMD-based feature extraction method impractical. In order to solve this problem, an improved EMD applying Singular Value Decomposition (SVD)-based feature extraction method was proposed in this study, which can extract the features coefficients of expansion based on all IMFs as accurately as possible, ignoring potentially linear dependence of IMFs. Experiments were conducted on four EEG databases for detecting depression. The improved EMD-based feature extraction method can extract feature from all three channels (Fp1, Fpz, and Fp2) on the four EEG databases. The average classification results of the proposed method on the four EEG databases including depressed patients and healthy subjects reached 83.27%,85.19%, 81.98% and 88.07%, respectively, which were comparable with the pre-proposed EMD-based feature extraction method.

Index Terms—depression, Empirical Mode Decomposition, EEG, feature extraction

1 Introduction

According to the World Health Organization (WHO), depression is a leading cause of mental illness and is predicted to become the major life-threatening illness in 2020 [1]. Depression is more common in females than in males, and its growth rate in females and males is 12% and 6.6%, respectively [2–4]. Depression can affect a person's thoughts, behavior, feelings, and sense of well-being.

As reported in [5], one million people commit suicide every year because of depression. Hence, depression can severely threaten human life [6]. Depression is treatable through available effective treatments such as medication, psychological counseling and other clinical methods, but due to ignorance, untimely detection or misdiagnosis, numerous people suffer from depression worldwide [7]. Thus, early detection is critical, which can directly reduce the social and economic pressures related to depression. Current clinical detection of depression is mostly based on self-reporting questionnaire and interview, but there are no objective assessment criteria in existing clinical practice [8]. With the growing number of depressed patients, it is important to trace the effect of treatment through early stage detection and assessments. Therefore, physiological data-based depression detection system could provide objective and rapid detection to enable timely advanced clinical treatment. Physiological data is more objective and accurate for determining the patient's physiological and mental state.

Electroencephalogram (EEG), which reflects the working status of human brain [9, 10], is regarded as the most excellent physiological data that can be used as a tool for the detection and diagnosis of depression. It is safe, low cost, noninvasive and easy to collect EEG across the surface of brain. EEG is widely used in brain function studies. Recently, many studies [7, 11, 12] demonstrated the relationship between depression and EEG. Many researches [11, 13–17] showed asymmetries in EEG of depressed patients over frontal cortex. Moreover, the EEG showed significant differences between healthy subjects and depressed patients in many researches [18–20].

Traditional feature extraction for EEG mostly include the following methods: (1) Fast Fourier transform (FFT) [21, 22]: The main disadvantage of FFT is that it only utilizes frequency information, but ignores time domain information. Nevertheless, a study [23] demonstrated that combining frequency and time domain information can improve the classification performance of EEG. (2)Autoregressive model [24]: It fits non-stationary signals, like EEG. This method

Corresponding Author
Jian Shen, Xiaowei Zhang, Bin Hu are with Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, Lanzhou, China.*
Email: {shenj17, zhangxw, hub17}@lzu.edu.cn
Gang Wang is with Beijing Anding Hospital of Capital Medical University, Beijing, China Email: gangwangdoc@gmail.com
Zhijie Ding is with The Third Peoples Hospital of Tianshui City, Tianshui,China Email: 407493919@qq.com
Bin Hu is with Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, Lanzhou, China. He is with CAS Center for Excellence in Brain Science and Institutes for Biological Sciences, Shanghai Institutes for Biological Sciences, Chines Academy of Sciences, Shanghai, China. And he is with Beijing Institute for Brain Disorders, Capital Medical University, Beijing, China.*
Email: bh@lzu.edu.cn

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2019.2934412, IEEE Transactions on Affective Computing

takes advantage of signal segmentation to estimate the parameters for each segment of EEG signals. (3) Time-frequency Analysis: This method outperformed conventional methods of frequency analysis in [25]. (4) Empirical Mode Decomposition (EMD) [26–28]: EMD is an efficient decomposition method introduced by Huang in recent years. It is suitable to process nonlinear and non-stationary signals and is an adaptive method for decomposing a signal into AM-FM modulated components. EMD can decompose complicated signals into several finite and simple functions, called Intrinsic Mode Functions (IMFs), which reflect the essential physical characteristics in the signal nature.

Due to nonlinear and non-stationary nature of most of physiological signals, EMD is an excellent choice to process physiological signals like EEG, which is one of the most complicated signals. EMD has been used in many fields such as de-noising and signal enhancement [29–31], and feature extraction [27, 28, 32, 33].

An EMD-based feature extraction method [27, 28], which used IMFs of reference signals in different class to represent EEG and regard coefficients of expansion based on all IMFs as a feature vector. The features represent the similarity between EEG and reference signals. Accordingly, the IMFs are the core component of the EMD-based feature extraction method. To solve the inverse of the matrix product of the IMFs and its transpose is the key step of the pre-proposed method. Unfortunately, according to Huang [26], the definition of orthogonality of IMFs is local and for some special data, the neighboring components of IMFs could certainly have sections of data carrying the same frequency at different time durations. Thus, the IMFs of reference signals could be linearly dependent, which can cause that the rank of the matrix product of the IMFs and its transpose is unfilled. In other words, the inverse for the matrix product does not exist. Therefore, the features coefficients of expansion based on all IMFs proposed in [27, 28] could not be extracted or be inaccurately extracted.

To solve the above problem, an improved EMD-based feature extraction method is proposed in this study. Firstly, the improved method applies Singular Value Decomposition (SVD) to decompose the matrix product of the IMFs and its transpose, getting one rectangular diagonal matrix with non-negative real numbers on the diagonal and two unitary matrices. Next, the improved method computes the pseudo-inverse of the diagonal matrix by replacing every non-zero diagonal entry by its reciprocal and transposing the resulting matrix. Then, the improved method calculates the pseudo-inverse of the matrix product of the IMFs and its transpose, which can be used in the following steps of EMD-based feature extraction method, through the pseudo-inverse of the diagonal matrix and the unitary matrices mentioned above. Therefore, the improved method can extract the features coefficients of expansion based on all IMFs as accurately as possible by ignoring the impact of the potentially linear dependence of the IMFs, which ensures the effectiveness of the EMD-based feature extraction method.

The rest of this paper is organized as follows: in Section II, materials and methods are introduced. Afterwards, the improved EMD-based feature extraction method is described in Section III. The experimental results are analyzed and compared in Section IV. Finally, the work is discussed and concluded.

2 Materials and methods

2.1 Data Acquisition

Considering that the prefrontal lobe has a strong correlation with emotional processes and psychiatric disorders [11, 13–15], we collected four EEG databases by three-electrode pervasive EEG collection device which used three electrodes located on the prefrontal lobe (Fp1, Fpz, and Fp2) [34] to detect depression in this study. The location of the three electrodes placement (Fp1, Fpz, and Fp2) and the three-electrode pervasive EEG collection device are shown in Fig. 1 and Fig. 2, respectively.

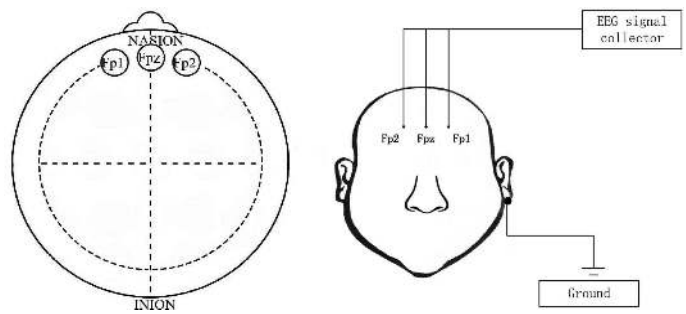

Fig. 1. Location of the three electrodes placement

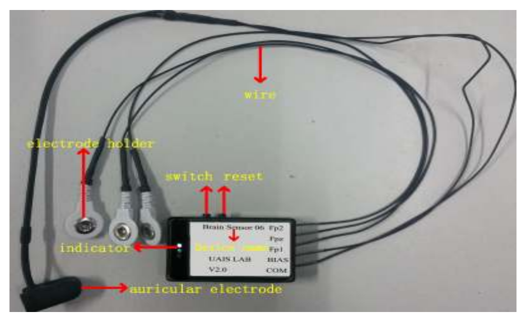

Fig. 2. Three-electrode pervasive EEG collection device

In the data acquisition phase, depressed patients and age-matched, gender-matched, education-matched healthy subjects were selected based on psychological questionnaires and being diagnosed by general practitioners (GPs) from Beijing Anding Hospital, Capital Medical University and The Third People's Hospital of Tianshui City, Tianshui, China. All subjects were limited on aged 18-55 years with normal hearing and intelligence. The educational level of all subjects was higher than primary school. In order to select depressed patients and healthy subjects, different questionnaires were used to evaluate the subjects' psychological status according to their depression severity, psychological irritability, stress levels, etc. The international questionnaires and evaluation standards mainly used were Life Event Scale (LES) [35], Mini-International Neuropsychiatric Interview (MINI) [36], Childhood Trauma Questionnaire (CTQ) [37] and Hamilton Rating Scale for Depression[38]. This study was approved by the Ethics Committee of Beijing Anding Hospital, Capital Medical University and The Third People's Hospital of Tianshui City, Tianshui, China. All subjects were informed about the aims and protocols of the data acquisition experiments before EEG recording.

The four EEG databases were called Dataset 1, Dataset 2, Dataset 3 and Dataset 4 in this study, respectively. EEG signals of Dataset 1 of 81 depressed patients and 89 healthy subjects with the resting state eye-closed was recorded for 40 seconds in Beijing Anding Hospital, Capital Medical University. Dataset 2 collected 160 depressed patients and 116 healthy subjects with the resting state eye-closed for 40 seconds in The Third People's Hospital of Tianshui City, Tianshui, China. Dataset 3 and Dataset 4 were collected based on affective auditory stimulus. Depressed patients were reported to have deficits in negative information sources cognitive processing [39, 40]. Therefore, we selected six affective auditory stimulus with different emotions from the International Affective Digitized Sounds (IADS-2) [41] to investigate the differences between depressed patients and healthy subjects in same affective auditory stimulus. The IADS-2 is a standardized database of 167 naturally occurring sounds that is widely used in the study of emotions. The six affective auditory stimulus and their Self-Assessment Manikin (SAM) [42] in the three affective dimensions of valence, arousal, and dominance are listed in Table 1. Every affective auditory stimuli lasts 6 seconds with subjects in resting state eye-closed. After each affective auditory stimuli, there is 6 seconds rest during which time the EEG signals are also collected. Thus, EEG signals of 105 depressed patients and 109 healthy subjects in Dataset 3 were collected in Beijing Anding Hospital, Capital Medical University. Dataset 4 collected 105 depressed patients and 70 healthy subjects in The Third People's Hospital of Tianshui City, Tianshui, China. Nevertheless, due to noise, the EEG signals during the sixth affective auditory stimuli Crowd and rest stage after Crowd were rejected.

The sampling frequency of the EEG signals was set to 250 Hz with 24 bit A/D convertor precision. All EEG signals were high-pass filtered with 1 Hz cutoff frequency and low-pass filtered with 40Hz cutoff frequency. We used the discrete wavelet transform and Kalman filtering [43] to remove the eye movement.

TABLE 1
The profile of six affective auditory stimulus

Stimulus	Valence		Arousal		Dominance		Property
	M	SD	M	SD	M	SD	
Cattle	5.01	1.85	6.04	1.85	4.56	1.75	neutral
Painting	4.96	1.68	5.37	1.68	5.06	1.82	neutral
Babies Cry	2.04	1.39	6.87	1.39	3.46	2.31	negative
Dentist Drill	2.89	1.67	6.91	1.67	2.92	2.03	negative
Baby	7.61	2.10	6.03	2.10	6.14	1.98	positive
Croud	7.65	1.58	7.12	1.58	6.09	2.18	positive

2.2 EMD Algorithm

EMD [26] can decomposes complicated signals into IMFs. Each IMF has the following specifications:

(1) The number of extrema (maxima and minima) is equal to the number of zero crossings of the signal or differs only by one;

(2) They are locally symmetric and the mean of top and bottom envelope of each IMF is zero.

The sifting process, which is an iterative algorithm, is the decomposition of the original signal into IMFs. The sifting process stops when reaches any of the following criteria:

(1) The residual signal energy becomes less than a predefined threshold;

(2) The residual signal is a monotonic function that cannot be decomposed into more IMFs.

The sifting process can be summarized into the following steps:

(1) Let $i = 1$;

(2) Find all extrema (maxima and minima) of the signal $x(t)$;

(3) Get the envelopes of minima ($e_{min}(t)$) and maxima($e_{max}(t)$) of the signal $x(t)$;

(4) Compute the mean of the maxima and minima envelops: $m(t) = \frac{e_{min}(t)+e_{max}(t)}{2}$;

(5) Compute the difference of the main signal and the mean signal: $h(t) = x(t) - m(t)$;

(6) Continue the steps 2–5 with $h(t)$ as a new signal or stop depending on stop criteria in [26], and let $c_i(t) = h(t)$, where $c_i(t)$ is the IMF. Then continue the process with the residual signal $x(t) - h(t)$ as a new $x(t)$ and increase i by one.

After the sifting process, the original signal can be written as the sum of IMFs and a residue [26]:

$$x(t) = \sum_{i=1}^{N} c_i(t) + r_n \tag{1}$$

The reference signal and its IMFs from Fpz channel of depressed patients in Dataset 3 during the affective auditory stimulus Cattle are shown in Fig. 3.

2.3 EMD-based Feature Extraction

All EEG signals for each channel of depressed patients and healthy subjects in training sets were respectively to obtain two reference signals $R_1(t)$, $R_2(t)$. $R_1(t)$ is the reference signal of depressed patients, and $R_2(t)$ is the reference signal of healthy subjects:

$$R_1(t) = \frac{1}{K_1}\sum_{i=1}^{K_1} x_{1,i}(t), R_2(t) = \frac{1}{K_2}\sum_{i=1}^{K_2} x_{2,i}(t) \tag{2}$$

K_1 and K_2 are the number of depressed patients and healthy subjects in training sets, respectively.

The choice of reference signals $R_1(t)$ and $R_2(t)$ is not same in different datasets. The EEG signals were segmented into 10-second epoch [34] in Dataset 1 and Dataset 2, and for each EEG epoch of all depressed patients and healthy subjects, we calculated the reference signals $R_1(t)$ and $R_2(t)$. Therefore, there are four different reference signals $R_1(t)$ and $R_2(t)$ for different epoches in Dataset 1 and Dataset 2. The EEG of each affective auditory stimuli and rest after each affective auditory stimuli of all depressed patients and healthy subjects were calculated as reference signals $R_1(t)$ and $R_2(t)$, respectively, in Dataset 3 and Dataset 4. Thus, there were 10 different reference signals $R_1(t)$ and $R_2(t)$ for different EEG of stimulus and rest after stimulus in Dataset 3 and Dataset 4.

Then EMD algorithm was applied to decompose the reference data into IMFs [26]:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2019.2934412, IEEE Transactions on Affective Computing

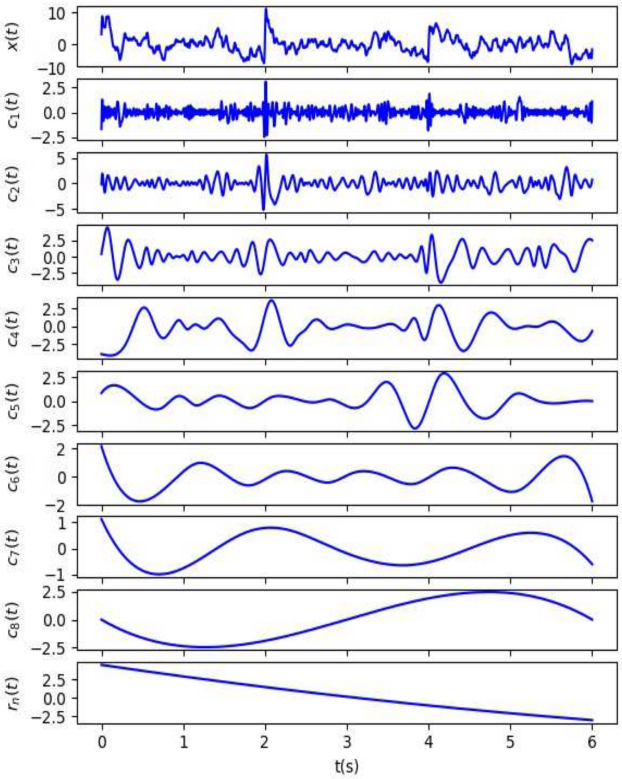

Fig. 3. The reference signal and its IMFs from Fpz channel of depressed patients

$$R_1(t) = \sum_{i=1}^{M_1} c_{1,i}(t) + r_{1,n}(t), \quad R_2(t) = \sum_{i=1}^{M_2} c_{2,i}(t) + r_{2,n}(t) \tag{3}$$

In Equation 3, M_1 and M_2 are the number of the IMFs of $R_1(t)$ and $R_2(t)$ respectively, $c_{1,i}(t)$ and $c_{2,i}(t)$ are the IMFs, $r_{1,n}(t)$ and $r_{2,n}(t)$ are the residues. Thus, each EEG record $x(t)$ can be expanded through $c_{1,i}(t)$ and $r_{1,n}(t)$ or $c_{2,i}(t)$ and $r_{2,n}(t)$:

$$\begin{aligned} x(t) &\simeq \sum_{i=1}^{M_1} b_{1,i}c_{1,i}(t) + b_{1,i+1}r_{1,n}(t) = \hat{x}_1(t) \\ x(t) &\simeq \sum_{i=1}^{M_2} b_{2,i}c_{2,i}(t) + b_{1,i+1}r_{2,n}(t) = \hat{x}_2(t) \end{aligned} \tag{4}$$

In Equation 4, the $b_{1,i+1}$ $(i = 0, 1, \cdots M_1)$ and $b_{2,i+1}$ $(i = 0, 1, \cdots M_2)$ are expansion coefficients based on IMFs and residue of reference signals $R_1(t)$ and $R_2(t)$. The coefficients can be calculated as follows [27]:

$$\begin{aligned} A_1 b_1 &= \hat{x}_1(t) \\ A_2 b_2 &= \hat{x}_2(t) \end{aligned} \tag{5}$$

Where A_1 and A_2 are the matrix of IMFs and residue of reference signals $R_1(t)$ and $R_2(t)$, respectively, b_1 and b_2 are the expansion coefficients:

$$\begin{aligned} A_1 &= \begin{bmatrix} c_{1,1}(1) & c_{1,2}(1) & \cdots & c_{1,M_1}(1) & r_{1,n}(1) \\ c_{1,1}(2) & c_{1,2}(2) & \cdots & c_{1,M_1}(2) & r_{1,n}(1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{1,1}(t) & c_{1,2}(t) & \cdots & c_{1,M_1}(t) & r_{1,n}(1) \end{bmatrix} \\ A_2 &= \begin{bmatrix} c_{2,1}(1) & c_{2,2}(1) & \cdots & c_{2,M_1}(1) & r_{2,n}(1) \\ c_{2,1}(2) & c_{2,2}(2) & \cdots & c_{1,M_2}(2) & r_{2,n}(1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{2,1}(t) & c_{2,2}(t) & \cdots & c_{2,M_2}(t) & r_{2,n}(1) \end{bmatrix} \end{aligned} \tag{6}$$

$$\begin{aligned} b_1 &= \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,M_1} & b_{1,M_1+1} \end{bmatrix} = (A_1^T A_1)^{-1} A_1^T x \\ b_2 &= \begin{bmatrix} b_{2,1} & b_{2,2} & \cdots & b_{2,M_1} & b_{2,M_2+1} \end{bmatrix} = (A_2^T A_2)^{-1} A_2^T x \end{aligned} \tag{7}$$

All the EEG signals of depressed patients and healthy subjects in training sets used the above equations to calculate expansion coefficients b_1 and b_2, which were used as feature for classification.

3 THE IMPROVED EMD BASED FEATURE EXTRACTION

According to Huang [26], the definition of orthogonality of IMFs is locally and for some special data, the neighboring components of IMFs could certainly have sections of data carrying the same frequency at different time durations. Thus, the IMFs of reference signals could be linearly dependent and the features of expansion coefficients based on all IMFs proposed in [27, 28] could not be extracted or inaccurately extracted, which means that the traditional EMD feature extraction method is not suitable for all types of data and has certain limitations.

Since the IMFs of reference signals maybe linearly dependent, the values of $|A_1^T A_1|$ and $|A_2^T A_2|$ in Equation 7 is zero or very near to zero, which would cause the inverse of $A_1^T A_1$ and $A_2^T A_2$ to be non-existent. Thus, the EMD-based feature vectors b_1 and b_2 can not be calculated using Equation 7.

In order to solve the above problem, we proposed an improved EMD-based feature extraction method. The first several steps were the same as Equations 3, 4, 5,and 6 after which $H_1 = A_1^T A_1$ and $H_2 = A_2^T A_1$. Then, calculated as follows:

$$\begin{aligned} rank_1 &= rank(H_1) \\ rank_2 &= rank(H_2) \end{aligned} \tag{8}$$

Where $rank_1$ and $rank_2$ are the rank of square matrices H_1 and H_2. If $rank_1$ is equal to $M_1 + 1$, the b_1 can be calculated using Equation 7. If $rank_1$ is less than $M_1 + 1$, the Equation 7 can be represented as follows:

$$A_1 b_1 = x \tag{9}$$

$$A_1^T A_1 b_1 = A_1^T x \tag{10}$$

The Equation 10 is equal to the following equation:

$$H_1 b_1 = A_1^T x \tag{11}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2019.2934412, IEEE Transactions on Affective Computing

Then, the SVD was introduced to solve the problem:

$$H_1 = U\Sigma_1 V^T \tag{12}$$

The pseudo-inverse of H_1 can be represented as follows:

$$H_1^+ = V\Sigma_1^+ U^T \tag{13}$$

$$H_1^+ H_1 = V\Sigma_1^+ U^T U \Sigma_1 V^T = \Sigma_1^+ \Sigma_1 \simeq E \tag{14}$$

Thus, b_1 can be calculated using Equations 5, 12, 13, 14 and 15:

$$A_1 b_1 = x \tag{15}$$

$$\Rightarrow A_1^T A_1 b_1 = A_1^T x \tag{16}$$

$$\Rightarrow H_1 b_1 = A_1^T x \tag{17}$$

$$\Rightarrow H_1^+ H_1 b_1 = H_1^+ A_1^T x \tag{18}$$

$$\Rightarrow b_1 = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,M_1} & b_{1,M_1+1} \end{bmatrix} = H_1^+ A_1^T x \tag{19}$$

If $rank_2$ is equal to M_2+1, b_2 can be calculated using Equation 7. If $rank_2$ is less than M_1+1, b_2 can be calculated in the same way as b_1:

$$A_2 b_1 = x \tag{20}$$

$$\Rightarrow A_2^T A_2 b_2 = A_2^T x \tag{21}$$

$$\Rightarrow H_2 b_2 = A_2^T x \tag{22}$$

$$\Rightarrow H_2^+ H_2 b_2 = H_2^+ A_2^T x \tag{23}$$

$$\Rightarrow b_2 = \begin{bmatrix} b_{2,1} & b_{2,2} & \cdots & b_{2,M_2} & b_{2,M_1+1} \end{bmatrix} = H_2^+ A_2^T x \tag{24}$$

Algorithm 1 Improved EMD-based feature extraction method

Input: Training EEG signals (the number of subjects is N); label of the signals.
Output: Extracted feature set
Method:

```
Calculate the reference signals R_1 and R_2 of different label respectively
Compute the IMFs and residues of R_1 and R_2 and let A_1 and A_2 be the set of IMFs and residues of R_1 and R_2 respectively
while i<=N do
    Extract the features of i − th subjects: A_1 b_{1,i} = x_i and A_2 b_{2,i} = x_i
    Compute H_1 = A_1^T A_1 and H_2 = A_2^T A_2, and compute rank rank_1 and rank_2 of H_1 and H_2 respectively
    if rank_1 = M_1 + 1 then
        b_{1,i} = (A_1^T A_1)^{-1} A_1^T x_i = H_1^{-1} A_1^T x_i
    else
        b_{1,i} = H_1^+ A_1^T x_i
    end if
    if rank_2 = M_2 + 1 then
        b_{2,i} = (A_1^T A_2)^{-1} A_2^T x_i = H_2^{-1} A_1^T x_i
    else
        b_{2,i} = H_2^+ A_2^T x_i
    end if
    The extracted feature of the EEG signals of i − th subject b_i is the combination of b_{1,i} and b_{2,i}
end while
return the extracted feature set b.
```

The steps of improved EMD-based feature extraction method are listed in Algorithm 1. The proposed improved EMD-based feature extraction method has a good generalization capability that can be applied to extract feature of various signals by ignoring the nonlinear and non-stationary nature. The robustness of the improved EMD-based feature extraction method is better than the traditional EMD method. In many situations and for many types of signals, the improved EMD-based feature extraction method is equivalent to EMD-based feature extraction method (using Equation 7 to extract features). Nevertheless, in some circumstances, especially, for some special signals, the proposed improved EMD-based feature extraction method can extract features that cannot be extracted by traditional EMD-based feature extraction method, as accurately as possible. Consequently, the performance of improved method is equal to the performance of the traditional method at least, but can be better than the performance of the traditional method in some special conditions. According to Algorithm 1, the computational complexity of the proposed improved EMD-based feature extraction method is equal to the computational complexity of traditional method. As compared to traditional method, the proposed method requires a little more time for computing $rank_1$ and $rank_2$ in each iteration. Thus, the proposed method is better than the traditional method in general.

4 Experiments and Results

4.1 Feature Extraction and Classification

In order to validate the effectiveness of the EMD-based feature extraction method, we selected three linear features of the EEG power spectrum: Max frequency, Mean frequency and Centroid frequency and three non-linear features: Permutation entropy [44], Shannon entropy [45] and LZ complexity [46] as traditional features. The first classification experiment was conducted on the four EEG databases using the traditional EMD-based feature extraction method and traditional features. To validate the effectiveness of the improved EMD-based feature extraction method, we compared the performance of the two methods. The second classification experiments were conducted using the effective features extracted through the improved EMD-based feature extraction method on the four EEG databases. In this study, we took advantage of SVM in the two classification experiments.

SVM can better solve small sample, high dimensional, nonlinear and local minima problems, while avoiding curse of dimensionality [47–49] and over-learning. SVM has a good generalization capability [49]. Consequently, in order to analyze the extracted features of the four EEG databases, SVM is an excellent choice to detect the depression. The Kernel RBF was applied in this study:

$$K(x_i, x_j) = \exp\left(\frac{-(x_i - x_j)^2}{2\sigma^2}\right) \tag{25}$$

Cross-validation was used to search the optimum values of σ and C. The optimum values of these parameters were estimated by grid-search using 10-fold cross-validation in this study for the four EEG databases. Trying exponentially

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2019.2934412, IEEE Transactions on Affective Computing

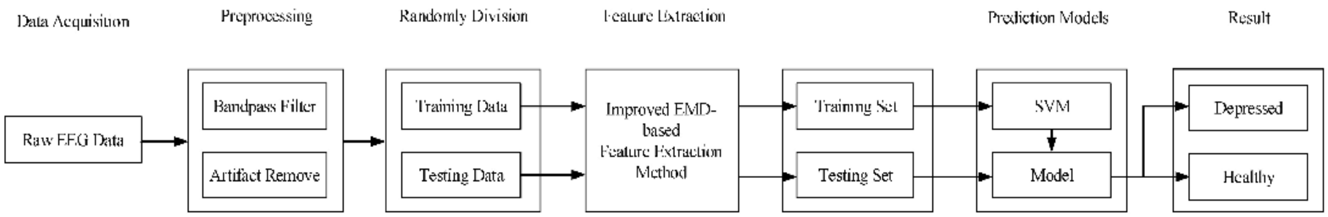

Fig. 4. Flowchart of the proposed depression detection method

growing sequences of C and σ was reported to be a practical method to search optimum values of parameters such as C = [2^{15}, 2^{14}, $\cdots$, 2^{-15}] and σ= [2^{15}, 2^{14}, $\cdots$, 2^{-15}] [50]. Each subject had four feature vectors (EEG signals of all depressed patients and healthy subjects were segmented into 10-second epoch) in Dataset 1 and Dataset 2, so the 10-fold cross-validation by subjects, in which the EEG feature sets were randomly divided into training sets and testing sets by the subject, and the train sets were randomly divided into training data and validation sets by subjects, was applied in this study. Thus, all four feature vectors of one subject can only be in testing sets or validation sets or training sets. The EEG signals of same affective auditory stimulus and same rest after each affective stimulus of all depressed patients and healthy subjects were divided into same set and for each set we took advantage of 10-fold cross-validation in Dataset 3 and Dataset 4.

For the $i-th$ attribute of feature x in training sets, it was normalized into [0,1] by the following Equation:

$$x_i^{new} = \frac{x_i - min_i}{max_i - min_i} \tag{26}$$

Moreover, the $i-th$ attribute of features in testing sets were also normalized through the above Equation, where the min_i and max_i are the minimum value and maximum value of the features of the $i-th$ attribute in corresponding training sets.

The general architecture of our proposed depression detection system is shown in Fig. 4.

The classification experiments were performed in the platform of 3.2 GHz CPU and 8 GB memory and Windows 10 operation system, the SVM was trained using LIBSVM [51] package of Java version.

4.2 Analysis and Comparison of Experimental Results

The average accuracy of traditional features and EMD-based feature extraction method on the four EEG databases shown in Table 2 were obtained in the same condition. According to Table 2, the results of EMD-based method was better than the traditional features since the traditional features only utilize frequency information by ignoring time domain information, whereas the EMD-based feature extraction method can reflect the essential physical characteristic existing in the signal nature, which means that both frequency information and time domain information of the signal are used to extract intrinsic mode features. The first experiment verified the effectiveness of the EMD-based feature extraction method and suggested that it was suitable for this study, as well as verified that the information, whether frequency information or time domain information, of signals can not be ignored. Therefore, while extracting features from signals, we must comprehensively consider their physical characteristics.

TABLE 2
Comparing the classification results of EMD-based feature extraction method and traditional features

	Average Accuracy
EMD	$\mathbf{79.95}\%$
Traditional Features	71.99%

The classification results of improved EMD-based feature extraction method and traditional EMD-based method on Dataset 1 and Dataset 2 of depressed patients and healthy subjects with the resting state eye-closed are shown in Table 3.

TABLE 3
Comparing the classification results of traditional EMD and the improved EMD-based feature extraction method on Dataset 1 and Dataset 2

	Measures	EMD	improved EMD
Dataset 1	Accuracy	82.65%	$\mathbf{83.27}\%$
	Sensitivity	$\mathbf{87.33}\%$	85.54%
	Specificity	77.63%	$\mathbf{82.04}\%$
	C	2^5	2^5
	σ	2^{-3}	2^{-2}
Dataset 2	Accuracy	83.03%	$\mathbf{85.19}\%$
	Sensitivity	89.28%	$\mathbf{89.58}\%$
	Specificity	76.78%	$\mathbf{81.67}\%$
	C	2^{-2}	2^4
	σ	2^{-1}	2^{-2}

As illustrated in Table 3, the measures of the improved EMD-based feature extraction method were more stable and balanced than measures of traditional EMD-based method. The accuracy of the improved EMD-based feature extraction method was better than the traditional EMD-based method. Sensitivity is very important in depression detection, which represents the percentage of depressed patients who are correctly identified as depressed. Thus, with the increase in sensitivity, the missed diagnosis rate of depression detection will reduce. The lower missed diagnosis rate can trace the effect of treatment on more depressed patients as well as save more lives. Specificity represents the percentage of healthy subjects who are correctly identified as healthy. High specificity is useful for detecting depression. With the increase

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2019.2934412, IEEE Transactions on Affective Computing

in specificity, the misdiagnosis rate will decrease as well as the effectiveness of treatment will increase. Therefore, both sensitivity and specificity are important in depression detection. Consequently, the improved EMD-based feature extraction method performed better than the traditional EMD-based feature extraction method on Dataset 1 and Dataset 2.

The classification results on Dataset 3 and Dataset 4 of depressed patients and healthy subjects with affective auditory stimulus are shown in Table 4. The best classification results in Dataset 3 and Dataset 4 were 82.18% and 88.07% during the rest after Dentist Drill and during affective auditory stimulus of Baby, respectively. Fig. 5 represents the box plot of three measures of EMD-based feature extraction method and improved EMD-based feature extraction method. The accuracy of improved EMD-based feature extraction method is better than that of EMD-based feature extraction method. Most of the specificity of improved EMD-based feature extraction method is better than that of EMD-based feature extraction method (the Q1, Q2 and Q3 values of improved EMD-based feature extraction method all are better than that of EMD-based feature extraction method). Moreover, the sensitivity of improved EMD-based feature extraction method was comparable to that of EMD-based feature extraction method. Thus, Table 4 and Fig.5 demonstrated that the improved EMD-based feature extraction method performed better than the traditional EMD-method.

In order to examine the effectiveness of EMD-based method and the improved EMD-based method, we applied Friedman test [52, 53] on Dataset 3 and Dataset 4. Thus, we defined that if at least two measures of one method was better than the other one in Table 4, this method was better than the other one. For example, the accuracy and sensitivity of EMD-based method during the affective auditory stimulus of Dentist Drill in Dataset 3 was better than the improved EMD-based method, thus, the EMD-based method performed better than the improved EMD-based method during the affective auditory stimulus of Dentist Drill in Dataset 3. Then, we calculated the difference between the two methods according to Friedman test. Friedman test revealed that the difference between the two methods was significant ($p < 0.01$) (Table 4), hence the improved EMD-based method is better than the traditional EMD-based method.

The results on Dataset 3 in Table 4 during affective auditory stimulus of Dentist Drill and Baby, and the rest after the two affective auditory stimulus were relatively higher, which was also observed in Dataset 4 in Table 4. Thus, the EEG collected during both negative and positive affective auditory stimulus can be used to detect depression effectively. The EEG responses on both negative and positive affective auditory stimulus of healthy subjects and depressed patients were obviously different. It was reported in [54] that healthy subjects showed greater rostral anterior cingulate cortex (ACC) activity when successfully inhibiting attention to positive stimulus, whereas depressed patients show greater activation when inhibiting attention to negative stimulus, which suggested that healthy subjects require more cognitive effort to divert attention away from positive stimulus, while depressed patients require more

TABLE 4
Comparing the classification results of traditional EMD and the improved EMD-based feature extraction method on Dataset 3 and Dataset 4

	Affective stimulus	Measures	EMD	improved EMD
Dataset 3	Cattle	Accuracy	75.25%	**78.01%**
		Sensitivity	**86.77%**	82.19%
		Specificity	68.45%	**75.08%**
	rest	Accuracy	72.27%	**73.79%**
		Sensitivity	**77.47%**	77.34%
		Specificity	67.42%	**69.13%**
	Painting	Accuracy	78.41%	**78.46%**
		Sensitivity	**84.80%**	78.42%
		Specificity	73.43%	**79.40%**
	rest	Accuracy	78.03%	**81.39%**
		Sensitivity	**87.66%**	80.17%
		Specificity	69.45%	**82.40%**
	Babies Cry	Accuracy	70.54%	**73.38%**
		Sensitivity	71.39%	**76.74%**
		Specificity	69.01%	**69.60%**
	rest	Accuracy	77.44%	**79.02%**
		Sensitivity	85.55%	**87.89%**
		Specificity	70.07%	**70.98%**
	Dentist Drill	Accuracy	**81.34%**	81.27%
		Sensitivity	**84.14%**	83.62%
		Specificity	78.45%	**79.47%**
	rest	Accuracy	**82.18%**	81.98%
		Sensitivity	**84.48%**	82.69%
		Specificity	**79.51%**	77.10%
	Baby	Accuracy	77.59%	**80.78%**
		Sensitivity	81.92%	**83.89%**
		Specificity	73.42%	**77.29%**
	rest	Accuracy	74.80%	**78.87%**
		Sensitivity	78.92%	**82.51%**
		Specificity	69.13%	**73.97%**
Dataset 4	Cattle	Accuracy	**85.62%**	85.13%
		Sensitivity	**93.64%**	86.97%
		Specificity	75.13%	**81.48%**
	rest	Accuracy	82.74%	**86.96%**
		Sensitivity	**92.01%**	87.48%
		Specificity	71.40%	**87.12%**
	Painting	Accuracy	78.92%	**79.54%**
		Sensitivity	86.33%	**88.03%**
		Specificity	**68.57%**	65.93%
	rest	Accuracy	79.90%	**81.08%**
		Sensitivity	86.11%	**89.81%**
		Specificity	**71.52%**	70.06%
	Babies Cry	Accuracy	**79.97%**	75.81%
		Sensitivity	85.08%	**85.73%**
		Specificity	**73.35%**	61.19%
	rest	Accuracy	80.58%	**82.32%**
		Sensitivity	**86.79%**	84.98%
		Specificity	71.53%	**78.51%**
	Dentist Drill	Accuracy	81.56%	**83.49%**
		Sensitivity	**86.30%**	85.14%
		Specificity	75.34%	**79.89%**
	rest	Accuracy	85.71%	**85.75%**
		Sensitivity	89.57%	**90.78%**
		Specificity	**80.74%**	77.98%
	Baby	Accuracy	86.33%	**88.07%**
		Sensitivity	**92.52%**	92.28%
		Specificity	80.83%	**81.64%**
	rest	Accuracy	84.01%	**86.83%**
		Sensitivity	90.01%	**90.42%**
		Specificity	75.51%	**79.65%**

According to the Friedman test, we calculated the $F_{1,19} = 10.69$, which showed $p < 0.01$, so the improved EMD-based method is better than the traditional EMD-based method.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2019.2934412, IEEE Transactions on Affective Computing

cognitive effort to divert attention away from negative stimulus. Therefore, the research [54] correlated with the results acquired in Table 4. Consequently, it can be concluded that both negative and positive affective auditory stimulus can easily and effectively distinguish depressed patients from healthy subjects.

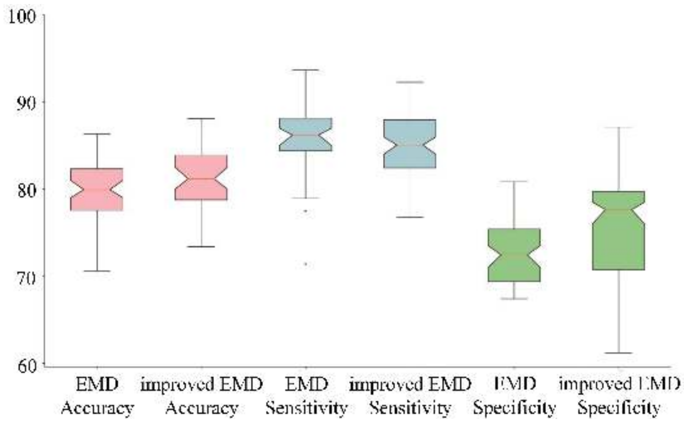

Fig. 5. The measures of traditional EMD and improved EMD-based feature extraction method in Table 4

As illustrated in Table 3 and Table 4, the results on Dataset 2 and Dataset 4 were better than Dataset 1 and Dataset 3, respectively, perhaps due to great regional differences in the incidence of depression [55, 56]. The Dataset 1 and Dataset 3 were collected in Beijing and the Dataset 2 and Dataset 4 were collected in Tianshui. The subjects in Dataset 1 and Dataset 3 suffered great stress everyday which can lead to depression and anxiety tendency, whereas, the subjects in Dataset 2 and Dataset 4 suffered much less stress than the subjects in Beijing. Thus, the EEG difference of depressed patients and healthy subjects in Tianshui was more significant than that in Beijing. According to Table 3 and Table 4, although the epoches in Dataset 4 are only 6-second, much shorter than 10-second epoches in Dataset 3, the results on Dataset 4 are comparable or even better than the results on Dataset 3. This could be because the affective auditory stimulus can induce more abnormal neural activities in depressed patients and healthy subjects, and different groups of subjects require different cognitive effort to divert attention away from different affective auditory stimulus (such as positive or negative auditory stimulus), which can lead to a significant difference on the EEG of depressed patients and healthy subjects. Moreover, the affective auditory stimulus can induce corresponding moods and reaction, which help to distinguish between depressed patients and healthy subjects.

5 Conclusion

This study presented an improved EMD-based EEG feature extraction method and its application in depression detection. EEG is a type of nonlinear, non-stationary and complicated physiological signals. Thus, it is very difficult for many popular feature extraction methods such as FFT to extract excellent features on EEG and many methods ignore the essential physical characteristic existing in the signal nature. EMD is an efficient decomposition method introduced by Huang. The EMD-based feature extraction method performed better than traditional features on Dataset 1. Thus, it is very important to comprehensively consider the physical characteristics of signals while extracting features on complicated signals.

The experimental results in Table 3 and Table 4 illustrated that the improved EMD-based EEG feature extraction method performed better than the traditional EMD-based method. The three measures: accuracy, sensitivity and specificity of the improved EMD-based method were performed better than traditional EMD-based method, indicating that the improved EMD-based method is more suitable in the field of depression detection, and can reduce both missed diagnosis and misdiagnosis rates. After applying the Friedman test on the results of Table 4, the improved EMD-based method was more significant ($p < 0.01$) than the traditional EMD-based method. Therefore, the improved EMD-based feature extraction method should be widely advocated for depression detection to save more lives in the future.

Acknowledgments

The authors are very grateful to the anonymous reviewers for their valuable comments and constructive suggestions which greatly improved the quality of this study. This work was supported by the National Key R&D Program of China(No.2017YFE0111900), the National Basic Research Program of China (973 Program) (No.2014CB744600), the National Natural Science Foundation of China (No.61402211, No.61210010), Program of Beijing Municipal Science & Technology Commission (No.Z171100000117005) and the Fundamental Research Funds for the Central Universities (lzujbky-2017-196, lzujbky-2017-kb08).

References

[1] World Health Organization. From: http://www.who.int/mental health/management/depression/definition/en. 2011.
[2] World Health Organization. World Suicide Prevention Day. From: http://www.who.int/mediacentre/events/annual/world_suicide_prevention_day/en. 2012.
[3] National Alliance on Mental Illness (NAMI). Latina Women and Depression FACT Sheet. 2012.
[4] M. Al Jazaery and G. Guo. Video-Based Depression Level Analysis by Encoding Deep Spatiotemporal Features. *IEEE Transactions on Affective Computing*, pages 1–1, 2018.
[5] U. R. Acharya, V. K. Sudarshan, H. Adeli, J. Santhosh, J. E. Koh, S. D. Puthankatti, and A. Adeli. A Novel Depression Diagnosis Index Using Nonlinear Features in EEG Signals. *European Neurology*, 74(1–2):79–83, 2015.
[6] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Hyett, G. Parker, and M. Breakspear. Multimodal Depression Detection: Fusion Analysis of Paralinguistic, Head Pose and Eye Gaze Behaviors. *IEEE Transactions on Affective Computing*, 9(4):478–490, 2018.
[7] U. R. Acharya, V. K. Sudarshan, H. Adeli, J. Santhosh, J. E. Koh, and A. Adeli. Computer-Aided Diagnosis

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2019.2934412, IEEE Transactions on Affective Computing

of Depression Using EEG Signals. *European Neurology*, 73(5–6):329–336, 2015.

[8] M. Bachmann, J. Lass, and H. Hinrikus. Single channel EEG analysis for detection of depression. *Biomedical Signal Processing & Control*, 31:391–397, 2017.

[9] L. Chen, E. Zhao, D. Wang, Z. Han, S. Zhang, and C. Xu. Feature extraction of EEG signals from epilepsy patients based on Gabor Transform and EMD Decomposition. In *International Conference on Natural Computation*, pages 1243–1247, 2010.

[10] W. Zheng. Multichannel EEG-Based Emotion Recognition via Group Sparse Canonical Correlation Analysis. *IEEE Transactions on Affective Computing*, 9(3):281–290, 2017.

[11] A. C. Deslandes, H. D. Moraes, F. A. M. S. Pompeu, P. Ribeiro, M. Cagy, C. Capitão, H. Alves, R. A. M. Piedade, and J. Laks. Electroencephalographic frontal asymmetry and depressive symptoms in the elderly. *Biological Psychology*, 79(3):317–322, 2008.

[12] H. Hinrikus, A. Suhhova, M. Bachmann, K. Aadamsoo, U. Võhma, H. Pehlak, and J. Lass. Spectral features of EEG in depression. *Biomedizinische Technik Biomedical Engineering*, 55(3):155–161, 2010.

[13] S. Debener, A. Beauducel, D. Nessler, B. Brocke, H. Heilemann, and J. Kayser. Is Resting Anterior EEG Alpha Asymmetry a Trait Marker for Depression? *Neuropsychobiology*, 41(1):31–37, 2000.

[14] Z. J. Koles, J. C. Lind, and P. Florhenry. A source-imaging (low-resolution electromagnetic tomography) study of the EEGs from unmedicated men with schizophrenia. *Psychiatry Research*, 130(2):171–190, 2004.

[15] C. Nissen, B. Feige, E. A. Nofzinger, U. Voderholzer, M. Berger, and D. Riemann. EEG slow wave activity regulation in major depression. *Somnologie*, 10(2):36–42, 2006.

[16] B. Hu, D. Majoe, M. Ratcliffe, Y. Qi, Q. Zhao, H. Peng, D. Fan, F. Zheng, M. Jackson, and P. Moore. EEG-Based Cognitive Interfaces for Ubiquitous Applications: Developments and Challenges. *IEEE Intelligent Systems*, 26(5):46–53, 2011.

[17] J. A. Coan and J. J. Allen. Frontal EEG asymmetry as a moderator and mediator of emotion. *Biological Psychology*, 67(1):7–49, 2004.

[18] V. A. Grin-Yatsenko, I. Baas, V. A. Ponomarev, and J. D. Kropotov. Independent component approach to the analysis of EEG recordings at early stages of depressive disorders. *Clinical Neurophysiology*, 121(3):281–289, 2010.

[19] Y. Kwan, C. Baek, S. Chung, T. H. Kim, and S. Choi. Resting-state quantitative EEG characteristics of insomniac patients with depression. *International Journal of Psychophysiology*, 124:26–32, 2018.

[20] M. Bachmann, J. Lass, and H. Hinrikus. Single channel EEG analysis for detection of depression. *Biomedical Signal Processing & Control*, 31:391–397, 2017.

[21] P. Carr and D. Madan. Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2(4):61–73, 1999.

[22] P. Welch. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73, 1967.

[23] B. D. Mensh, J. Werfer, and H. S. Seung. Combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals. *IEEE Transactions on Biomedical Engineering*, 51(6):1052–1056, 2004.

[24] G. Mohammadi, P. Shoushtari, B. Molaee Ardekani, and M. B. Shamsollahi. Person identification by using AR model for EEG signals. In *Proceeding of World Academy of Science, Engineering and Technology*, volume 11, pages 281–285, 2006.

[25] B. Boashash, M. Mesbah, and P. B. Colditz. Time-frequency detection of EEG abnormalities. *Amsterdam*, pages 663–670, 2003.

[26] N. E. Huang, Z. Shen, S. R. Long, Manli C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. T. Chi, and H. H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings Mathematical Physical & Engineering Sciences*, 454(1971):903–995, 1998.

[27] T. Solis-Escalante, G. G. Gentiletti, and O. Yanez-Suarez. Single trial P300 detection based on the empirical mode decomposition. In *Conference proceedings: IEEE Engineering in Medicine and Biology Society*, pages 1157–1160, 2006.

[28] A. Arasteh, M. H. Moradi, and A. Janghorbani. A Novel Method Based on Empirical Mode Decomposition for P300-Based Detection of Deception. *IEEE Transactions on Information Forensics & Security*, 11(11):2584–2593, 2016.

[29] N. Chatlani and J. J. Soraghan. Adaptive Empirical Mode Decomposition for Signal Enhancement with application to speech. In *International Conference on Systems, Signals and Image Processing*, pages 101–104, 2008.

[30] M. Blanco-Velasco, B. Weng, and K. E. Barner. ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in Biology & Medicine*, 38(1):1–13, 2008.

[31] T. Y. Sun, C. C. Liu, J. H. Jheng, and T. Y. Tsai. An efficient noise reduction algorithm using empirical mode decomposition and correlation measurement. In *International Symposium on Intelligent Signal Processing and Communications Systems*, pages 1–4, 2009.

[32] J. Tang, X. Yang, J. Xu, Y. Tang, Q. Zou, and X. Zhang. The Algorithm of R Peak Detection in ECG Based on Empirical Mode Decomposition. In *Fourth International Conference on Natural Computation*, pages 624–627, 2008.

[33] S. K. Hadjidimitriou and L. J. Hadjileontiadis. EEG-Based Classification of Music Appraisal Responses Using Time-Frequency Analysis and Familiarity Ratings. *IEEE Transactions on Affective Computing*, 4(2):161–172, 2013.

[34] J. Shen, S. Zhao, Y. Yao, Y. Wang, and L. Feng. A novel depression detection method based on pervasive EEG and EEG splitting criterion. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1879–1886, 2017.

[35] B. E. Compas, G. E. Davis, C. J. Forsythe, and B. M.

Wagner. Assessment of major and daily stressful events during adolescence: the Adolescent Perceived Events Scale. *Journal of Consulting & Clinical Psychology*, 55(4):534–541, 1987.

[36] D. V. Sheehan, Y. Lecrubier, K. H. Sheehan, P. Amorim, J. Janavs, E. Weiller, T. Hergueta, R. Baker, and G. C. Dunbar. The Mini-International Neuropsychiatric Interview (M.I.N.I): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, 59(Suppl 20):22–33, 1998.

[37] D. Bernstein, L. Fink, and D. Bernstein. *Childhood trauma questionnaire: A retrospective self-report: Manual*. Psychological Corporation, 1998.

[38] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M. Torres, S. Scherer, G. Stratou, R. Cowie, and M. Pantic. Avec 2016: Depression, mood, and emotion recognition workshop and challenge. In *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, pages 3–10. ACM, 2016.

[39] R. J. Davidson, W. Irwin, M. J. Anderle, and N. H. Kalin. The neural substrates of affective processing in depressed patients treated with venlafaxine. *American Journal of Psychiatry*, 160(1):64–75, 2003.

[40] C. L. Fales, D. M. Barch, M. M. Rundle, M. A. Mintun, A. Z. Snyder, J. D. Cohen, J. Mathews, and Y. I. Sheline. Altered emotional interference processing in affective and cognitive-control brain circuitry in major depression. *Biological Psychiatry*, 63(4):377–384, 2008.

[41] M. M. Bradley and P. J. Lang. The International Affective Digitized Sounds Affective Ratings of Sounds and Instruction Manual. *University of Florida*, 2007.

[42] M. M. Bradley and P. J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, 1994.

[43] Y. Chen, Q. Zhao, B. Hu, J. Li, H. Jiang, W. Lin, Y. Li, S. Zhou, and H. Peng. A method of removing ocular artifacts from EEG using discrete wavelet transform and Kalman filtering. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1485–1492, 2016.

[44] C. Bandt and B. Pompe. Permutation entropy: a natural complexity measure for time series. *Physical Review Letters*, 88(17):174102, 2002.

[45] Shannon C. E. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[46] A. Lempel and J. Ziv. On the Complexity of Finite Sequences. *IEEE Transactions on Information Theory*, 22(1):75–81, 1976.

[47] C. Cortes and V. N. Vapnik. Support-Vector Networks. *Machine Learning*, 20(3):273–297, 1995.

[48] K. P. Burnham and D. R. Anderson. Model Selection and Multimodel Inference. *Technometrics*, 33(2):105–124, 2004.

[49] S. Maldonado, A . Flores, T. Verbraken, B. Baesens, and R. Weber. Profit-based feature selection using support vector machines–General framework and an application for customer retention. *Applied Soft Computing*, 35:740–748, 2015.

[50] C. W. Hsu, C. C. Chang, and C. J. Lin. A practical guide to support vector classification. 2003.

[51] C. C. Chang and C. J. Lin. LIBSVM: a library for support vector machines. *ACM transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

[52] M. Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701, 1937.

[53] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7(1):1–30, 2006.

[54] S. G. Disner, C. G. Beevers, E. A. Haigh, and A. T. Beck. Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience*, 12(8):467, 2011.

[55] World Health Organization et al. Depression and other common mental disorders: global health estimates. 2017.

[56] A. P. Polednak. US regional differences in death rates from depression. *Social Psychiatry & Psychiatric Epidemiology*, 47(12):1977–1983, 2012.

Jian Shen is a PhD Candidate at the Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University since 2017. He received his master's degree from College of Computer Science and Engineering at Northwest Normal University in June 2016. His research interests include Data Mining, Affective Computing and Pervasive Computing.

Xiaowei Zhang received Ph.D. degree in Computer Application Technology from Lanzhou University, China. He is currently an Associate Professor at School of Information Science and Engineering, Lanzhou University, China. He is interested in research fields about Affective Computing, Multi-model Fusion, and Machine Learning.

Bin Hu is a Master degree candidate at the Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University Since September 2016. He received a Bachelor's degree from Department of Nuclear Science and Technology, Lanzhou University in June 2017. His research interests include Multi-model fusion and Affective computing.

Gang Wang received Ph.D. degree in computer science from Institute of mental health, Peking University, China in 2000. Since 2016, he has been the Dean of Beijing Anding Hospital, Capital Medical University. His research interests include Depression detection and Therapeutics of bipolar affective disorder.

Zhijie Ding is director of The Third People's hospital of Tianshui City, archiater of psychiatry. He is vice-director of the Gansu Provincial Medical Association Psychiatry Branch and vice-director of the psychiatric branch of Gansu society of integrated traditional Chinese and Western Medicine.

Bin Hu received Ph.D. degree in computer science from Institute of Computing Technology, Chinese Academy of Science, China in 1998.Since 2008, he has been a professor and the Dean of School of Information Science and Engineering, Lanzhou University, China. He had been also guest professorship in ETH Zurich, Switzerland till 2011.His research interests include Pervasive Computing, Computational Psychophysiology, and Data Modeling.