COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 136 (2016) 151–161

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

EEG-based mild depressive detection using feature selection methods and classifiers

*Xiaowei Li, Bin Hu *, Shuting Sun, Hanshu Cai*

School of Information Science and Engineering, Lanzhou University, Lanzhou, China

A R T I C L E I N F O

Article history:
Received 9 December 2015
Received in revised form
8 August 2016
Accepted 16 August 2016

Keywords:
Mild depression
EEG
Search methods
Classification algorithms

A B S T R A C T

Background and objective: Depression has become a major health burden worldwide, and effectively detection of such disorder is a great challenge which requires latest technological tool, such as Electroencephalography (EEG). This EEG-based research seeks to find prominent frequency band and brain regions that are most related to mild depression, as well as an optimal combination of classification algorithms and feature selection methods which can be used in future mild depression detection.

Methods: An experiment based on facial expression viewing task (Emo_block and Neu_block) was conducted, and EEG data of 37 university students were collected using a 128 channel HydroCel Geodesic Sensor Net (HCGSN). For discriminating mild depressive patients and normal controls, BayesNet (BN), Support Vector Machine (SVM), Logistic Regression (LR), k-nearest neighbor (KNN) and RandomForest (RF) classifiers were used. And BestFirst (BF), GreedyStepwise (GSW), GeneticSearch (GS), LinearForwordSelection (LFS) and RankSearch (RS) based on Correlation Features Selection (CFS) were applied for linear and non-linear EEG features selection. Independent Samples T-test with Bonferroni correction was used to find the significantly discriminant electrodes and features.

Results: Data mining results indicate that optimal performance is achieved using a combination of feature selection method GSW based on CFS and classifier KNN for beta frequency band. Accuracies achieved 92.00% and 98.00%, and AUC achieved 0.957 and 0.997, for Emo_block and Neu_block beta band data respectively. T-test results validate the effectiveness of selected features by search method GSW. Simplified EEG system with only FP1, FP2, F3, O2, T3 electrodes was also explored with linear features, which yielded accuracies of 91.70% and 96.00%, AUC of 0.952 and 0.972, for Emo_block and Neu_block respectively.

Conclusions: Classification results obtained by GSW + KNN are encouraging and better than previously published results. In the spatial distribution of features, we find that left parietotemporal lobe in beta EEG frequency band has greater effect on mild depression detection. And fewer EEG channels (FP1, FP2, F3, O2 and T3) combined with linear features may be good candidates for usage in portable systems for mild depression detection.

* *Corresponding author.* School of Information Science and Engineering, Lanzhou University, Lanzhou, China. Fax: +86 931 8912779.
E-mail address: bh@lzu.edu.cn (B. Hu).

http://dx.doi.org/10.1016/j.cmpb.2016.08.010

1. Introduction

Depression is a common mental disorder that already affects more than 350 million people worldwide [1]. Furthermore, it is estimated by the World Health Organization that depression will become the second leading cause of illness by the year 2020 [2]. In particular, college students suffer from various factors which could lead to depression, poor academic performance, lower socioeconomic status, and stressful tasks such as examinations [3–5]. Ibrahim et al. [5] suggested that the depression rate among university students ranged from 10% to 85%, and found the prevalence of moderate depression in university students from Egypt was as high as 37% [6]. Othieno et al. [3] and Oppong Asante and Andoh-Arthur [7] found similar results in students from universities in Kenyan and Ghana, with prevalences of 35.7% and 31.1% respectively. While research papers were addressing the issue of high occurrence of depression among college students, there were few that provided the solutions. As we all know, current diagnostic techniques of depression have the obvious disadvantages [8], which are associated with patient denial, poor sensitivity, subjective biases and inaccuracy. Hence, one of the most difficult challenges would be to find an easy, accurate and practical method of detection depression. With a deepening sense of urgency, we are exploring such method with EEG.

Electroencephalography (EEG) is an objective and reliable method for the evaluation of brain function which is often used in auxiliary diagnosis of illnesses such as depression [9], seizure and schizophrenia [10]. The advantages of EEG are sensitivity, relatively low-cost and convenience of recording. Henriques and Davidson [11] recorded baseline resting (closed eyes) electroencephalogram activity from 5 clinically depressed and 13 control participants. Analysis of the EEG indicated that depressed subjects had less left-sided activation than normal control subjects. Fingelkurts et al. [12] recorded resting EEG data from 12 severely depressive patients and 10 normal subjects to study the composition of EEG brain oscillations. This study found that severe depression affects brain activity across nearly the whole cortex and embodies in considerable restructuring brain oscillations in a wide range of frequency: 0.5–30 Hz. Furthermore, in recent years, there are many studies using various classification techniques and feature selection methods based on EEG signals to discriminate depressed patients and normal controls. For example, Erguzel et al. [13] used a genetic algorithm (GA) for feature selection and back-propagation neural network (BPNN) for classification which was tested on 147 severely depressed patients with an accuracy of 89.12%. Hosseinifard et al. [14] showed that a combination of a GA for feature selection and a support vector machine (SVM) for classification could achieve an overall accuracy of 88.6%. Hosseinifard et al. [15] used power ratings of four EEG bands and four nonlinear features for classifying 45 depressed patients and 45 normal subjects and applied k-nearest neighbor (KNN), linear discrimination analysis (LDA) and Logistic regression (LR) for classification. This approach achieved the highest classification accuracy of 90% given by a nonlinear feature selection and LR classifier. Spyrou et al. [16] used Random Forest (RF), Random Tree, Multilayer Perceptron (MPL Network) and SVM to identify 34 participants suffering from both cognitive impairment and geriatric depression (mean age 69.81) and 32 control subjects (mean age 70.33) using synchronization and oscillatory features. Results indicated that RF gained the highest accuracy (95.5%). However, these methods mainly aim at detecting depression, there are few studies that provide effective detection means for mild depression, so as to help mild depression take precautions and avoid mild depression evolving into major depression. In this paper we want to find out the more suitable and effective method for mild depression detection.

In addition, several previous studies have demonstrated that depression can affect the ability to recognize different facial emotions [17]. The results suggest that depressed patients pay more attention to facial images expressing negative emotions than neutral emotions, when the images were presented at the same time [18,19]. And relevant researchers [20,21] have indicated that the use of emotional faces expressions rather than words could obtain more consistent results in studies of attention biases in depression. We therefore adopted a similar paradigm to discriminate between mild depressive patients and normal controls. The sensory data from the patients were extracted from 17 EEG features including eight linear features and nine non-linear features from 16 electrodes, and five classifiers were then applied. The challenge is then to find an optimal combination of features and classifiers that can be effective in portable applications with limited channels and processing time.

Firstly we used five classifiers including BayesNet (BN), SVM, LR, KNN, and RF for classifying 10 mildly depressed patients and 10 healthy participants. To eliminate the redundant and less discriminant features, and hence maximize classification performance, we adopted five feature search methods including BestFirst (BF), GreedyStepwise (GSW), GeneticSearch (GS), LinearForwordSelection (LFS) and RankSearch (RS) based on Correlation Features Selection (CFS), for linear and non-linear EEG features selection. We then compared their performance to find the best selection method and classifier combination to identify the mildly depressive patients. A second element of our work is that we evaluated the feature distribution across the two brain hemispheres to identify which brain regions and frequency bands show significant differences between mildly depressed and normal subjects. Then we tried to find the fewer electrodes and features to detect mild depressive patients effectively to apply in real time system and portable devices.

2. Methods

2.1. Subjects

Aimed at the high occurrence of depression in current campus, we did the questionnaire survey for more than 200 undergraduates to record socio-demographic information in Lanzhou University (Lanzhou, Gansu province of China). By questionnaire screening only 37 right-handed volunteers participated in the study. All participants had no prior history of psychopathology and had normal or corrected-to-normal vision. Before the start of the experiment participants were asked to complete the Beck Depression Inventory test-II (BDI-II) [22], BDI scores ranging from 14 to 28 and 0 to 13 were considered to correspond to mildly depressed and normal states respectively. The number of mildly depressed subjects was ten (4 females, 6 males) whose age ranged from 18 to 24 years (Mean = 20.60, SD = 1.74). To obtain an equal number of subjects in each group, we selected 10 subjects (2 females, 8 males) from the normal group, with BDI scores ranging from 6 to 13, and ages corresponding to those of the mildly depressed group. The total number of subjects was therefore 20. All volunteers gave their consent and were rewarded for their participation.

2.2. Experiment stimulus and procedure

The stimuli consisted of 60 facial expression images selected from the Chinese Facial Affective Picture System (CFAPS) [23] which in total has 870 representative facial expression pictures of 7 emotion types, including 74 anger faces, 47 disgust faces, 64 fear faces, 95 sadness faces, 120 surprise faces, 248 happiness faces and 222 neutral facial expressions. So the 15 negative pictures we used were randomly selected from anger, disgust, fear and sadness. The experiment contained two blocks: Neu_block and Emo_block. Each block contained 15 trials giving a total of 30. For the Neu_block each trial contained two pictures of neutral Chinese facial expressions in parallel. For Emo_block each trial contained two pictures of Chinese facial expression, one neutral and one emotional (sadness, angry, disgust, and fear), the location of each picture was presented on the left or right side of the screen randomly. Furthermore, each block was viewed sequentially in its entirety, with each block randomly selected to be viewed first.

Participants sat 60 cm from the screen and viewed the two blocks. Before the beginning of each block, instructions were displayed on the screen to every participant. Each trial was presented for 6 s, followed by a gray background presented for 2 s, hence, participants completed the whole experiment in approximately 5 minutes.

2.3. EEG data recording and preprocessing

The EEG data were recorded with a 128 channel HydroCel Geodesic Sensor Net (HCGSN). 16 electrodes (Fp1, Fp2, F3, F4, F7, F8, C3, C4, T3, T4, P3, P4, T5, T6, O1, O2) were chosen in reference to Cz, according to the standard international 10/20 system. As a note, we used 16 electrodes as opposed to 128 due to time performance and computational efficiency reasons; in addition previous studies [15,24,25] have widely used these electrodes for depression research. The sampling frequency was 250 Hz with Net Station acquisition software and Electrical Geodesics amplifiers. Electrode impedance was kept below 70 kΩ [26].

In our experiment, EEG signals were recorded as stimuli were presented to each participant, trial by trial. Each participant completed both Neu_block and Emo_block, as each block contained 15 trials, this gave a total of 30 trials. EEG signals were divided into 30 segments according to the TTL marks in time series for each subject. Neu_block and Emo_block data were processed separately generating 300, 6 second long, samples from the 20 subjects. To provide noise reduction, high-pass and low-pass filters were set with 0.5 Hz and 70 Hz cutoff frequencies respectively. A notch filter was used to remove 50 Hz frequency noise. The Net Station waveform tool was used to discard Artifacts due to eye movements and muscle activity. Since ocular artifacts (OAs) are presented in the frequency band from 0 to 16 Hz, thus overlapping with the alpha rhythm frequency band of 8–13 Hz. This study therefore used FastICA as an initial stage of de-noising as it has been shown to be effective in delineating overlapping frequency bands [27]. Data processing tool was Matlab R2010a.

2.4. Feature extraction

The EEG signals were filtered with Hanning Filter to extract the three frequency bands, theta (4–8 Hz), alpha (8–13 Hz) and beta (13–30 Hz), and used for further feature extraction. These frequency bands had been confirmed to have great differences between depressive patients and normal controls by many researchers [28–31]. In this study, we adopted adaptive AR (Auto Regressive) model to calculate the PSD (power spectrum density). Max power spectrum density and Sumpower (Appendix A) were based on AR model. Activity, mobility and complexity were calculated based on time-varying Hjorth [32] parameters, and we also calculated other features (variance, meanSquare, etc.). Hence we got 8 linear features in total. Brain is typically a non-linear system. So we extracted 9 non-linear features based on previous research [15,25,33–36], like Approximate entropy (ApEn), Kolmogorov entropy (Kol), Correlation dimension (RC), Lyapunov exponent (LLE), C0-complexity (C0), Lempel–Ziv complexity (LZC), Permutation entropy (Per_en), Singular-value deposition entropy (SVDen) and Spectral. Hence there were 17 EEG features for 16 electrodes, for 3 frequency bands, totaling 816 features.

2.5. Feature selection

Feature selection was used to combat what has been termed “the curse-of-dimensionality” [37]. Previous studies have indicated that if the number of training samples is smaller than the number of feature vectors, then the classifier tends to give poor results [38]. In this study, we have 300 data samples compared to 272 dimensional feature vectors. These numbers are relatively close, so it is vital to identify the significant features that can be enhanced to improve the performance of the classifier.

In this paper, we used five search methods: BestFirst (BF), GreedyStepwise (GSW), GeneticSearch (GS), LinearForwordSelection (LFS) and RankSearch (RS) based on Correlation Features Selection (CFS) [39], hence we obtained five feature selection methods. Weka software [40] was used to implement the search methods as described below.

BestFirst (BF) [41]: Searches the space of attribute subsets by greedy hill climbing augmented with a backtracking facility. Setting the number of consecutive non-improving nodes allowed control of the level of backtracking done. In our study, Best first starts with the empty set of attributes and searches forward (by considering all possible single attributes additions and deletions at a given point).

GreedyStepwise (GSW): Performs a greedy hill climbing forward search through the space of attribute subsets. It starts with no attributes. We selected generateRanking to be true, so a ranked list of attributes was produced by traversing the space from one side to the other and recording the order that attributes are selected.

GeneticSearch (GS) [42]: Performs a search using a simple genetic algorithm.

LinearForwardSelection (LFS) [43]: Extension of BestFirst. Takes a restricted number of k attributes into account. The search uses either the initial ordering to select the top k attributes, or performs a ranking. The search direction is forward.

RankSearch (RS) [44]: Uses an attribute/subset evaluator to rank all attributes. If a subset evaluator is specified, then a forward selection search is used to generate a ranked list. From the ranked list of attributes, subsets of increasing size are evaluated, i.e. the best attribute plus the next best attribute, etc. The best attribute set is reported. RankSearch is linear in the number of attributes because the attribute evaluator used is GainRatioAttributeEval.

2.6. Classification

Since there is no uniform classification method best for all applications, usually it may be useful to test multiple methods. And considering computation time, interpretability, complexity, flexibility of different classification methods and previous applications by other studies [14,15,45,46]. In our experiments, we chose the representative classifiers based on different categories including method based on the graph model of probability (BN), method based on statistical learning theory (SVM), functional method (LR), method based on distance (KNN) and ensemble learning method (RF) to classify EEG data. Prior to classification, all features were normalized in the range[0,1] using the Min–Max algorithm. These classifiers were then used to classify the 10 mildly depressed patients with respect to the 10 normal controls for Neu_block and Emo_block data using the Weka tool with leave-one-out cross validation (LOOCV). The classifiers were set to use their default parameter values as implemented in Weka, except for SVM. We chose linear kernel for SVM and the main reasons are: (1) when the number of samples is less than the number of features, non-linear learning methods do not significantly affect the results and it may be better to simply use linear learning method [47]. (2) To decide whether the nature of the data is linear or non-linear. Generally, if the data size is small, it is better to select linear method to avoid over-fitting [48]. (3) There was evidence [49] indicated that linear SVM had better robustness than polynomial and RBF SVM. And compared with polynomial kernel and RBF kernel for SVM (Appendix B), results proved that SVM with linear kernel outperformed SVM with RBF kernel and polynomial kernel. So SVM with linear kernel was chosen as the basic classifier in this paper.

3. Results and discussion

3.1. Sample characteristics

Means and standard deviations for age, and BDI scores for the mildly depressed and normal group are shown in Table 1. In respect to demographic variables such as age and gender there are no significant differences between the groups (p = ns, ns means difference is not significant). The results show a significant difference in BDI ($p < 0.001$) between the two groups, which infer that our machine learning algorithms were actually predicting mild depression rather than age and gender.

Table 1 – Basic information of mild depression group and control group.

	Mild depression group (n = 10)		Control group (n = 10)		P value
	Mean	SD (range)	Mean	SD (range)	
Age	20.60	1.74 (18–24)	20.30	1.85 (18–24)	p = 0.71 (ns)
Gender	4 Females		2 Females		p = 0.44 (ns)
BDI	18.40	4.29	7.70	1.00	p < 0.001

Boldface indicates Search method GSW is the optimal method.

3.2. Combination of search methods based on CFS and classifiers

Tables 2 and 3 show the mean number of generated features obtained from the five search methods based on CFS from all three EEG frequency bands (816 features), using five classifiers (BN, SVM, LR, KNN and RF). Because the main purpose of our study is to find the prominent EEG frequency band and brain regions to detect mild depression, hence these search methods based on CFS and classifiers were also applied to alpha, beta and theta bands respectively. Both tables indicate that the best search method is GSW which obtained 63 features and 93.93% mean accuracy for Emo_block data and 42 features and 96.27% mean accuracy for Neu_block data, using all features. For each frequency band the best search method is also GSW, in addition, the mean classification accuracies of beta band are better than alpha and theta bands, which suggest that beta band has more relationship with depression state. And previous studies have indicated significant differences between depressive and normal controls. Researchers [28,30] indicated that beta frequency band significantly increases in beta power in depression. However, there is an existing opposite conclusion. Omel'chenko and Zaika [29] found statistically significant decrease in relative beta band in depression compare to controls. For Emo_block and Neu_block the mean accuracies are 88.67% and 90.60% respectively, which are not much less than that of all the features. So the performance evaluation results of combination of five classifiers and five feature selection methods based on Emo_block beta band data and Neu_block beta band data are shown in Figs. 1 and 2.

From Fig. 1 for both Emo_block and Neu_block beta data, we can conclude that the classification accuracy of the combination GSW based on CFS with any one of the classifiers (BN, SVM, LR, KNN and RF) is better than combination the other search methods based on CFS with any one of the classifiers. Particularly GSW + KNN and GSW + RF can obtain the best classification accuracies. From Fig. 2 the area under the receiver operating characteristic (ROC) curve (AUC) of KNN and RF is almost similar, which is more than 0.950. However, the features

Table 2 – Comparison search methods based on CFS of Emo_block data.

Search methods	Alpha (8–13 Hz)		Beta (13–30 Hz)		Theta (4–8 Hz)		All (4–30 Hz)	
	Mean accuracy	Mean features number	Mean accuracy	Mean features number	Mean accuracy	Mean features number	Mean accuracy	Mean features number
None	77.27%	272	82.67%	272	72.93%	272	86.80%	816
BF	80.67%	25	83.47%	29	79.53%	15	91.53%	60
GS	78.73%	115	81.47%	53	76.20%	67	82.47%	103
LFS	79.80%	15	83.27%	14	78.67%	13	90.67%	26
RS	85.07%	80	85.20%	75	78.53%	48	90.20%	55
GSW	**86.20%**	55	**88.67%**	47	**82.53%**	41	**93.93%**	63

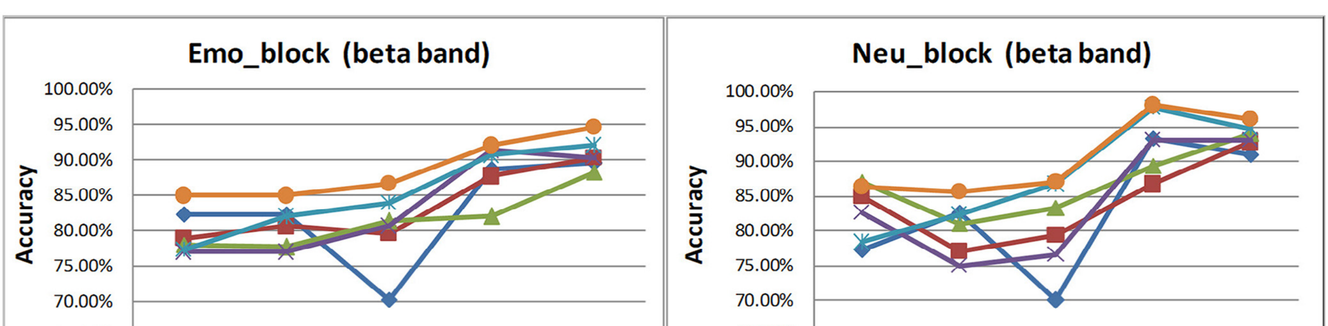
(1) Mean Accuracy: mean accuracy of five classifiers (BN, SVM, LR, KNN and RF) classifying 10 mild depressive patients and 10 normal controls. (2) Mean Features Number: Because each classifier obtains highest accuracy needed features number is different using a search method. Mean Features Number represents the mean number of needed features that five classifiers get highest accuracies using each search method based on CFS.
Boldface indicates Search method GSW is the optimal method.

Table 3 – Comparison search methods based on CFS of Neu_block data.

Search methods	Alpha (8–13 Hz)		Beta (13–30 Hz)		Theta (4–8 Hz)		All (4–30 Hz)	
	Mean accuracy	Mean features number	Mean accuracy	Mean features number	Mean accuracy	Mean features number	Mean accuracy	Mean features number
None	81.27%	272	82.87%	272	71.93%	272	92.47%	816
BF	87.13%	36	84.13%	20	81.67%	25	93.40%	56
GS	84.33%	66	86.93%	117	79.07%	63	86.67%	135
LFS	86.20%	25	84.07%	14	83.13%	19	93.67%	26
RS	86.93%	69	87.93%	106	82.47%	51	93.33%	41
GSW	**88.80%**	37	**90.60%**	58	**85.27%**	29	**96.27%**	42

(1) Mean Accuracy: mean accuracy of five classifiers (BN, SVM, LR, KNN and RF) classifying 10 mild depressive patients and 10 normal controls. (2) Mean Features Number: Because each classifier obtains highest accuracy needed features number is different using a search method. Mean Features Number represents the mean number of needed features that five classifiers get highest accuracies using each search method based on CFS.
Boldface indicates Search method GSW is the optimal method.

number of RF is more than KNN, and the RF is more time consuming than KNN. So GSW + KNN is the optimal combination. The reason why GSW is the best search method is that it has the steepest ascent search, as the search is forced to the far side of the search space by adding the single best attributes at each step. Furthermore, as GSW searches the space of attribute subsets by hill climbing, it is generally more efficient than the other search algorithms, so is more suitable when the



	BN	SVM	LR	KNN(3)	RF
None	82.33%	82.33%	70.33%	88.67%	89.67%
BF	79.00%	80.67%	79.67%	87.67%	90.33%
GS	78.00%	77.67%	81.33%	82.00%	88.33%
LFS	77.00%	77.00%	80.67%	91.33%	90.33%
RS	77.33%	82.00%	84.00%	90.67%	92.00%
GSW	85.00%	85.00%	86.67%	92.00%	94.67%

	BN	SVM	LR	KNN(3)	RF
None	77.33%	82.67%	70.00%	93.33%	91.00%
BF	85.00%	77.00%	79.33%	86.67%	92.67%
GS	87.00%	81.00%	83.33%	89.33%	94.00%
LFS	82.67%	75.00%	76.67%	93.00%	93.00%
RS	78.33%	82.33%	86.67%	97.67%	94.67%
GSW	86.33%	85.67%	87.00%	98.00%	96.00%

Fig. 1 – Combination of five search methods based on CFS and five classifiers based on Emo_block and Neu_block beta data.

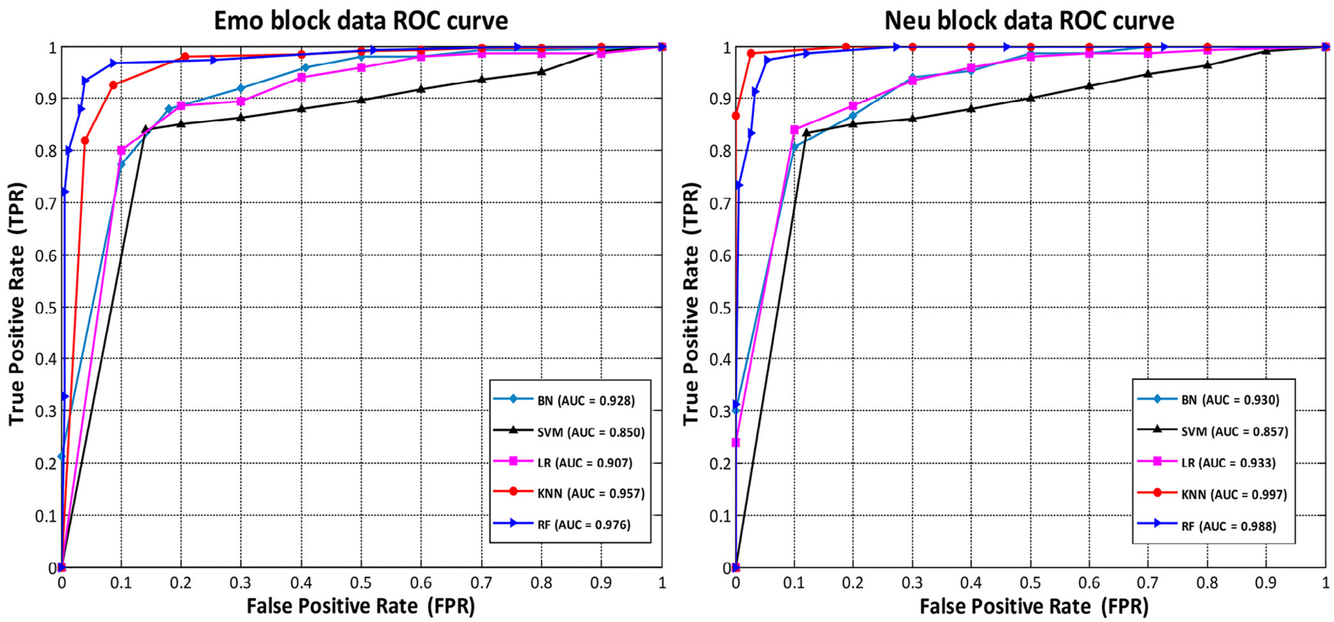


Fig. 2 – ROC Curves of mild depressive patients in five classifiers based on Emo_block and Neu_block beta data.

time is limited to implement a search, such as in real-time systems. It can also return an effective solution even if it is interrupted at any time before it ends. Hence GSW was selected as our search method as it is robust and efficient.

For both normal and mildly depressed subjects we used k = 1, 2, 3 for KNN, and found that k = 3 gave the best performance. It can be seen that KNN achieves the highest accuracy with mean 92.00% and AUC of 0.957 for Emo_block data, 91.30% for mild depressive patients, and with selected features 15. For the Neu_block data, the mean accuracy and AUC also are the highest, with mean accuracy of 98.00% and AUC of 0.997, 98.70% for mild depressive patients, with the number of selected features 84. These accuracies are better than comparable studies using EEG data, they also concur with the general relative performance of the classifiers reported in these studies. For example, Cieslak and Chawla [50] explained that the good performance of KNN was due to its insensitivity to errors produced by non-stationary, which have a ubiquitous presence in EEG signals [51]. Parvin et al. [52]demonstrated that KNN has the ability to cope with discriminant analysis of difficult probability densities which made it effective for classifying EEG data. Yu et al. [53] also found that KNN was the most effective classifier with respect to detecting emotion sickness using EEG data. And it is also worth mentioning that KNN computation time is relatively short. Above all, we consider GSW + KNN as the most effective method to discriminate mild depressive patients and normal controls.

3.3. *Analysis of feature distributions in the brain regions*

We analyzed the feature set (15 features from Emo_block beta data and 84 features from Neu_block beta data obtained by search method GSW based on CFS) distribution across the brain regions in three ways. The results are shown in Fig. 3. For Emo_block LH features number are more than RH features, and especially LPA features number are much more than RPA features number, so we can conclude that electrodes sites P3, T3 and T5 may have greater effect to discriminate mild depressive patients and normal controls (Fig. 3a and 3b). For Neu_block we can consider that LH electrodes may have significant difference to classify mild depressive patients and normal controls, and there is greater difference between LPA and RPA brain regions than between LAN and RAN brain regions (Fig. 3a and 3c), so we also can consider that electrodes P3,T3 and T5 have the significant effect. And through linear and nonlinear features statistics, linear features number is much more than nonlinear features number for both Emo_block and Neu_block, so linear features make greater effect to discriminate mild depressive and normal groups. Previous studies [28,54–56] have demonstrated that linear analysis of EEG could be an efficient method for identifying depressed patients from normal subjects.

3.4. *Statistical analysis results*

In the previous section, for Emo_block and Neu_block beta data we analyzed the features distribution characteristic which achieved highest accuracy obtained by GSW + KNN, however, we cannot clearly point out which electrodes have the greater effect. So independent samples T-test with Bonferroni correction (a method was applied to eliminate false positive errors in case of multiple comparison) is used to find the significantly discriminant electrodes between mild depressive patients and normal controls. The results are shown in Tables 4 and 5. For Emo_block significant difference brain regions are that left frontal lobe (FP1, F3), left temporal lobe (T3), right frontal lobe (FP2), and right occipital lobe (O2) from Table 4. For Neu_block significant difference brain regions are that left frontal lobe (FP1, F3), left parietotemporal lobe (P3, T3), right frontal lobe (FP2), right occipital lobe (O2), and right temporal lobe (T4) from Table 5. These results are almost consistent with the results of Fig. 3. And Li et al. [57] have demonstrated that

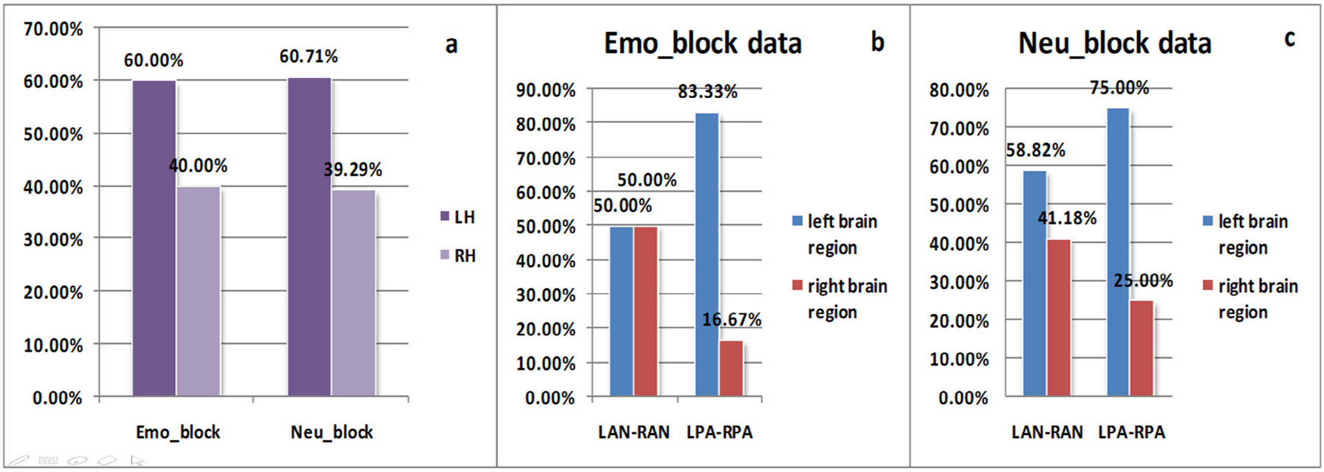


Fig. 3 – Analysis of features distribution across the brain region in three ways, higher values indicate greater differences between mild depressive patients and normal controls (a) left hemisphere (LH) and right hemisphere (RH) features number comparison–with sites paired with (In (left)-In (right)) FP1-FP2, F3-F4, F7-F8, C3-C4, O1-O2, P3-P4, T3-T4 and T5-T6; (b, c) left anterior (LAN) and right anterior (RAN) hemisphere features number comparison–with anterior sites paired with (In (left)-In (right)) FP1-FP2, F3-F4 and F7-F8; and left parietotemporal (LPA) and right parietotemporal (RPA) hemisphere features number comparison–with parietotemporal sites paired with (In (left)-In (right)) P3-P4, T3-T4 and T5-T6.

mild depressed group exhibited a higher activation in temporal lobe of beta band. Koo et al. [58] indicated that compared to healthy controls depressive patients increased beta band power mainly in frontal, central and centroparietal regions, which were often found have the relation to have trouble in falling sleep, and inner restlessness in MDD patients. In addition for both Emo_block and Neu_block, linear features have the significant difference between mild depressive patients and normal controls, which are also consistent with the previous section result. Hence we can conclude that GSW + KNN are the efficient combination to select significant difference features and achieve high accuracy.

3.5. *Simplified system results*

128 channels EEG system, such as the one used in our experiment is an expensive equipment, which can only be available in specialized laboratories or hospitals. It also requires well-trained technicians to operate, since all the electrodes have to be placed in the proper spot. And each complete procedure could take as long as an hour. All these reasons are making these complex systems impractical to large scale. Therefore, a worthwhile EEG system of screening mild depression for large populations, for instance all the university students, has to be a portable system that is cheaper, simpler and easier to use.

Table 4 – Independent samples T-test ($p < 0.05$) of Mild depressive patients and normal controls for Emo_block data, with Bonferroni correction (p value).

	FP1	FP2	F3	F4	C3	C4	P3	P4	O1	O2	F7	F8	T3	T4	T5	T6
ApEn	0.057	0.237	0.015	0.995	0.909	0.785	0.756	0.028	0.000	0.004	0.002	0.780	0.301	0.131	0.000	0.261
C0	0.917	0.905	0.166	0.314	0.894	0.481	0.347	0.729	0.576	0.344	0.430	0.388	0.222	0.928	0.256	0.113
Kol	0.114	0.131	0.628	0.095	0.065	0.481	***0.000***	0.332	0.776	0.598	0.704	0.189	0.004	0.483	0.346	0.291
LLE	0.000	***0.000***	0.001	0.134	0.177	0.082	0.002	0.963	0.042	0.659	0.427	0.781	0.607	0.015	0.108	0.445
LZC	0.052	0.330	0.624	0.735	0.730	0.399	0.361	0.442	0.735	0.029	0.645	0.765	0.498	0.698	0.882	0.008
Per_en	0.026	0.463	0.146	0.601	0.925	0.263	0.749	0.117	0.271	0.855	0.175	0.503	0.687	0.126	0.470	0.011
RC	0.017	***0.000***	0.030	0.041	0.452	0.349	0.020	0.520	0.965	0.090	0.763	0.453	0.317	0.083	0.121	0.194
SVDen	0.793	0.208	0.140	0.444	0.059	0.025	0.022	0.039	0.678	0.853	0.894	0.160	0.165	0.131	0.051	0.007
Spectral	0.053	0.101	0.072	0.742	0.844	0.064	0.049	0.004	0.401	0.961	0.024	0.012	0.084	0.157	0.003	0.036
ppmean	***0.000***	***0.000***	***0.000***	0.002	0.295	***0.000***	0.016	0.194	0.819	***0.000***	0.739	0.003	***0.000***	0.001	0.026	0.002
meanSquare	***0.000***	***0.000***	***0.000***	0.001	0.306	0.256	0.025	0.318	0.498	***0.000***	0.445	0.002	***0.000***	0.007	0.015	0.167
variance	***0.000***	***0.000***	***0.000***	0.001	0.446	0.246	0.029	0.322	0.508	0.002	0.395	0.003	***0.000***	0.006	0.014	0.157
activity	***0.000***	***0.000***	***0.000***	0.010	0.590	0.004	0.008	0.266	0.738	***0.000***	0.957	0.005	***0.000***	0.005	0.013	0.029
mobility	0.001	***0.000***	0.005	0.845	0.781	0.257	0.421	0.819	0.157	0.830	0.003	0.008	0.030	0.037	0.753	0.004
complexity	0.942	0.116	0.078	0.059	0.016	0.317	0.026	0.797	0.257	0.338	0.011	0.064	0.619	0.441	0.270	0.386
maxp	0.004	0.002	***0.000***	0.012	0.516	0.052	0.004	0.006	0.088	***0.000***	0.858	0.003	0.001	0.003	***0.000***	0.024
sumpower	***0.000***	***0.000***	***0.000***	0.002	0.366	0.235	0.030	0.308	0.489	***0.000***	0.335	0.002	***0.000***	0.008	0.017	0.146

(1) Bold, Italic and Underline values represent significant difference values after Bonferroni correction, $p < 0.0002$ (0.05/272). (2) 10th–17th rows are linear features. ppmean: peak to peak amplitude; maxp: max power spectrum density.

Table 5 – Independent samples T-test ($p < 0.05$) of Mild depressive patients and normal controls for Neu_block data, with Bonferroni correction (p value).

	FP1	FP2	F3	F4	C3	C4	P3	P4	O1	O2	F7	F8	T3	T4	T5	T6
ApEn	0.040	0.004	0.475	0.395	0.667	0.223	0.554	0.705	***0.000***	0.020	0.322	0.832	0.450	0.106	0.051	0.167
C0	0.312	0.857	0.067	0.267	0.359	0.278	0.103	0.741	0.749	0.841	0.389	0.220	0.108	0.325	0.470	0.809
Kol	0.004	***0.000***	0.041	0.010	0.013	0.819	***0.000***	0.233	0.045	0.119	0.604	0.051	0.005	0.072	0.004	0.001
LLE	0.085	***0.000***	0.089	0.001	0.419	0.250	0.125	0.045	0.003	0.947	0.087	0.520	0.233	0.832	0.001	0.745
LZC	0.152	0.203	0.424	0.222	0.221	0.643	0.494	0.074	0.485	0.104	0.848	0.223	0.066	0.077	0.142	0.015
Per_en	0.848	0.589	0.781	0.328	0.078	0.619	0.118	0.017	0.234	0.074	0.305	0.719	0.245	0.771	0.172	0.025
RC	0.555	0.018	0.382	0.038	0.148	0.741	0.062	0.024	0.422	0.011	0.105	0.369	0.620	0.202	0.551	0.249
SVDen	0.746	0.752	0.660	0.208	0.514	0.367	0.058	0.268	0.810	0.167	0.403	0.885	0.631	0.745	0.293	0.194
Spectral	0.008	0.102	0.449	0.785	0.059	0.227	0.630	0.320	0.456	0.157	0.716	0.261	0.503	0.114	0.940	***0.000***
ppmean	***0.000***	***0.000***	***0.000***	***0.000***	0.018	0.001	0.170	0.220	0.608	***0.000***	0.484	0.012	***0.000***	***0.000***	0.004	0.032
meanSquare	***0.000***	***0.000***	***0.000***	0.000	0.715	0.001	***0.000***	0.107	0.777	***0.000***	0.250	0.002	***0.000***	***0.000***	0.002	0.025
variance	***0.000***	***0.000***	***0.000***	0.000	0.725	***0.000***	***0.000***	0.206	0.717	***0.000***	0.300	0.001	***0.000***	***0.000***	0.002	0.024
activity	***0.000***	***0.000***	***0.000***	0.004	0.548	0.002	0.016	0.219	0.673	***0.000***	0.494	0.020	***0.000***	***0.000***	0.002	0.023
mobility	0.001	***0.000***	0.001	0.018	0.301	0.042	0.090	0.723	0.748	0.142	0.002	0.028	0.020	0.001	0.726	***0.000***
complexity	0.411	0.158	***0.000***	0.043	0.001	0.230	***0.000***	0.454	0.071	0.332	0.495	0.499	0.029	0.001	0.149	0.701
maxp	***0.000***	0.000	***0.000***	0.017	0.596	0.004	0.003	***0.000***	0.053	***0.000***	0.536	0.040	***0.000***	0.000	***0.000***	***0.000***
sumpower	***0.000***	***0.000***	***0.000***	0.000	0.675	***0.000***	***0.000***	0.115	0.807	***0.000***	0.240	0.002	***0.000***	***0.000***	0.002	0.035

(1) Bold, Italic and Underline values represent significant difference values after Bonferroni correction, $p < 0.0002$ (0.05/272). (2)10th–17th rows are linear features. ppmean: peak to peak amplitude; maxp: max power spectrum density.

Table 6 – Fewer channel electrodes (FP1, FP2, F3, O2, T3) classification accuracy.

	Dep (Recall)	Nor (Recall)	Mean Accuracy	AUC
Emo_block	91.30%	92.00%	91.70%	0.952
Neu_block	96.70%	95.30%	96.00%	0.972

In other words, it must have fewer electrodes. Above results demonstrated the system using our selected 16 electrodes could achieve very high accuracy in differentiating mild depression patients with normal, while reducing the decisive number of electrodes from 128 to 16. However, 16-electrodes system is still a fairly complicated-to-use system.

We further explore the system with only 5 electrodes. Since the brain is a complex system, to ensure the better spatial and temporal analysis performance, we carefully chose the electrodes FP1, FP2, F3, O2 and T3, which had significant difference between mild depressive patients and normal controls. They were also the common electrodes selected from Emo_block and Neu_block beta data. To meet the real-time system requirement, we only chose the linear features of each electrode. The classifier we used was KNN, with $k = 1, 2, 3$, where $k = 3$ gave the best results. The feature selection method was GSW based on CFS. The classification results are shown in Table 6.

Despite only using 5 electrodes, the classification accuracies of our simplified system are 91.70% and 96.00% for the Emo_block and Neu_block respectively. The results are slightly lower than our 16-electrodes system, but still better than previous results [13–15].

4. Conclusion and future work

In this paper we collected EEG data during Emo_block and Neu_block of facial expression viewing. With the purpose of providing a more efficient method for detecting mild depressed patients, we tried a combination of five feature selection methods and five classification algorithms. It was found that GSW based on CFS and KNN had the optimal performance, and beta frequency band played a more important role in detection mild depression than alpha and theta frequency bands, with the classification accuracy above 92% and AUC above 0.950 for beta frequency bands of Emo_block and Neu_block. So this combination was considered as the more efficient, more accurate and more robust method to discriminate mild depressive patients and normal controls. And it was worth mentioning that training time and running time of GSW + KNN was relatively short, which makes it a more suitable choice for implementation of real-time system. We also analyzed the selected feature set distribution across the brain regions, results indicated that left hemisphere; especially left parietotemporal hemisphere brain region had significant differences between mildly depressed and normal subjects.

To meet the real-time system design requirement, we also testified the performance of EEG based depression recognition, with only linear features and fewer electrodes used. 5 channel (FP1, FP2, F3, O2, T3) electrodes were selected on the basis of Independent samples T-test results. The results were encouraging, accuracy was higher than 91% and AUC was above 0.950 for beta frequency bands of Emo_block and Neu_block. Hence the selected electrodes and linear features indicated a good choice for portable device and real-time system.

This work can contribute to mild depression detection of college students and EEG based real-time system implementation. However, one limitation of this study is the sample size for the mild depressed group, which was relatively small, and the other is only two depression levels that were used. To extend our research, in the future, we will increase the number of subjects to improve the validation of results and to classify more depressed states such as severe depression. It is hoped that these findings may have the generalizability to provide an effective approach for auxiliary diagnosis of mild depression and to help mild depressed patients take precautions early.

Acknowledgment

This work was supported by the National Basic Research Program of China (973 Program) (No.2014CB744600), the National Natural Science Foundation of China (grant No.60973138, grant No.61003240), the International Cooperation Project of Ministry of Science and Technology (No.2013DFA11140), the National Basic Research Program of China (973 Program) (No.2011CB711000).

Appendix A

$$P(e^{j\omega}) = \sigma_\omega^2 \left| \frac{1}{1+\sum_{i=1}^{M}\alpha(i)e^{-j\omega i}} \right|^2 \quad (1)$$

$$\text{Max power spectrum} = \max\{p(e^{j\omega}), \omega \in [0, \pi]\} \quad (2)$$

$$\text{Sumpower} = \sum_{\omega} p(e^{j\omega}) \quad (3)$$

Note: $P(e^{j\omega})$ is signal power spectral density of ω, M is the order of AR model, σ_ω^2 is white noise variance, α is the coefficient of AR model.

Appendix B

Table B1 – Comparison classification accuracies of SVM with different kernels using Emo_block data.

Search Methods	SVM (Kernel function type)	Theta	Alpha	Beta	ALL
None	Linear kernel	**73.00%**	**78.33%**	**82.33%**	**83.00%**
	Polynomial kernel	53.33%	71.00%	57.67%	67.00%
	RBF kernel	68.00%	73.00%	55.33%	71.00%
BF	Linear kernel	**73.33%**	**72.33%**	**80.67%**	**88.67%**
	Polynomial kernel	50.67%	64.00%	59.67%	52.33%
	RBF kernel	70.67%	70.00%	65.33%	82.00%
GS	Linear kernel	**72.33%**	**72.00%**	**77.67%**	**80.00%**
	Polynomial kernel	50.67%	66.00%	54.67%	59.00%
	RBF kernel	69.33%	70.33%	60.00%	76.33%
LFS	Linear kernel	**72.67%**	**71.67%**	**77.00%**	**88.00%**
	Polynomial kernel	50.67%	62.67%	51.67%	63.33%
	RBF kernel	71.33%	69.33%	71.33%	84.00%
RS	Linear kernel	**75.00%**	**83.00%**	**82.00%**	**90.00%**
	Polynomial kernel	51.67%	59.67%	54.00%	54.67%
	RBF kernel	71.67%	74.67%	59.67%	68.00%
GSW	Linear kernel	**76.67%**	**83.00%**	**85.00%**	**91.00%**
	Polynomial kernel	57.33%	68.33%	58.33%	74.33%
	RBF kernel	73.00%	75.67%	79.33%	86.33%

Boldface represent that SVM with linear kernel outperformed SVM with RBF kernel and polynomial kernel.

Table B2 – Comparison classification accuracies of SVM with different kernels using Sil_block data.

Search Methods	SVM (Kernel function type)	Theta	Alpha	Beta	ALL
None	Linear kernel	**72.67%**	**84.00%**	**82.67%**	**94.00%**
	Polynomial kernel	56.67%	77.00%	55.00%	63.33%
	RBF kernel	72.67%	80.00%	62.67%	78.33%
BF	Linear kernel	**81.00%**	**86.33%**	**77.00%**	**93.00%**
	Polynomial kernel	53.33%	63.33%	54.67%	74.33%
	RBF kernel	78.00%	84.00%	72.67%	90.33%
GS	Linear kernel	**80.67%**	**82.33%**	**81.00%**	**89.00%**
	Polynomial kernel	53.67%	73.67%	55.00%	65.00%
	RBF kernel	71.67%	77.67%	64.00%	79.33%
LFS	Linear kernel	**81.00%**	**84.00%**	**75.00%**	**94.00%**
	Polynomial kernel	57.33%	60.00%	55.00%	58.33%
	RBF kernel	76.67%	82.00%	71.00%	91.33%
RS	Linear kernel	**82.67%**	**84.00%**	**82.33%**	**94.00%**
	Polynomial kernel	54.33%	54.00%	55.00%	57.33%
	RBF kernel	71.67%	83.00%	69.33%	90.33%
GSW	Linear kernel	**84.67%**	**87.33%**	**85.67%**	**96.00%**
	Polynomial kernel	64.33%	76.00%	60.67%	85.67%
	RBF kernel	78.67%	85.00%	76.00%	92.00%

Boldface represent that SVM with linear kernel outperformed SVM with RBF kernel and polynomial kernel.

REFERENCES

[1] M. Marcus, M.T. Yasamy, M. van Ommeren, Depression: A Global Public Health Concern, WHO Department of Mental Health and Substance Abuse, 2012.

[2] G.H. Brundtland, From the World Health Organization. Mental health: new understanding, new hope, JAMA 286 (19) (2001) 2391.

[3] C.J. Othieno, R.O. Okoth, K. Peltzer, S. Pengpid, L.O. Malla, Depression among university students in Kenya: prevalence and sociodemographic correlates, J. Affect. Disord. 165 (2014) 120–125.

[4] K. Peltzer, S. Pengpid, S. Olowu, M. Olasupo, Depression and associated factors among university students in Western Nigeria, J. Psychol. Afr. 23 (3) (2013) 459–465.

[5] A.K. Ibrahim, S.J. Kelly, C.E. Adams, C. Glazebrook, A systematic review of studies of depression prevalence in university students, J. Psychiatr. Res. 47 (3) (2013) 391–400.

[6] A.K. Ibrahim, S.J. Kelly, C. Glazebrook, Analysis of an Egyptian study on the socioeconomic distribution of depressive symptoms among undergraduates, Soc. Psychiatry Psychiatr. Epidemiol. 47 (6) (2012) 927–937.

[7] K. Oppong Asante, J. Andoh-Arthur, Prevalence and determinants of depressive symptoms among university students in Ghana, J. Affect. Disord. 171c (2014) 161–166.

[8] M. Sung, C. Marci, A. Pentland, Objective Physiological and Behavioral Measures for Identifying and Tracking Depression State in Clinically Depressed Patients, Massachusetts Institute of Technology Media Laboratory, Cambridge, MA, 2005, p. 595. Technical Report (TR).

[9] P. Giannakopoulos, P. Missonnier, E. Kovari, G. Gold, A. Michon, Electrophysiological markers of rapid cognitive decline in mild cognitive impairment, Front. Neurol. Neurosci. 24 (2009) 39–46.

[10] E. Parvinnia, M. Sabeti, M. Zolghadri Jahromi, R. Boostani, Classification of EEG Signals using adaptive weighted distance nearest neighbor algorithm, J. King Saud Univ. Comput. Inf. Sci. 26 (1) (2014) 1–6.

[11] J.B. Henriques, R.J. Davidson, Left frontal hypoactivation in depression, J. Abnorm. Psychol. 100 (4) (1991) 535–545.
[12] A.A. Fingelkurts, A.A. Fingelkurts, H. Rytsala, K. Suominen, E. Isometsa, S. Kahkonen, Composition of brain oscillations in ongoing EEG during major depression disorder, Neurosci. Res. 56 (2) (2006) 133–144.
[13] T.T. Erguzel, S. Ozekes, O. Tan, S. Gultekin, Feature selection and classification of electroencephalographic signals: an artificial neural network and genetic algorithm based approach, Clin. EEG Neurosci. 46 (2015) 321–326.
[14] B. Hosseinifard, M.H. Moradi, R. Rostami, Classifying depression patients and normal subjects using machine learning techniques, in: Electrical Engineering (ICEE), IEEE, Tehran, 2011, pp. 1–4.
[15] B. Hosseinifard, M.H. Moradi, R. Rostami, Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal, Comput. Methods Programs Biomed. 109 (3) (2013) 339–345.
[16] I.-M. Spyrou, C. Frantzidis, C. Bratsas, I. Antoniou, P.D. Bamidis, Geriatric depression symptoms coexisting with cognitive decline: a comparison of classification methodologies, Biomed. Signal Process. Control 25 (2016) 118–129.
[17] M.S. George, T. Huggins, W. McDermut, P.I. Parekh, D. Rubinow, R.M. Post, Abnormal facial emotion recognition in depression: serial testing in an ultra-rapid-cycling patient, Behav. Modif. 22 (2) (1998) 192–204.
[18] I.H. Gotlib, E. Krasnoperova, D.N. Yue, J. Joormann, Attentional biases for negative interpersonal stimuli in clinical depression, J. Abnorm. Psychol. 113 (1) (2004) 121–135.
[19] J.L. Kellough, C.G. Beevers, A.J. Ellis, T.T. Wel, Time course of selective attention in clinically depressed young adults: an eye tracking study, Behav. Res. Ther. 46 (11) (2008) 1238–1243.
[20] M. Rinck, E.S. Becker, A comparison of attentional biases and memory biases in women with social phobia and major depression, J. Abnorm. Psychol. 114 (1) (2005) 62.
[21] I.H. Gotlib, E. Krasnoperova, D.N. Yue, J. Joormann, Attentional biases for negative interpersonal stimuli in clinical depression, J. Abnorm. Psychol. 113 (1) (2004) 127.
[22] A.T. Beck, R.A. Steer, R. Ball, W.F. Ranieri, Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients, J. Pers. Assess. 67 (3) (1996) 588–597.
[23] X. Gong, Y. Huang, Y. Wang, Y. Luo, Revision of the Chinese facial affective picture system, Chin. Ment. Health J. 25 (2011) 40–46.
[24] G.C. Blackhart, J.A. Minnix, J.P. Kline, Can EEG asymmetry patterns predict future development of anxiety and depression? A preliminary study, Biol. Psychol. 72 (1) (2006) 46–50.
[25] S.A. Akar, S. Kara, S. Agambayev, V. Bilgiç, Nonlinear analysis of EEGs of patients with major depression during different emotional states, Comput. Biol. Med. 67 (2015) 49–60.
[26] T.C. Ferree, P. Luu, G.S. Russell, D.M. Tucker, Scalp electrode impedance, infection risk, and EEG data quality, Clin. Neurophysiol. 112 (3) (2001) 536–544.
[27] B. Hu, D. Majoe, M. Ratcliffe, Y. Qi, Q. Zhao, H. Peng, et al., EEG-based cognitive interfaces for ubiquitous applications: developments and challenges, IEEE Intell. Syst. 26 (5) (2011) 46–53.
[28] V. Knott, C. Mahoney, S. Kennedy, K. Evans, EEG power, frequency, asymmetry and coherence in male depression, Psychiatry Res. 106 (2) (2001) 123–140.
[29] V. Omel'chenko, V. Zaika, Changes in the EEG-rhythms in endogenous depressive disorders and the effect of pharmacotherapy, Hum. Physiol. 28 (3) (2002) 275–281.
[30] D. Begić, V. Popović-Knapić, J. Grubišin, B. Kosanović-Rajačić, I. Filipčić, I. Telarović, et al., Quantitative electroencephalography in schizophrenia and depression, Psychiatr. Danub. 23 (4) (2011) 355–362.
[31] J. Ricardo-Garcell, J.J. González-Olvera, E. Miranda, T. Harmony, E. Reyes, L. Almeida, et al., EEG sources in a group of patients with major depressive disorders, Int. J. Psychophysiol. 71 (1) (2009) 70–74.
[32] B. Hjorth, Time domain descriptors and their relation to a particular model for generation of EEG activity, in: CEAN-Computerized EEG Analysis, 1975, pp. 3–8.
[33] F.-Y. Fan, Y.-J. Li, Y.-H. Qiu, Y.-S. Zhu, Use of ANN and complexity measures in cognitive EEG discrimination, in: 27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005, IEEE, 2006.
[34] M. Bachmann, K. Kalev, A. Suhhova, J. Lass, H. Hinrikus, Lempel Ziv complexity of EEG in depression, in: 6th European Conference of the International Federation for Medical and Biological Engineering, Springer, 2015.
[35] Z. Liang, Y. Wang, X. Sun, D. Li, L.J. Voss, J.W. Sleigh, et al., EEG entropy measures in anesthesia, Front. Comput. Neurosci. 9 (2015).
[36] N. Nicolaou, J. Georgiou, Detection of epileptic electroencephalogram based on permutation entropy and support vector machines, Expert Syst. Appl. 39 (1) (2012) 202–209.
[37] J. Elder, D. Pregibon, A statistical perspective on KDD, in: Advances in Knowledge Discovery and Data Mining, 1996, pp. 83–116.
[38] F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, B. Arnaldi, A review of classification algorithms for EEG-based brain-computer interfaces, J. Neural Eng. 4 (2) (2007) R1–R13.
[39] M.A. Hall, L.A. Smith, Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper, in: FLAIRS Conference, 1999.
[40] Weka machine learning tool. http://www.cs.waikato.ac.nz/ml/weka/downloading.html.
[41] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artif. Intell. 97 (1–2) (1997) 273–324.
[42] D.E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, 1989.
[43] M. Gütlein, E. Frank, M. Hall, A. Karwath, Large scale attribute selection using wrappers, in: Proceedings of IEEE Symposium on Computational Intelligence and Data Mining, 2009, pp. 332–339.
[44] M.A. Hall, G. Holmes, Benchmarking attribute selection techniques for discrete class data mining, IEEE Trans. Knowl. Data Eng. 15 (6) (2003) 1437–1447.
[45] A.T. Sohaib, S. Qureshi, J. Hagelbäck, O. Hilborn, P. Jerčić, Evaluating classifiers for emotion recognition using EEG, in: Foundations of Augmented Cognition, Springer., 2013, pp. 492–501.
[46] C. Lehmann, T. Koenig, V. Jelic, L. Prichep, R.E. John, L.-O. Wahlund, et al., Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG), J. Neurosci. Methods 161 (2) (2007) 342–350.
[47] C.-W. Hsu, C.-C. Chang, C.-J. Lin, A Practical Guide to Support Vector Classification, 2003.
[48] M.J. Patel, A. Khalaf, H.J. Aizenstein, Studying depression using imaging and machine learning methods, Neuroimage Clin. 10 (2016) 115–123.
[49] X.-W. Wang, D. Nie, B.-L. Lu, Emotional state classification from EEG data using machine learning approach, Neurocomputing 129 (2014) 94–106.
[50] D.A. Cieslak, N.V. Chawla, A framework for monitoring classifiers' performance: when and why failure occurs?, Knowl. Inf. Syst. 18 (1) (2009) 83–108.

[51] P. Shenoy, M. Krauledat, B. Blankertz, R.P. Rao, K.R. Muller, Towards adaptive classification for BCI, J. Neural Eng. 3 (1) (2006) R13–R23.
[52] H. Parvin, H. Alizadeh, B. Minaei-Bidgoli, MKNN: modified k-nearest neighbor, in: Proceedings of the World Congress on Engineering and Computer Science, Citeseer, 2008.
[53] Y.-H. Yu, P.-C. Lai, L.-W. Ko, C.-H. Chuang, B.-C. Kuo, C.-T. Lin, An EEG-based classification system of Passenger's motion sickness level by using feature extraction/selection technologies, in: The 2010 International Joint Conference on Neural Networks (IJCNN), IEEE, 2010.
[54] S. Mantri, D. Patil, P. Agrawal, V. Wadhai, Non invasive EEG signal processing framework for real time depression analysis, in: SAI Intelligent Systems Conference (IntelliSys), 2015, IEEE, 2015.
[55] H. Alhaj, G. Wisniewski, R. McAllister-Williams, The use of the EEG in measuring therapeutic drug action: focus on depression and antidepressants, J. Psychopharmacol. 25 (2011) 1175–1191.
[56] A. Baskaran, R. Milev, R.S. McIntyre, The neurobiology of the EEG biomarker as a predictor of treatment response in depression, Neuropharmacology 63 (4) (2012) 507–513.
[57] X. Li, B. Hu, T. Xu, J. Shen, M. Ratcliffe, A study on EEG-based brain electrical source of mild depressed subjects, Comput. Methods Programs Biomed. 120 (3) (2015) 135–141.
[58] P. Koo, C. Berger, J. Bartz, P. Wybitul, J. Höppner, P124. QEEG and CSD power analysis in depression, Clin. Neurophysiol. 126 (8) (2015) e151–e152.