Computer Methods and Programs in Biomedicine 164 (2018) 169–179

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Attentional bias in MDD: ERP components analysis and classification using a dot-probe task

Xiaowei Li[a], Jianxiu Li[a], Bin Hu[a,b,c,*], Jing Zhu[a], Xuemin Zhang[d,e,f], Liuqing Wei[d], Ning Zhong[g], Mi Li[g], Zhijie Ding[h], Jing Yang[i], Lan Zhang[i]

[a] Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, China
[b] CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, China
[c] Beijing Institute for Brain Disorders, Capital Medical University, China
[d] Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education (Beijing Normal University), Faculty of Psychology, Beijing Normal University, Beijing, China
[e] State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China
[f] Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, China
[g] International WIC Institute, Beijing University of Technology, Beijing, China
[h] The Third People's Hospital of Tianshui City, Tianshui, China
[i] Department of Child Psychology, Lanzhou University Second Hospital, Lanzhou, China

ARTICLE INFO

Article history:
Received 22 January 2018
Revised 27 June 2018
Accepted 16 July 2018

Keywords:
Major depressive disorder
Attentional bias
Event-related potentials
Feature selection
Classification

ABSTRACT

Background and Objective: Strands of evidence have supported existence of negative attentional bias in patients with depression. This study aimed to assess the behavioral and electrophysiological signatures of attentional bias in major depressive disorder (MDD) and explore whether ERP components contain valuable information for discriminating between MDD patients and healthy controls (HCs).

Methods: Electroencephalography data were collected from 17 patients with MDD and 17 HCs in a dot-probe task, with emotional-neutral pairs as experimental materials. Fourteen features related to ERP waveform shape were generated. Then, Correlated Feature Selection (CFS), ReliefF and GainRatio (GR) were applied for feature selection. For discriminating between MDDs and HCs, *k*-nearest neighbor (KNN), C4.5, Sequential Minimal Optimization (SMO) and Logistic Regression (LR) were used.

Results: Behaviorally, MDD patients showed significantly shorter reaction time (RT) to valid than invalid sad trials, with significantly higher bias score for sad-neutral pairs. Analysis of split-half reliability in RT indices indicated a strong reliability in RT, while coefficients of RT bias scores neared zero. These behavioral effects were supported by ERP results. MDD patients had higher P300 amplitude with the probe replacing a sad face than a neutral face, indicating difficult attention disengagement from negative emotional faces. Meanwhile, data mining analysis based on ERP components suggested that CFS was the best feature selection algorithm. Especially for the P300 induced by valid sad trials, the classification accuracy of CFS combination with any classifier was above 85%, and the KNN ($k=3$) classifier achieved the highest accuracy (94%).

Conclusions: MDD patients show difficulty in attention disengagement from negative stimuli, reflected by P300. The CFS over other methods leads to a good overall performance in most cases, especially when KNN classifier is used for P300 component classification, illustrating that ERP component may be applied as a tool for auxiliary diagnosis of depression.

1. Introduction

Depression is one of the most prevalent psychiatric disorders [1], with an estimated 300 million affected individuals globally (World Health Organization, [2]). Specifically, major depressive disorder (MDD) has become a serious mental health issue. The cardinal features of MDD are abnormal cognitive processing and brain activation in response to emotional information. In particular, indi-

* Corresponding author at: School of Information Science and Engineering, Lanzhou University, Lanzhou, 222 South Tianshui Rd., Lanzhou 730000, China.
E-mail addresses: lixwei@lzu.edu.cn (X. Li), jxli2015@lzu.edu.cn (J. Li), bh@lzu.edu.cn (B. Hu), zhujing@lzu.edu.cn (J. Zhu), xmzhang@bnu.edu.cn (X. Zhang), zhong@maebashi-it.ac.jp (N. Zhong), limi@bjut.edu.cn (M. Li), 407493919@qq.com (Z. Ding).

https://doi.org/10.1016/j.cmpb.2018.07.003
0169-2607/

viduals with MDD tend to exhibit increased attentional processing for negative stimuli relative to positive events, which is considered a core cognitive risk factor for depression [3,4]. Importantly, Clasen et al. [5] found that attentional bias to negative information is indeed associated with the persistence of sad mood in depression. Moreover, depression may be characterized by difficulties in attention disengagement from negative thoughts, memories and events [6,7]. In turn, susceptibility to negative stimuli adversely affects the ability of patients to respond to daily life needs. Understanding the detailed time course of attentional bias could help delineate which stages of affective processing are affected by the negativity bias encountered in MDD.

The dot-probe task has been widely used to examine abnormal attentional bias in depressed subjects [8–10]. It is based on the assumption that individuals tend to respond faster to a probe stimulus presented to an area in the visual field being attended to. Faster reactive times (RTs) to targets replacing emotional faces and slower RTs to probes substituting neutral faces indicate the participant's priority attention to the emotional faces [11]. In a study by Joormann and Gotlib [12], subjects were instructed to identify the locations of targets that appeared at the emotionally neutral and salient sides, respectively, formerly depressed participants showed positive bias scores for sad faces. In addition, Zhao et al. [10], using the dot probe task, found that abstinent heroin addicts (AHAs) have shorter RTs for drug-related stimuli compared with neutral-related stimuli, confirming attentional bias of AHAs in the drug-related context. In contrast, other studies using this task via RT index failed to evidence negative attentional bias [8,9]. The poor reliability of RT indices associated with attentional bias has also been demonstrated by recent studies [9,13], which increases the error risk using the RT index to assess attentional bias in depression. Besides, behavioral measures can only provide a snapshot of attention deployment, rather than explicitly reflecting the distribution of attention over a period of time [14].

1.1. ERP correlates of emotional processing

Electroencephalography provides a direct measure of postsynaptic potentials with millisecond temporal resolution, and can be used to investigate information processing and functional interactions in the human brain [15]. Particularly, the high temporal resolution of event-related potentials (ERPs) compared with alternative brain imaging tools, i.e. functional magnetic resonance imaging (fMRI), represents a great advantage in examining sequential cognitive processing states involved in a task [10]. For instance, early visual components approximately 80–250 ms after stimulus onset, P100 or N170, are particularly associated with perceptual processing, including visual processing of the stimulus and structural encoding of faces [16,17]. A study reported that the N170 is not affected by emotional expressions, supporting the notion that structural encoding and expression analysis are independent processes [18]. However, another study reported a contrasting finding that an early automatic encoding of emotional expressions is reflected by the N170 [19]. The role of the N170 in emotional facial expression cognition remains debatable. The P100 is thought to provide evidence for covert orienting of visuospatial attention [20]. Mingtian et al. [20] found that HCs have significantly large P100 amplitude to valid positive–neutral pairs in the right parietal–occipital region than invalid positive–neutral pairs in the dot-probe task. Indication of abnormal neurophysiological processing in depressive individuals was found in early perception stages covered by decreased P100 latency over right posterior regions to sad faces compared with healthy subjects [21]. These findings support the right hemisphere of the brain is mainly involved in emotional processing. In order to explore visual processing of stimuli in MDD, the P100 component in the early stage was the focus of the current research.

Late ERP components, such as P200, N200 and P300, were thought to reflect higher-order cognitive processes [22,23], especially the P300 component, providing an indicator of attention disengagement from emotional stimuli [24,25]. The peak latency of the P300 at approximately 300 ms after stimulus onset is highest at the centro-parietal sites along the midline [10,21]. Using the dot-probe task, Leutgeb et al. [25] found that phobias exhibit higher P300 to valid probes following a spider picture, but not a neutral image. A larger probe-induced P300 is thought to reflect the need for greater cognitive resources to get rid of the previously involved stimulus location, indicating difficult disengagement from previously involved locations [25]. Importantly, this design is allowed to assess differences in attentional disengagement across emotional stimuli in MDDs and HCs.

1.2. Application based on ERP classification

Classification algorithms for identification have been developed in a variety of applications using ERPs. For example, Sveinsson et al. [26] proposed a new architecture for classification based on ERPs by extracting features from both chronic schizophrenic patients and HCs, with an overall classification accuracy reaching 90%. In [27], a genetic algorithm (GA) with the Fuzzy ARTMAP (FA) classifier was proposed for the classification of alcoholics and non-alcoholics using brain rhythms extracted during visual stimuli, and the best average classification accuracy reached 95.9%. A computer-based classification system was designed for segregating depression patients from HCs using the ERP component P600, which extracts seventeen features related to the waveform shape. After combination with the optimum support vector machine (SVM) classifier, classification accuracy reached 94% [28]. However, attention-related ERP components in the dot-probe task as a tool for auxiliary diagnosis of depression has not been studied.

1.3. Goals of the current study

Therefore, in the current study, a dot-probe task was employed to explore attentional bias to emotional faces in MDD patients. At the behavioral level, we predicted that MDDs would have faster reaction to valid sad stimuli compared with invalid sad stimuli. At the neurological level, we expected MDD patients would demonstrate certain abnormal neurophysiological indicators, which might be manifested as difficult attention disengagement from sad faces by enhanced P300 amplitude to valid sad trials. Consistent with previous studies [9,13], we raised concerns about the reliability of RT indices, and evaluated their split-half reliability. Additionally, whether ERP components contain valuable information for discriminating between MDD patients and healthy controls was examined. Classification accuracy was compared for different feature selection techniques and classifiers, in order to identify the best selection method and classifier combination for detecting MDD.

2. Methods

2.1. Participants

A total of 19 patients diagnosed with MDD and 20 healthy controls were included in the current study. Three HCs were excluded due to heavy artifact contamination of EEG data. To ensure match between MDD patients and HCs in terms of sex, age and education level, we removed two patients with MDD. As a result, 17 MDD patients (age, 32.9 ± 10 years; 8 females and 9 males; education level 11.50 ± 2.9 years) and 17 HCs (age, 29.9 ± 9.3 years; 8 females and 9 males; education level 12.55 ± 2.4 years) were further assessed. All participants were right-handed, with normal or

corrected-to-normal visual acuity. Depression diagnosis was performed by psychiatrists using the Mini International Neuropsychiatric Interview (MINI) [29], and patients who met the Diagnostic and Statistical Manual of Mental Disorders criteria (DSM-IV) for depression were enrolled in this study. Inclusion criteria for depressive patients were: (a) primary education level or above; (b) Patient Health Questionnaire-9 (PHQ-9) score ≥ 5 points; (c) before enrollment, no antipsychotics and antidepressants received in the past two weeks. Exclusion criteria were: (a) past or present experience of psychotic episodes; (b) depressive disorder caused by organic mental disorder; (c) current experience of severe and unstable physical diseases; (d) abuse of alcohol or psychoactive substances in the past year; (e) pregnancy or breastfeeding in women; (f) serious suicidal thoughts; (g) a history of neurological disorders, including epilepsy, head injury, and loss of consciousness; (h) a history of electroconvulsive therapy and magnetic stimulation in the past three months. HC individuals had no brain illness history, psychiatric disorders or medications. Meanwhile, the PHQ-9 was used to assess the severity of depressive symptoms in both groups.

Independent *t*-test was used to assess age, education level, and PHQ-9 score, respectively, while group difference in gender was examined by the Chi-squared test. No significant differences between the two groups were found in age ($t(32) = -0.92$, $p = 0.81$), gender ($\chi^2(1) = 0.75$, $p = 0.50$) and education level ($t(32) = 0.33$, $p = 0.89$). PHQ-9 scores in the MDD group were significantly higher than those of HC individuals (17.9 ± 4.3 for MDDs; 3 ± 2.6 for HCs; $t(32) = -12.22$, $p < 0.005$).

All participants provided signed informed consent before enrolment in this study, which was approved by the Lanzhou University Second Hospital's ethics committee. Each participant received a reimbursement of approximately USD $16 for participation after the experiment.

2.2. Dot-probe task

2.2.1. Emotional face stimuli

Previous studies have demonstrated that people better recognize facial expressions from their own country or race [30]. Therefore, the face stimuli used in this study were obtained from the native Chinese Facial Affective Picture System (CFAPS), which can better fit emotional inducement for Chinese native subjects [31]. The face stimulus set comprised photographs of ten males and ten females displaying two kinds of emotions (happy and sad), as well as 40 neutral faces, yielding a total of 80 expressions. The picture size was 5.16 cm × 5.95 cm, and all non-facial features were trimmed (i.e., no hair or clothing). The MATLAB software was used to equate mean pixel luminance, contrast, and centro-spatial frequency in all faces, whose pictures were converted into 8-bit greyscale images.

Each sad or happy face was paired with a neutral face, creating two emotional blocks: happy-neutral and sad-neutral faces. Thus, we obtained emotional-face pair stimuli consisting of 20 happy-neural and 20 sad-neutral faces. Each block had 160 trials for a total of 480 trials. These face pairs acting as different valences randomly appeared in the left or right side of the screen, with equal number of occurrences in each block, as did the targets.

2.2.2. Design and implementation of the dot-probe task

Here we applied a dot-probe task to assess differences in behavior and neural responses to attentional bias between MDD patients and HCs. IBM compatible computers and DELL 17-I with 1024 × 768 pixels resolution and a refresh rate of 75 Hz were used to present this task, which was programmed with the E-Prime version 2.0 software (Psychology Software Tools Inc., Pittsburgh PA, USA). Stimulus presentation was synchronized with the screen refresh rate.

The typical sequence of events is presented in Fig. 1. At the beginning of each trial, a central-fixation cross was displayed for 300–600 ms in the middle of the black screen. The central-fixation cross subtended a visual angle of 1.6° × 1.61°(1.9 cm × 1.9 cm) at a 60 cm viewing distance. Followed by the emergence of CFAPS face pairs (happy-neutral or sad-neutral), a face appeared on the left and another on the right of the central-fixation cross for 500 ms, and the distance between the fixation cross and the center of each face was 5 cm. After randomly selected intervals of 100, 150, 200, 250, or 300 ms of central-fixation cross, a probe with a radius of 1.5 cm appeared in the center of one of the previously presented faces for 150 ms. The participants were instructed to press keys as quickly and accurately as possible to determine the probe' location within a specific time of 2000 ms. Pressing the response key indicated that the response window was terminated; inter-trial interval was 600 ms.

During the whole trial, the participants were required to maintain visual fixation on the fixation cross, so as to minimize eye movement artifacts during EEG recording. Before the real experiment, the subjects were required to complete 10 practice trials. The experiment took place in a sound attenuated, dark and air conditioned room.

2.3. EEG data acquisition and preprocessing

EEG was recorded at a sampling rate of 250 Hz with the Net Station software (version 4.5.4) on a 128-channel HydroCel Geodesic Sensor Net. The scalp impedance of each sensor was kept below 70kΩ, and EEG data were re-referenced off-line to the average of all electrodes.

Signal processing and analysis of EEG data was performed with the EEGLAB [32] and ERPLAB [33] toolboxes in MATLAB. EEG signals were filtered by a high-pass filter with a cutoff frequency of 0.3 Hz and a low-pass filter with a cutoff frequency of 30 Hz. A small section of EEG data was excluded by view inspection, including swallowing and coughing, which resulted in large artifacts. In addition, independent component analysis (ICA) was computed for all EEG channels to remove ocular and muscle artifacts [34]. ICs associated with eye movements and blinking activities were identified by visual inspection as frontopolar maxima for blinks/vertical saccades and lateral frontal maxima with different polarities for horizontal saccades.

2.4. Analysis of ERP components

To assess the ERP components, ICA-corrected EEG data were segmented for each trial from 100 ms pre-stimulus to 600 ms post-stimulus, and baseline corrected using 100 ms pre-stimulus data. Approximately 8% of all trials for each emotional face in valid and invalid conditions were excluded because of large movement artifacts (signal exceeding ± 100μV). Statistical analysis showed no significant difference in the number of trials remaining for each condition ($p > 0.05$).

This study examined two ERP components associated with attention processing, including cue-evoked P100 and target-evoked P300. P100 amplitude is determined by averaging the amplitudes of electrodes near the left and right occipital regions [20]. Therefore, in the present study, mean P100 amplitude was obtained by averaging left and right parieto-occipital electrode data (PO7/O1 and PO8/O2) between 80–160 ms after cue presentation [20,35]. The P300 was defined at the electrodes Cz and Pz with a time window between 300 and 600 ms after stimulus presentation [21]. Since the cue-evoked response lasted relatively long, the early components of the target-evoked response were severely distorted by the cue-evoked response. Therefore, early ERP components elicited by target stimuli were not considered in this study.

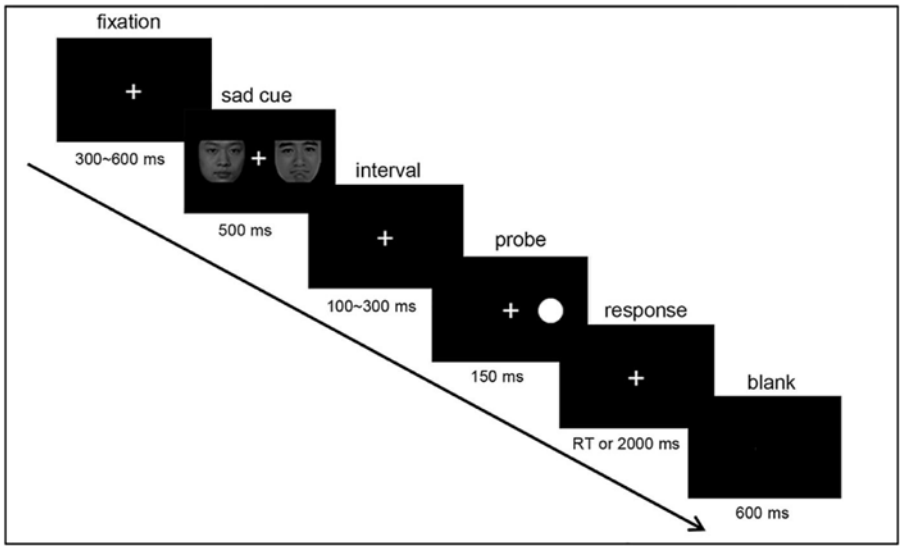

Fig. 1. A typical trial for the dot probe task.

2.5. Statistical analyses

2.5.1. Behavioral data

Reaction time (RT) was defined as response latency from the probe onset. To reduce the impact of outliers, RTs more than 1000 ms or less than 100 ms (or incorrect) were eliminated from the current analysis [12,36]. Behavioral data were analyzed by repeated-measures ANOVA with group (MDD patients vs. HCs) as between-subjects factor and two within-subjects factors, including type (happy vs. sad) and validity (valid, emotional face and probe appeared on same side vs. invalid, probe appeared in a location different from the emotional face).

Moreover, another measure of attentional bias is to calculate bias scores of emotional faces using the following equation [13]:

$$Attentionalbiasscore = 1/2[(RpLe - RpRe) + (LpRe - LpLe)], \quad (1)$$

where R and L are the right and the left sides of the central-fixation cross respectively, p is the probe, and e is the emotional face. Thus, RpLe represents average latency when the probe appears in the right position with the emotional face in the left position, for example. Positive bias scores suggest preferential attention to emotional faces, while negative ones indicate avoidance of emotional faces.

2.5.2. ERP data

To evaluate the characteristics of attentional bias in MDD, ERP components were compared between MDD patients and HCs. In particular, the analyses mainly focused on the P100 and P300 components at the early and late stage stages, respectively. The mean amplitude and latency of P100 were submitted to assessment by repeated-measures ANOVA with group, type and hemisphere (left vs. right) as factors, while repeated-measures ANOVA was conducted on mean amplitude and latency of P300, with four factors, including group, type, validity and electrode site (Cz vs. Pz).

2.6. Classification methodology

2.6.1. Feature generation

Scalp topography showed that the main P100 and P300 activation areas were in the occipital area and centro-parietal sites, respectively. A total of 14 electrodes from the occipital area and centro-parietal sites were selected, including PO3, PO4, PO7, PO8, O1, O2, and Oz for the P100 component, and C2, C1, Cz, CPz, P1, P2 and Pz for the P300 component. For ERP average in each subject, we extracted the following 14 time domain features for the P100 (at 80–160 ms time interval) and P300 (at 300–600 ms time interval) components.

(1) Latency (LAT, $t_{s_{\max}}$)––the ERP's latency time, i.e. the time where the maximum signal value occurs:

$$t_{s_{\max}} = \{t | s(t) = s_{\max}\} \quad (2)$$

(2) Amplitude (AMP, $s_{\max}$)––the maximum signal value:

$$s_{\max} = \max\{s(t)\} \quad (3)$$

(3) Latency/amplitude ratio (LAR, $t_{s_{\max}}/s_{\max}$).

(4) Absolute amplitude (AAMP, $|s_{\max}|$).

(5) Absolute latency/amplitude ratio (ALAR, $|t_{s_{\max}}/s_{\max}|$).

(6) Positive area (PAR, A_P)––the sum of positive signal values for the P100 component from 80 to 160 ms and for the P300 component from 300 to 600 ms:

$$A_{P1} = \sum_{t=80ms}^{160ms} 0.5(s(t) + |s(t)|) \quad (4)$$

$$A_{P3} = \sum_{t=300ms}^{600ms} 0.5(s(t) + |s(t)|) \quad (5)$$

(7) Negative area (NAR, A_n)––the sum of negative signal values for the P100 component from 80 to 160 ms and for the P300 component from 300 to 600 ms:

$$A_{n1} = \sum_{t=80ms}^{160ms} 0.5(s(t) - |s(t)|) \quad (6)$$

$$A_{n3} = \sum_{t=300ms}^{600ms} 0.5(s(t) - |s(t)|) \quad (7)$$

(8) Total area (TAR, A_{Pn}):

$$A_{Pn} = A_P + A_n \quad (8)$$

(9) Mean of peak-to-peak (PP, PPmean):

$$PPmean = (s_{\max} - s_{\min})/2 \quad (9)$$

where $s_{\max}$ and $s_{\min}$ are maximum and minimum signal values, respectively:

$$s_{\max} = \max\{s(t)\}, s_{\min} = \min\{s(t)\} \quad (10)$$

(10) Mean square:

$$meanSquare=\frac{1}{N}\sum_{t=1}^{N}s(t)^2 \tag{11}$$

(11) Variance:

$$Variance=\frac{1}{N-1}\sum_{t=1}^{N}\left(s(t)-\overline{s(t)}\right)^2 \tag{12}$$

(12) The Hjorth parameter, in the time domain analysis method [37], defines three time domain parameters, including activity, mobility and complexity corresponding to the amplitude, slope and rate of change, which is characteristic of EEG data in the time domain. The formula for n moments calculations are as follows:

$$m_0=\int_{-\infty}^{+\infty}s(\varpi)d\varpi=\frac{1}{T}\int_{-\infty}^{+\infty}s^2(t)dt \tag{13}$$

$$m_2=\int_{-\infty}^{+\infty}\varpi^2 s(\varpi)d\varpi=\frac{1}{T}\int_{t-T}^{t}\left(\frac{ds}{dt}\right)^2 dt \tag{14}$$

$$m_4=\int_{-\infty}^{+\infty}\varpi^4 s(\varpi)d\varpi=\frac{1}{T}\int_{t-T}^{t}\left(\frac{d^2s}{dt^2}\right)^2 dt \tag{15}$$

where $s(\varpi)$ represents the density spectral function of energy, $s(t)$ is the EEG signal of a period of time, and m_i represents the i moment. Then, the Hjorth parameters are derived as:

$$Activity=m_0 \tag{16}$$

$$Mobility=\sqrt{m_2/m_0} \tag{17}$$

$$Complexity=\sqrt{m_4/m_2-m_2/m_0} \tag{18}$$

Activity, mobility and complexity were extracted as features in the study. So, each participant had 1176 features (14 electrodes × 14 features × 6 stimulus conditions). Six stimulus conditions consisted of happy-neutral and sad-neutral pairs for the P100 component, as well as happy-neutral valid, happy-neutral invalid, sad-neutral valid and sad-neutral invalid trials for the P300 component. All the features were normalized, i.e. the obtained data were mapped to the [0, 1] interval. The conversion function was as follows:

$$s(t)'=(s(t)-s_{\min})/(s_{\max}-s_{\min}) \tag{19}$$

2.6.2. Feature selection

As a preprocessing step of data mining, feature selection is effective in reducing dimensionality, removing irrelevant data, increasing learning accuracy and improving comprehensibility. In this study, 34 data samples with 98 dimensional feature vectors were obtained for each condition of each ERP component. The high feature dimension affects the efficiency of abnormal detection to a great extent. Therefore, we chose three feature selection algorithms, including Correlated Feature Selection (CFS) [38], ReliefF [39], and GainRatio (GR) [40] to reduce the dimension of feature space.

The Correlation Feature Selection (CFS) algorithm evaluates subsets of features based on the hypothesis that "Good feature subsets contain features that are highly correlated with the class, yet uncorrelated to each other". CFS first calculates the correlation matrix of feature-class and feature-feature from the training data set; the GreedyStepwise search method was used in this study. According to evaluation criteria, the impact of each feature on classification was defined as follows:

$$Merit_s(k)=\frac{k\overline{r_{cf}}}{\sqrt{k+k(k-1)\overline{r_{ff}}}} \tag{20}$$

where $Merit_S(k)$ represents an evaluation of S, a subset of features that contain k features; $\overline{r_{cf}}$ is the average feature-class correlation; $\overline{r_{ff}}$ is the average feature-feature inter-correlation.

The ReliefF algorithm, a well-known distance-based feature ranking method, was first developed by Robnik-Šikonja and Kononenko [39]. Its main idea is to rank features in terms of distance by searching nearest neighbors of samples for each class and then weighting features in terms of how well they differentiate samples from different classes. For each target sample, when the weight is less than a certain threshold, the target feature is removed. As an important machine learning algorithm, ReliefF has a wide range of applications in high-throughput data feature selection, and classification.

The GainRatio (GR) algorithm is an extension of information gain, proposed by Quinlan in [40]. It overcomes the shortcomings of information gain, which bias the selection of attributes with many different values. The GR uses $SplitInfo_F(S)$ to normalize information gain:

$$SplitInfo_F(S)=-\sum_{i=1}^{v}\frac{|Sv|}{|S|}*\log_2\left(\frac{|Sv|}{|S|}\right) \tag{21}$$

This value represents the potential information generated by decomposing the training data set S into v partitions, corresponding to v outcomes of a test on the feature F. The GR is obtained as:

$$GainRatio(F,S)=\frac{Gain(F)}{SplitInfo_F(S)} \tag{22}$$

In (22), the ratio becomes unstable when $SplitInfo_F(S)$ approaches zero. In order to avoid this situation, information gain of the selected test must be at least as large as the average gain of all the tests examined. Eventually, feature with the maximum gain ratio is selected as the split feature.

2.6.3. Classifiers

Because there is no uniform classification for all applications, testing multiple methods is often useful. Taking into account time, interpretation, complexity, and flexibility of different classification methods as well as other research applications [40,41], four famous and most frequently used classifiers, including k-nearest neighbor (KNN), C4.5, Sequential Minimal Optimization (SMO) and Logistic Regression (LR), were adopted to generate classification accuracy for the datasets with selected features obtained by different feature selection methods. Weka (Waikato environment for knowledge analysis) was selected as the classification platform; it is publicly available online at http://www.cs.waikato.ac.nz/ml/weka. In the k-nearest neighbor method, k was set as 1, 2, 3, 4 and 5, respectively, to find which k value could achieve the best classification performance. For other classifiers, we used default parameter settings.

We used Leave-One-Out Cross Validation (LOOCV) to assess classification performance. On each fold of the LOOCV, all but one of the samples was used to train the classifier, while the remaining sample was saved for validating the results. This procedure proceeded until each sample had performed once as the test set. Since no random sampling is involved its results are deterministic [42]. This procedure is an attractive one for three reasons. Firstly, LOOCV allows us to train with the greatest amount of data each time, which increases the chance that the classifier is accurate. Secondly, LOOCV can eliminate the influence of random factors, to ensure the

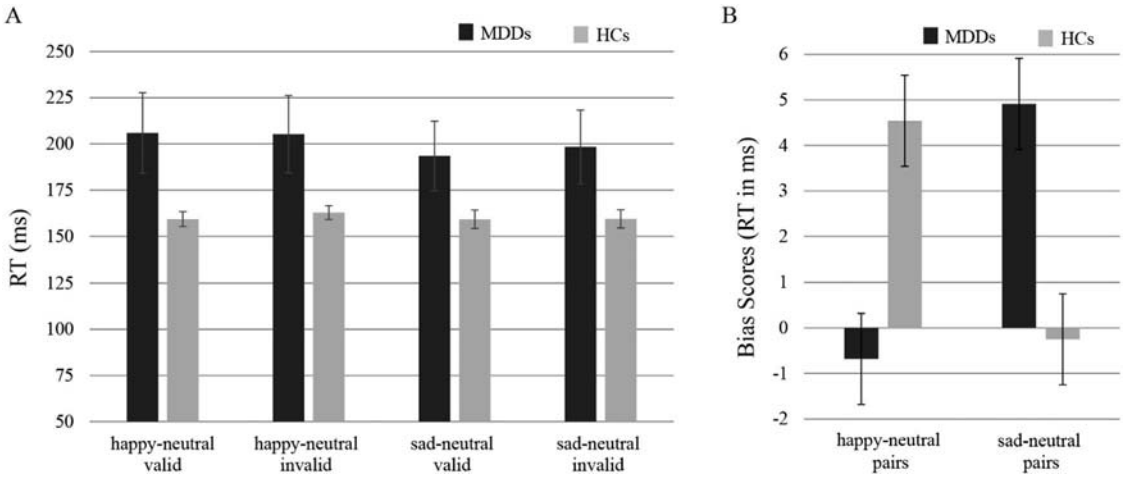

Fig. 2. Means and standard errors of reaction times (A) and attentional bias scores (B) in the dot-probe task, for MDD patients and HCs.

Table 1
Split-half reliability measures of each trial type in MDD patients and HCs.

	Reaction times			Bias scores		
Group	Happy-neutral valid	Happy-neutral invalid	Sad-neutral valid	Sad-neutral invalid	Happy-neutral pairs	Sad-neutral pairs
MDDs	0.97	0.92	0.94	0.94	−0.18	−0.18
HCs	0.8	0.72	0.92	0.94	0.04	−0.12

validation process can be completely repeated. Finally, the LOOCV is more suitable for the classification of low number of samples, and can better prevent overfitting [43,44].

3. Results and analysis

3.1. Behavioral measures

3.1.1. Comparisons of mean RTs between the two groups in valid/invalid trials

Average RTs and reliability coefficients in both two groups under different conditions are shown in Fig. 2. Statistical analysis showed that main effect of validity (Fig. 2(A), $F_{1,32}=6.86$, $p<0.02$, $\eta^2=0.002$) and group × type × validity interaction ($F_{1,32}=4.48$, $p<0.05$, $\eta^2=0.005$) were significant. No other main effects or interactions obtained significance. To further assess the 3-way interaction, a simple effect test was performed. The results showed that the validity effect was significant for sad-neutral pairs in MDDs ($F_{1,32}=9.60$, $p<0.005$). No other significant effects were found. These data showed that MDD patients had shorter RTs for valid (193.48 ± 75.45 ms) than invalid (198.40 ± 79.66 ms) sad trials ($p<0.05$). Table 1 shows reliability coefficients for each trial type in MDD patients and HCs. As illustrated in Table 1, reliability coefficients of RTs in the four conditions were all above 0.7.

3.1.2. Attentional bias scores in the two groups for emotional faces

To verify that MDD patients would selectively attend to sad faces, repeated-measures ANOVA was conducted with type and group to assess bias scores. Only a group × type interaction showed significance (Fig. 2(B), $F_{1,\ 32}=4.52$, $p<0.05$, $\eta 2=0.47$). Follow-up tests revealed that bias scores in the MDD group for sad faces were significantly higher than those of the HC group ($t(32)=5.24$, $p<0.03$). However, group differences in bias scores hardly determined which group showed bias; this may be caused by an independent bias from one of the groups, or varying degrees of bias from both groups [36]. To identify the group showing bias, one-sample *t*-test was performed by comparing with zero. In MDD patients, bias scores of sad faces were significantly higher than zero ($t(16)=2.45$, $p<0.05$), which indicated the MDD group showed bias toward negative pictures. In HCs, no bias score was different from zero. However, reliability coefficients of bias scores for all face-pairs showed poor reliability (Table 1).

3.1.3. Cue-evoked ERP component

P100 measures are summarized in Table 2. No significant main and interaction effects were found for P100 amplitude. In P100 latency, there was a significant group difference ($F_{1,32}=84.68$, $p<0.0005$, $\eta^2=0.35$), indicating that P100 latency in MDD patients was significantly shorter than that of HC individuals. In addition, a main effect of hemisphere ($F_{1,32}=41.62$, $p<0.0005$, $\eta^2=0.12$), and hemisphere × group interaction effect ($F_{1,32}=84.33$, $p<0.0005$, $\eta^2=0.25$) were observed. For the above two-way interaction, follow-up simple effects test revealed that MDD patients had significantly decreased P100 latency over right parietal-occipital than HCs for all face-pairs (Fig. 3, $F_{1,32}=269.38$, $p<0.0005$).

3.1.4. Target-evoked ERP component

P300 measures are summarized in Table 2. For P300 amplitude, the main effects of type ($F_{1,32}=4.93$, $p<0.04$, $\eta^2=0.01$) and validity ($F_{1,32}=4.32$, $p<0.05$, $\eta^2=0.002$), as well as validity × group interaction ($F_{1,32}=9.51$, $p<0.005$, $\eta^2=0.005$), margin type × validity interaction ($F_{1,32}=3.48$, $p<0.08$, $\eta^2=0.01$), and margin interaction effect of type × validity × electrode site ($F_{1,32}=3.89$, $p<0.06$, $\eta^2=0.0007$) were significant. No other main effects or interactions showed significance. Then, a separate RM-ANOVA was conducted for each group. The main effect of validity ($F_{1,32}=14.44$, $p<0.003$, $\eta^2=0.04$) and type × validity × electrode site interaction ($F_{1,16}=5.76$, $p<0.04$, $\eta^2=0.001$) were significant only in MDD group. For the type × validity × electrode site interaction, follow-up simple effects analysis revealed that MDD patients had greater P300 amplitude to valid than invalid sad trials at the Cz ($F_{1,16}=10.57$, $p<0.01$) and Pz (Fig. 4, $F_{1,16}=8.03$, $p<0.02$) channels.

For P300 latency, a main effect of electrode site ($F_{1,32}=6.73$, $p<0.02$, $\eta^2=0.05$) and a significant interactive effect with

Table 2
Peak latencies and mean amplitudes for the P100 and P300 components.

Group	Type	P1 Am*(SD)		P1 La**(SD)		Val***	P3 Am*(SD)		P3 La**(SD)	
		OL	OR	OL	OR		Cz	Pz	Cz	Pz
MDDs	happy-neutral	1.33	0.73	138.82	109.76	valid	3.41(4.37)	4.55(4.03)	356.92(35.06)	350.46(35.64)
		−0.46	−0.47	−2.63	−1.9	invalid	4.53(5.06)	4.42(3.81)	355.69(29.36)	353.23(37.96)
	sad-neutral	1.38	1.1	140.71	109.18	valid	5.57(3.33)	6.36(3.95)	345.23(29.68)	343.08(30.86)
		−0.47	−0.47	−2.89	−1.6	invalid	3.98(3.51)	4.14(4.27)	354.77(26.04)	350.77(34.19)
HCs	happy-neutral	2	2.17	142.59	148.12	valid	2.87(2.98)	2.23(3.12)	364.29(57.03)	327.71(31.43)
		−0.46	−0.47	−2.63	−1.9	invalid	3.39(3.73)	2.79(2.45)	359.43(47.94)	330.57(39.74)
	sad-neutral	2.28	2.34	143.18	148.24	valid	4.26(3.83)	3.64(3.33)	358.29(55.46)	326(38.67)
		−0.47	−0.47	−2.89	−1.6	invalid	3.16(4.18)	3.53(2.86)	358(51.24)	323.71(40.16)

SD, standard deviation; *Am, Amplitude; **La, Latency; ***Val, Validity; OL = ERP average of channels PO7 and O1; OR = ERP average of channels PO8 and O2;

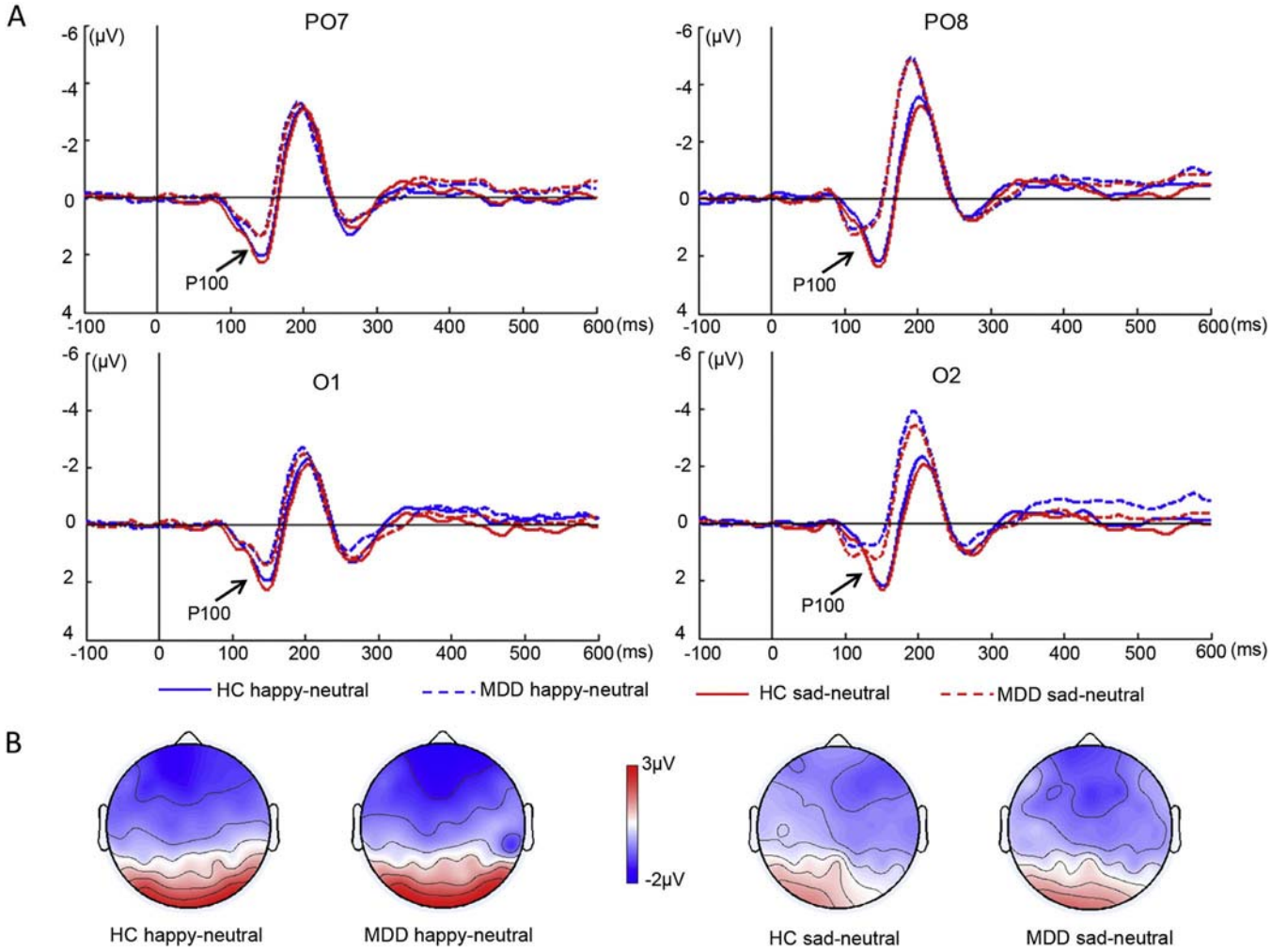

Fig. 3. Grand average ERPs elicited by sad-neutral and happy-neutral pairs at PO7, PO8, O1 and O2 sites in MDD patients and HCs (A), along with scalp topographies of the P100 component (140–160 ms) in MDD patients and HCs, displaying a maximum at occipital electrodes (B).

group × electrode site ($F_{1,32} = 4.25$, $p = 0.05$, $\eta^2 = 0.03$), were detected. Simple effects analysis revealed that P300 latency at channel Pz was shorter than that at channel Cz in MDD individuals, which was not observed in HC individuals.

3.2. Classification performance

Tables 3 and 4 report the numbers of features selected by each algorithm and the corresponding average classification accuracy (KNN, C4.5, SMO, LR) for the P100 component in two different conditions, and for the P300 component in four different conditions, respectively. The accuracy performances of the four classifiers were greatly improved after features selection by the proposed heuristics. As indicated in Table 3, CFS was the most effective feature selection algorithm for happy-neutral pairs, obtaining 6 features and a mean accuracy of 85.29%. Classification performance was improved by about 14% compared with that obtained using all features. For sad-neutral pairs, the best method for screening features was ReliefF, which obtained 6 features and a mean accuracy of 69.78%. Classification performance was improved by about 20% compared with that recorded with all features used. However, CFS

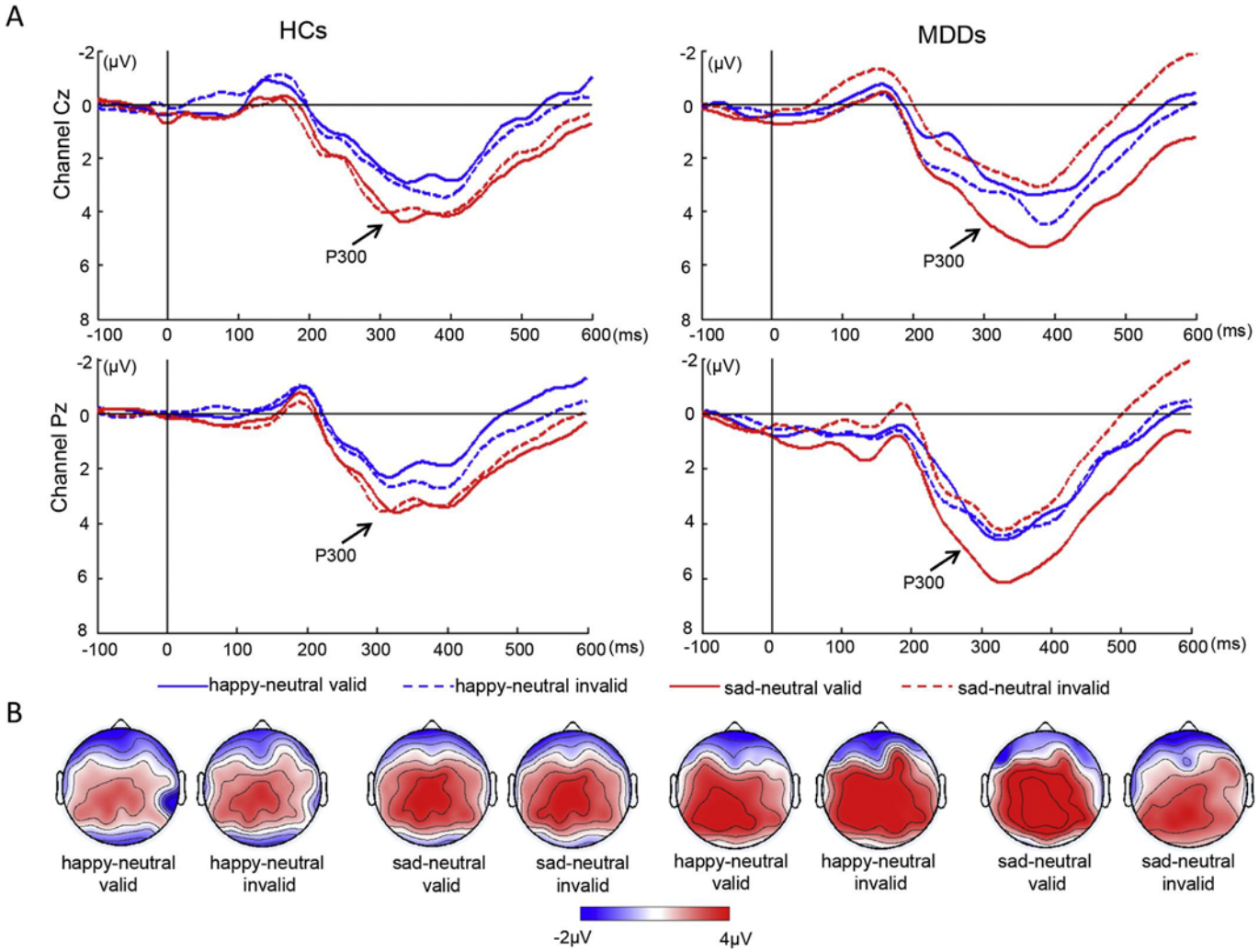

Fig. 4. Grand-average ERPs were elicited by validly and invalidly probe stimuli at the Cz and Pz sites in MDD patients and healthy controls (A), and scalp topographies of the P300 component (350–370 ms) displayed a maximum at centro-parietal electrodes (B).

Table 3
Comparison of different feature selection algorithms based on the P100 component.

Feature selection methods	Happy-neutral pairs		Sad-neutral pairs	
	Mean accuracy	Mean features number	Mean accuracy	Mean features number
None	71.35%	98	49.41%	98
CFS	**85.29%**	6	66.18%	3
ReliefF	79.41%	9	**69.78%**	6
GainRatio	83.82%	8	67.65%	8

Mean Accuracy, mean accuracy of the four classifiers (KNN, C4.5, SMO and LR); mean features number, mean number of needed features for the four classifiers to obtain highest accuracies using each feature selection method (CFS, ReliefF and RG).

Table 4
Comparison of different feature selection algorithms based on the P300 component.

Feature selection methods	Happy-neutral valid		Happy-neutral invalid		Sad-neutral valid		Sad-neutral invalid	
	Mean accuracy	Mean features number	Mean accuracy	Mean features number	Mean accuracy	Mean features number	Mean accuracy	Mean features number
None	71.33%	98	54.12%	98	69.12%	98	70.00%	98
CFS	**82.35%**	9	**68.24%**	3	**88.94%**	10	**87.03%**	19
ReliefF	79.41%	10	62.94%	3	80.25%	14	83.92%	15
GainRatio	80.29%	12	64.71%	8	86.00%	8	80.46%	3

Mean accuracy, mean accuracy of the four classifiers (KNN, C4.5, SMO and LR); mean features number, mean number of needed features for the four classifiers to obtain highest accuracies using each feature selection method (CFS, ReliefF and RG).

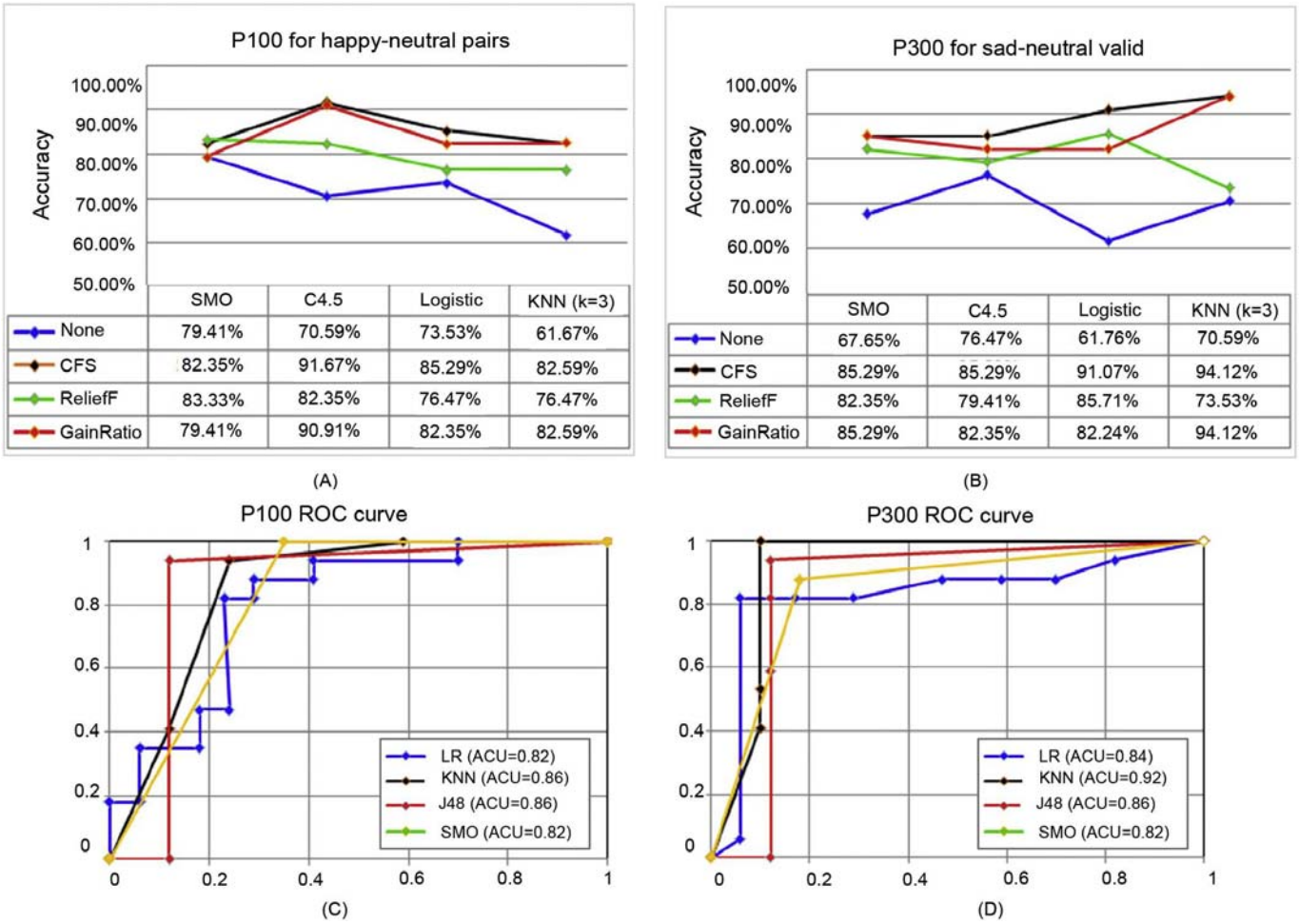

Fig. 5. Classification accuracies of KNN, C4.5, SMO, and LR for the P100 component elicited by happy-neutral pairs (A), and P300 component elicited by sad-neutral valid trials (B). Using CFS as a feature selection method, ROC curves of MDD patients in the four classifiers based on the P100 evoked by happy-neutral pairs and P300 evoked by sad-neutral valid trials, shown in (C) and (D), respectively.

used fewer features (only 3), and reached an accuracy of 66.18%. Compared with ReliefF and GR, CFS took less time in feature selection.

As shown in Table 4, the best feature selection method was CFS for the P300 evoked by the four conditions. The mean classification accuracy of the P300 evoked by valid sad trials was higher than those of other conditions, deriving 10 features with a mean accuracy of 88.94%.

Fig. 5 shows the classification accuracy for the three feature extraction algorithms (CFS, ReliefF and RG) in combination, with different types of classifiers (KNN, C4.5, SMO, LR) using LOOCV method. As shown in Fig. 5(A), in the P100 component elicited by happy-neutral pairs, using CFS as a feature selection method, C4.5 had the highest classification accuracy, with classification accuracy reaching 91.67%. The area under the receiver operating characteristic (ROC) curve (AUC) of C4.5 was 0.86 (Fig. 5(C)). For the P300 component elicited by sad-neutral valid pairs, combination of KNN ($k=3$) and CFS or GR yielded an accuracy of 94.12% (Fig. 5(B)) and an AUC of 0.92 for CFS (Fig. 5(D)). These findings indicated that combined C4.5 and CFS achieved the highest accuracy for the cue-ERP component, and combined KNN and CFS yielded the highest accuracy for the target-ERP component. Above all, we consider that CFS is the most effective feature selection method in discriminating between MDD patients and HCs.

4. Discussion

The present study assessed attentional bias to emotional faces in MDD patients by examining ERP components in a dot-probe task. With the purpose of finding a more efficient method for detecting MDD, a combination of three feature selection methods and four classification algorithms was evaluated. The following findings emerged. Firstly, we assessed the behavioral aspects of attentional mechanisms, as reflected in RT. The results confirmed the notion that MDD individuals have significantly faster RT to valid sad trials compared to invalid sad trials. Specifically, strongest split-half reliability was obtained for RTs. Using a dot-probe task with emotional faces as stimuli, Gotlib et al. [45] found attentional bias to negative faces in depression. Joormann and Gotlib [12] and Joormann et al. [36] reported similar findings in previously depressed participants and girls at high risk of depression, respectively. We did replicate Gotlib et al. findings in depression, demonstrating negative attentional bias by the RT index in the dot-probe task (i.e., difference in RT for valid sad and invalid sad trials) [45]. Additionally, a bias score significantly larger than zero was observed for sad-neutral pairs in MDD patients, but not HCs. However, split-half reliability of RT bias scores neared zero. This is consistent with previous findings that the RT bias scores of attentional bias exhibit poor reliability [13]. Behavioral results indicated that MDD patients were bi-

ased to sad faces in contrast to HCs, which corroborates a previous study [45].

Secondly, early P100 component analysis showed that MDD patients had relatively shorter latency for all face pairs over right parietal-occipital regions compared with HC individuals. Electrophysiological recording and neuroimaging studies reported that facial emotion processing mainly occurs in the right hemisphere of the brain. Sato et al. [46] found that the occipital and temporal cortices, especially in the right hemisphere, have higher activation during the process of watching dynamic pictures by fMRI. A large number of studies also pointed out that the right hemisphere has advantages in the field of facial processing, e.g. in face recognition [19] and face-selective activation [47]. According to a study by Zhao et al. [48], early perceptual processing abnormality in the depression group only occurs at the right posterior sites. Additionally, the decreased P100 latency for all facial expression indexes in MDD patients may imply that cognitive biases to emotional facial stimuli might not have arisen yet during the earliest stages of preconscious processing.

The P100 induced by happy-neutral pairs showed higher classification performance compared with sad-neutral pairs. The classification performance of C4.5 and CFS combination reached 91.67% using the LAR and Ppmean features from channels Oz and O2. The AUC of KNN was 0.86, which was equal to that of C4.5, but C4.5 used fewer features. Right hemisphere deficiency in depression during perceptual processing has been reported previously [47,48]. The current experimental results showed that classification accuracy is highest for distinguishing between MDD patients and HCs when features are selected from the right parietal–occipital region.

Thirdly, the P300 amplitude of MDD patients was significantly greater when the probe replaced sad faces than when it was shown after neutral faces, with the maximal value at parietal recording sites. After presentation of probe stimuli, ERP analysis helped examine how attention bias affects cognitive processing [10,20,21]. Previous studies considered that the P300 component reflect an activity involved in attentional resource allocation [24,25]. The ERP task induced by the P300 is the most commonly used ‘oddball task’ [49,50]. To the best of our knowledge, there are no published studies on P300 modulation to a visual probe in MDD patients using a dot-probe task. Enhanced P300 amplitude in response to valid sad trials in MDD patients, reflected the difficulty in attentional disengagement across negative stimuli, which supports previous findings suggesting that depressed individuals have difficulty in disengaging attention from negative events [6,7]. Meanwhile, negativity bias may excessively enhance memory for negative materials in depression.

The accuracy of CFS and KNN combination achieved 94.12% and an AUC of 0.92 using 3 features, including PAR, Mobility and Latency, which were extracted at channels Pz and C1. In addition, the classification performance of the GR and KNN combination also reached 94.12%. For KNN, both GR and CFS provided a small number of feature sets and obtained good classification accuracy. Compared to previous studies [28], we used 3 features from two electrodes to achieve good classification performance. The current findings indicated that the P300 induced by valid sad trials in MDD has more sensitive information, and may be helpful for assisted diagnosis of depression.

Finally, classification performance after feature selection was improved obviously compared with those using all the features, especially for CFS. Classification accuracy of the combination of CFS with all classifiers has increased by more than 11%. The advantage of CFS over other methods is that it can automatically determine the number of distinguishing features [51]. Hall and Holmes [38] showed that CFS performs consistently well on different data sets, suggesting that CFS is a good overall performer.

The present findings supported the notion that depressive participants display attentional bias to negative stimuli in emotion processing in MDD. However, because of the relatively small sample size of bidirectional bipolar disorder, high-risk of depression and schizophrenia, we could not compare the different characteristics of attentional bias in various diseases when dealing with emotional stimuli. This is a limitation of the present work. We are still collecting data on the above mood disorders. In addition, due to the high trial-to-trial variability and the unfavorable ratio between signal (ERP) and noise, implementation classification using ERP on a single trial basis, and extracting effective features for real-time diagnosis of depression are also challenging problems. In future studies, effective feature extraction algorithms should be developed to identify and diagnose depression.

5. Conclusion

Event-Related Potentials (ERPs) measured on the human scalp reflect brain electrical activity related to stimulus processing. To assess attentional bias in major depression, we focused on reaction time, and P100 and P300 components. Reaction time with the strongest reliability indicated that MDD patients selectively attend to sad faces. At the neurological level, MDD individuals exhibit difficulty in disengaging attention from sad faces, which was reflected in the P300. Furthermore, data mining indicated that CFS is the most effective feature selection method to discriminate between MDD patients and HCs. The CFS and C4.5 combination achieved the highest accuracy, with a classification accuracy above 92% for the P100 induced by happy-neutral pairs. The CFS and KNN ($K=3$) combination achieved the highest accuracy, with a classification accuracy above 94% for the P300 induced by valid sad trials. The P300 showed better classification results compared with the P100. These findings not only provide novel insights into emotional negativity biases in depression, but also reveal an auxiliary method for identifying depression in a portable system.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

This work was supported by the National Basic Research Program of China (973 Program) (No. 2014CB744600); the National Natural Science Foundation of China (Nos. 61632014, 61210010); the International Cooperation Project of Ministry of Science and Technology (No. 2013DFA11140); and the Program of Beijing Municipal Science & Technology Commission (No. Z171100000117005).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2018.07.003.

Reference

[1] M.B. First, Diagnostic and statistical manual of mental disorders, 5th edition, and clinical utility, J. Nerv. Mental Dis. 201 (2013) 727–729.
[2] World Health Organization (WHO. 2018); http://www.who.int/en/news-room/fact-sheets/detail/depression.
[3] R. De Raedt, E.H.W. Koster, Understanding vulnerability for depression from a cognitive neuroscience perspective: a reappraisal of attentional factors and a new conceptual framework, Cogn. Affect. Behav. Neurosci. 10 (2010) 50–70, doi:10.3758/CABN.10.1.50.
[4] N.A.S. Farb, J.A. Irving, A.K. Anderson, Z.V. Segal, A two-factor model of relapse/recurrence vulnerability in unipolar depression, J. Abnorm. Psychol. 124 (2015) 38–53.

[5] P.C. Clasen, T.T. Wells, A.J. Ellis, C.G. Beevers, Attentional biases and the persistence of sad mood in major depressive disorder, J. Abnorm. Psychol. 122 (2013) 74–85.

[6] G.J. Siegle, S.R. Steinhauer, M.E. Thase, V.A. Stenger, C.S. Carter, Can't shake that feeling: event-related fMRI assessment of sustained amygdala activity in response to emotional information in depressed individuals, Biol. Psychiat. 51 (2002) 693.

[7] S.G. Disner, C.G. Beevers, E.A.P. Haigh, A.T. Beck, Neural mechanisms of the cognitive model of depression, Nat. Rev. Neurosci. 12 (2011) 467.

[8] H. Brown, T. Eley, S. Broeren, C. MacLeod, M. Rinck, J. Hadwin, K. Lester, Psychometric properties of reaction time based experimental paradigms measuring anxiety-related information-processing biases in children, J. Anxiety Disord. 28 (2014) 97–107. https://doi.org/10.1016/j.janxdis.2013.11.004.

[9] E.S. Kappenman, J.L. Farrens, S.J. Luck, G.H. Proudfit, Behavioral and ERP measures of attentional bias to threat in the dot-probe task: poor reliability and lack of correlation with anxiety, Front. Psychol. 5 (2014), doi:10.3389/fpsyg.2014.01368.

[10] Q. Zhao, H. Li, B. Hu, Y. Li, C. Gillebert, D. Mantini, Q. Liu, Neural correlates of drug-related attentional bias in heroin dependence, Front. Hum. Neurosci. 11 (2017) 646.

[11] E. Fox, R. Russo, K. Dutton, Attentional bias for threat: evidence for delayed disengagement from emotional faces, Cognit. Emot. 16 (2002) 355–379.

[12] J. Joormann, I.H. Gotlib, Selective attention to emotional faces following recovery from depression, J. Abnorm. Psychol. 116 (2007) 80–85, doi:10.1037/0021-843X.116.1.80.

[13] B.E. Gibb, S.D. Pollak, G. Hajcak, M. Owens, Attentional biases in children of depressed mothers: an event-related potential (ERP) study, J. Abnorm. Psychol. 125 (2016) 1166, doi:10.1037/abn0000216.

[14] K. Horley, L.M. Williams, C. Gonsalvez, E. Gordon, Face to face: visual scanpath evidence for abnormal processing of facial expressions in social phobia, Psychiatry Res. 127 (2004) 43–53. https://doi.org/10.1016/j.psychres.2004.02.016.

[15] A.E. Whitton, S. Deccy, M.L. Ironside, P. Kumar, M. Beltzer, D.A. Pizzagalli, Electroencephalography source functional connectivity reveals abnormal high-frequency communication among large-scale functional networks in depression, Biol. Psychiatry Cogn. Neurosci. Neuroimaging 3 (2018) 50–58.

[16] P. Maurage, P. Philippot, P. Verbanck, X. Noel, C. Kornreich, C. Hanak, S. Campanella, Is the P300 deficit in alcoholism associated with early visual impairments (P100, N170)? An oddball paradigm, Clin. Neurophysiol. 118 (2007) 633–644.

[17] N. Jaworska, P. Blier, W. Fusee, V. Knott, The temporal electrocortical profile of emotive facial processing in depressed males and females and healthy controls, J. Affect. Disord. 136 (2012) 1072–1081.

[18] M. Eimer, A. Holmes, F.P. McGlone, The role of spatial attention in the processing of facial expression: an ERP study of rapid brain responses to six basic emotions, Cognit. Affect. Behav. Neurosci. 3 (2003) 97–110.

[19] M. Batty, M.J. Taylor, Early processing of the six basic facial emotional expressions, Cognit. Brain Res. 17 (2003) 613–620.

[20] Z. Mingtian, Z. Xiongzhao, Y. Jinyao, Y. Shuqiao, R.A. Atchley, Do the early attentional components of ERPs reflect attentional bias in depression? It depends on the stimulus presentation time, Clin. Neurophysiol. 122 (2011) 1371–1381, doi:10.1016/j.clinph.2010.09.016.

[21] W. Yang, X. Zhu, X. Wang, D. Wu, S. Yao, Time course of affective processing bias in major depression: an ERP study, Neurosci. Lett. 487 (2011) 372–377, doi:10.1016/j.neulet.2010.10.059.

[22] A. Ibanez, M. Melloni, D. Huepe, E. Helgiu, A. Rivera-Rei, A. Canales-Johnson, P. Baker, A. Moya, What event-related potentials (ERPs) bring to social neuroscience? Social Neurosci. 7 (2012) 632–649.

[23] K. Kompatsiari, G. Candrian, A. Mueller, Test-retest reliability of ERP components: a short-term replication of a visual Go/NoGo task in ADHD subjects, Neurosci. Lett. 617 (2016) 166–172.

[24] H.M. Fichtenholtz, J.B. Hopfinger, R. Graham, J.M. Detwiler, K.S. Labar, Happy and fearful emotion in cues and targets modulate event-related potential indices of gaze-directed attentional orienting, Soc. Cogn. Affect. Neur. 2 (2007) 323–333, doi:10.1093/scan/nsm026.

[25] V. Leutgeb, M. Sarlo, F. Schöngassner, A. Schienle, Out of sight, but still in mind: electrocortical correlates of attentional capture in spider phobia as revealed by a 'dot probe'paradigm, Brain Cogn. 93 (2015) 26–34.

[26] J.R. Sveinsson, J.A. Benediktsson, S.B. Stefansson, K. Davidsson, Parallel principal component neural networks for classification of event-related potential waveforms, Med. Eng. Phys. 19 (1997) 15–20. https://doi.org/10.1016/S1350-4533(96)00035-5.

[27] R. Palaniappan, R. Paramesran, Using genetic algorithm to identify the discriminatory subset of multi-channel spectral bands for visual response, Appl. Soft Comput. 2 (2002) 48–60. https://doi.org/10.1016/S1568-4946(02)00028-5.

[28] I. Kalatzis, N. Piliouras, E. Ventouras, C.C. Papageorgiou, A.D. Rabavilas, D. Cavouras, Design and implementation of an SVM-based computer classification system for discriminating depressive patients from healthy controls using the P600 component of ERP signals, Comput. Meth. Prog. Bio. 75 (2004) 11–22. https://doi.org/10.1016/j.cmpb.2003.09.003.

[29] Y. Lecrubier, D.V. Sheehan, E. Weiller, P. Amorim, I. Bonora, K.H. Sheehan, J. Janavs, G.C. Dunbar, The mini international neuropsychiatric interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI, Eur Psychiat 12 (1997) 224–231, doi:10.1016/S0924-9338(97)83296-8.

[30] D. Matsumoto, P. Ekman, American-Japanese cultural differences in intensity ratings of facial expressions of emotion, Motiv. Emot. 13 (1989) 143–157.

[31] B. Lu, M. Hui, H. Yu-Xia, The development of native chinese affective picture system–a pretest in 46 college students, Chin. Mental Health J. 19 (11) (2005) 719–722.

[32] A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, J. Neurosci. Meth. 134 (2004) 9–21, doi:10.1016/j.jneumeth.2003.10.009.

[33] J. Lopezcalderon, S.J. Luck, ERPLAB: an open-source toolbox for the analysis of event-related potentials, Front. Hum. Neurosci. 8 (2014) 213.

[34] T.P. Jung, S. Makeig, C. Humphries, T.W. Lee, M.J. McKeown, V. Iragui, T.J. Sejnowski, Removing electroencephalographic artifacts by blind source separation, Psychophysiology 37 (2000) 163–178.

[35] S.M. Sass, W. Heller, J.E. Fisher, R.L. Silton, J.L. Stewart, L.D. Crocker, J.C. Edgar, K.J. Mimnaugh, G.A. Miller, Electrophysiological evidence of the time course of attentional bias in non-patients reporting symptoms of depression with and without co-occurring anxiety, Front. Psychology 5 (2014) 301.

[36] J. Joormann, L. Talbot, I.H. Gotlib, Biased processing of emotional information in girls at risk for depression, J. Abnorm. Psychol. 116 (2007) 135.

[37] B. Hjorth, EEG analysis based on time domain properties„ Electroencephalogr. Clin. Neurophysiol. 29 (1970) 306–310.

[38] M.A. Hall, G. Holmes, Benchmarking attribute selection techniques for discrete class data mining, IEEE Trans. Knowl. Data Eng. 15 (2002) 1437–1447.

[39] M. Robnik-Šikonja, I. Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF, Mach. Learning 53 (2003) 23–69.

[40] J. Han, J. Pei, Kamber M., Data Mining Concepts and Techniques, Elsevier, 2011.

[41] X. Li, B. Hu, S. Sun, H. Cai, EEG-based mild depressive detection using feature selection methods and classifiers, Comput. Methods Prog. Biomed. 136 (2016) 151–161. https://doi.org/10.1016/j.cmpb.2016.08.010.

[42] G. Castellano, S.D. Villalba, A. Camurri, Recognising human emotions from body movement and gesture dynamics, affective computing and intelligent interaction, in: Second International Conference, ACII 2007, Lisbon, Portugal, September 12-14, 2007, Proceedings, 2007, pp. 71–82.

[43] B.A. Johnston, B. Mwangi, K. Matthews, D. Coghill, J.D. Steele, Predictive classification of individual magnetic resonance imaging scans from children and adolescents, Eur. Child Adolesc. Psychiatry 22 (2013) 733–744.

[44] B. Mwangi, K.M. Hasan, J.C. Soares, Prediction of individual subject's age across the human lifespan using diffusion tensor imaging: a machine learning approach, Neuroimage 75 (2013) 58–67.

[45] I.H. Gotlib, E. Krasnoperova, D.N. Yue, J. Joormann, Attentional biases for negative interpersonal stimuli in clinical depression, J. Abnorm. Psychol. 113 (2004) 121–135, doi:10.1037/0021-843X.113.1.121.

[46] W. Sato, T. Kochiyama, S. Yoshikawa, E. Naito, M. Matsumura, Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study, Brain Res. Cogn. Brain Res. 20 (2004) 81–91.

[47] A.D. Engell, G. McCarthy, Probabilistic atlases for face and biological motion perception: an analysis of their reliability and overlap, Neuroimage 74 (2013) 140–151. https://doi.org/10.1016/j.neuroimage.2013.02.025.

[48] Q. Zhao, Y. Tang, S. Chen, Y. Lyu, A. Curtin, J. Wang, J. Sun, S. Tong, Early perceptual anomaly of negative facial expression in depression: an event-related potential study, Neurophysiol. Clin./Clin. Neurophysiol. 45 (2015) 435–443, doi:10.1016/j.neucli.2015.09.011.

[49] T.A. Ortiz, J. Perez-Serrano, C.Z. Zaglul, J.G. Coullaut, R.G. Coullaut, J.R. Criado, P300 clinical utility in major depression, Actas Esp. de Psiquiatria 30 (2002) 1–6.

[50] J.G. Coullaut-Valera, I.D. del Rio Arbaiza, R.G. Coullaut-Valera, T. Ortiz, Alterations of P300 wave in occipital lobe in depressive patients, Actas Espanolas de Psiquiatr. 35 (2007) 243–248.

[51] H. Liu, J. Li, L. Wong, A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns, Genome Inf. 13 (2002) 51–60.