

## Research Article

# Method of Depression Classification Based on Behavioral and Physiological Signals of Eye Movement

Mi Li <sup>1,2</sup>, Lei Cao <sup>1,2</sup>, Qian Zhai,<sup>3,4</sup> Peng Li <sup>1,2</sup>, Sa Liu <sup>1,2</sup>, Richeng Li <sup>1,2</sup>, Lei Feng,<sup>3,4</sup>  
Gang Wang,<sup>3,4</sup> Bin Hu,<sup>1,5</sup> and Shengfu Lu <sup>1,2</sup>

<sup>1</sup>Department of Automation, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

<sup>2</sup>The Beijing International Collaboration Base on Brain Informatics and Wisdom Services, Beijing 100124, China

<sup>3</sup>The National Clinical Research Center for Mental Disorders & Beijing Key Laboratory of Mental Disorders, Beijing Anding Hospital, Capital Medical University, Beijing, China

<sup>4</sup>The Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing, China

<sup>5</sup>Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, Lanzhou, China

Correspondence should be addressed to Shengfu Lu; [lusf@bjut.edu.cn](mailto:lusf@bjut.edu.cn)

Received 16 August 2019; Accepted 20 December 2019; Published 14 January 2020

Guest Editor: Michela Gelfusa

Copyright © 2020 Mi Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a method of depression recognition based on direct measurement of affective disorder. Firstly, visual emotional stimuli are used to obtain eye movement behavior signals and physiological signals directly related to mood. Then, in order to eliminate noise and redundant information and obtain better classification features, statistical methods (FDR corrected *t*-test) and principal component analysis (PCA) are used to select features of eye movement behavior and physiological signals. Finally, based on feature extraction, we use kernel extreme learning machine (KELM) to recognize depression based on PCA features. The results show that, on the one hand, the classification performance based on the fusion features of eye movement behavior and physiological signals is better than using a single behavior feature and a single physiological feature; on the other hand, compared with previous methods, the proposed method for depression recognition achieves better classification results. This study is of great value for the establishment of an automatic depression diagnosis system for clinical use.

## 1. Introduction

Depression is a psychiatric disorder characterized by significant and persistent loss of pleasure, anhedonia, and decreased interest. To date, no specific biological markers have been found for the diagnosis of depression. Therefore, in clinic, the diagnosis of depression is mainly carried out by psychiatrists through structured interviews based on diagnostic manuals (e.g., DSM-IV). With the development of artificial intelligence technology, pattern recognition technology based on machine learning has been widely studied in the recognition or diagnosis of depression.

Recently, a lot of research has been done on the depression classification based on resting-state fMRI brain image signals. For example, Cao et al. classified the severe

depression patients based on functional connections of resting-state fMRI by feature selection and SVM and obtained 84.21% classification accuracy [1]. Qin et al. used SVM to identify patients with severe depression based on the diffusion tensor imaging (DTI) of resting-state fMRI, and the highest classification accuracy was 83.05% [2]. Sato et al. classified the patients with severe depression based on fMRI signals and achieved an accuracy of 78.26%, sensitivity of 72.00%, and specificity of 85.71% [3]. Bhaumik et al. used SVM to classify remitted major depressive patients based on the functional connectivity of resting-state fMRI and achieved an accuracy of 76.1%, sensitivity of 81.5%, and specificity of 68.9% [4]. Schnyer et al. used DTI and SVM to identify patients with severe depression, and the highest classification accuracy, specificity, and sensitivity were

74.0%, 80.0%, and 68.0% [5]. Ramasubbu et al. used SVM based on the voxel space of resting-state fMRI to classify mild-moderate, severe, and very severe patients, and the classification accuracy was 58%, 52%, and 66%, respectively [6].

In addition to using fMRI signals to identify depression, in recent years, EEG-based depression recognition has also been widely studied. For example, Liao et al. used SVM to classify severe depression patients based on resting-state EEG signals, and 80% of the classification accuracy is obtained [7]. Bairy et al. used the decision tree algorithm to classify based on EEG signal, and the classification accuracy, sensitivity, and specificity were 94.30%, 91.46%, and 97.45% [8]. Mumtaz et al., based on resting-state EEG using SVM for classification, achieved an accuracy of 98.4%, sensitivity of 96.66%, and specificity of 100% [9]. Acharya et al. used deep convolution neural network to recognize depression based on EEG signals and achieved the highest accuracy of 96% [10].

Compared with EEG and fMRI signals, the behavior data such as expression and voice are easier to obtain. Therefore, a lot of research has been done on the method of depression classification based on behavior data. For example, Valstar et al. used linear SVM with random gradient descent to classify depression based on video expression, audio, and multimodality, and the F1 scores were 0.583, 0.889, and 0.467, respectively [11]. Ma et al. proposed a depth classification model (DepAudioNet) combining convolutional neural network (CNN) with long-term and short-term memory (LSTM) based on audio signals, and the optimal F1 score is 0.52 (0.70), classification accuracy is 0.35 (1.00), and recall is 1.00 (0.54) (nondepressive results in parentheses) [12].

In addition to using a single physiological or behavioral signal for depression classification, many studies have recognized depression based on multimodal data. For example, Zhao et al. used the multimodal data of resting-state fMRI and DTI for classification with an accuracy of 80.95% [13]. Le et al. fused audio, video, language, and sleep multimodal data at feature level and obtained F1 scores of 57.1% and 87.7% in depression patients and normal control group, respectively, by using decision tree algorithm [14]. Al Hanai et al. used LSTM to classify depression based on multimodal data of audio/text, and the optimal F1 score and recall rate were 0.77 and 0.83, respectively [15]. Haque et al. used causal convolution neural network (C-CNN) to identify depression in patients with severe depression based on multimodality of expression, voice, and linguistic (text), and 83.3% sensitivity and 82.6% specificity were obtained [16].

Compared with EEG/fMRI, behavioral signals are easier to collect and low in cost from the source of classified signals. Therefore, the method of depression recognition based on behavioral signals and deep learning methods has attracted the attention of researchers. However, whether it is physiological signal (fMRI, EEG) or behavioral signal (expression, voice, etc.), the common characteristics of these signals are acquired in the natural state of the individual; for example, resting-state EEG and fMRI signals are all acquired when the individual is in the state of closing eyes, relaxing, and not

thinking about anything. Behavioral signals such as expressions and voices are natural state signals obtained during interviews (usually using virtual agents). That is to say, at present, the classification signals collected in the natural state are not directly related to affective disorders but indirect measurement of affective disorders. However, the core symptoms of depression patients are low and bad mood is caused by emotional injury, so the signals obtained by direct measurement of emotions are of great value for the study of depression recognition.

Eye movement technology opens up a new way for the study of automatic detection of depression. The eye is the window of the mind and the important organ for humans to observe the world directly. Eye movement signals, such as gaze objects, gaze time, gaze shift, and pupil size, acquired by eye tracking technology, are all direct reflections of the brain's information processing demand and can quantitatively characterize emotional perception. Therefore, eye movement signals are a direct measurement of emotional state. Alghowinem et al., firstly based on the eye movement features (horizontal, vertical, and eyelid movement) obtained by emotional language stimulation to classify 30 depressive patients and 30 normal people, achieved 75% of the classification accuracy by using the Gauss mixture model and SVM [17]. Li et al. used random forest algorithm to classify 9 depressive patients and 25 normal people based on the eye movement features (pupil size, gaze position, gaze time, etc.) obtained by emotional expression picture stimulation, and the classification accuracy was 80.1% [18]. It can be seen that the above studies using eye movement behavior signals and physiological signals (pupil diameter) as classification features still do not get a high classification accuracy. Although the pupil diameter signal can reflect the change of emotion, the physical size of the pupil diameter varies greatly and cannot be directly used for emotional measurement. Therefore, this study further processed behavior and pupil diameter signals of eye movement in order to extract classification features that can better reflect the differences between depressed patients and normal people.

Many previous eye movement studies have shown that, on the one hand, depressed patients tend to have emotional attention bias [19], i.e., reduced attention to happy emotions and excessive attention to sad emotions, which represented that positive attention bias scores (the difference between the total fixation time at happy expressions and the total fixation time at calm expressions) of depressed patients were significantly greater than those of the normal people, while the negative attention bias scores (the difference between the total fixation time at sad expressions and the total fixation time at calm expressions) of depressed patients were significantly smaller than those of the normal people [20]. On the other hand, the processing of emotional information in the brain can cause changes in pupil diameter. The change in pupil diameter directly reflects the change of people's emotion. Positive emotional visual stimulation can cause people to have happy emotional experience, accompanied by the enlargement of pupil diameter, whereas negative emotional visual stimulation can cause people to have sad emotional experience, accompanied by the smaller pupil

diameter [21–23]. Studies have shown that there are significant differences in pupil diameter between depressive patients and normal people under different emotional stimuli [24, 25].

In this study, two kinds of classification features are extracted from eye movement signals directly related to emotion when evoked by different emotional pictures: one is the attention bias score [20] which reflects eye movement behavior signals, and the other is the emotional bandwidth [26] which reflects physiological signals based on pupil diameter. Noise and redundant information often exist in the acquired physiological and behavioral signal features, so feature selection and feature extraction are needed.

In addition, the performance of classification model directly affects the classification results. The data collected in this study are tabular data, and the data scale is small, which is not suitable for deep learning methods. For relatively small-scale data, SVM implements the suboptimal solution learning method, while the extreme learning machine (ELM) [27, 28] shows better classification performance. ELM is a feedforward neural network model based on a single hidden layer, which has been widely used for its advantages of simple calculation process, fast speed, and good generalization performance. At present, basic ELM and kernel ELM have developed. Compared with basic ELM, kernel ELM has the advantages of fewer adjustable parameters and no need to set hidden layer nodes manually. Therefore, to improve the recognition accuracy of depression, this study used kernel ELM as a classifier. At the same time, to compare the impact of different classification models on the classification results, we also used SVM, KNN, and random forest to classify depression.

## 2. Materials and Methods

**2.1. Signal Acquisition.** There were 96 participants in the data collection, including 48 depressed patients and 48 normal people. From the international standard expression library, the NimStim set [29], three facial expressions (neutral, happy, and sad) images of 36 people (18 males and 18 females), was selected. On the one hand, to reduce the eye movement caused by nonexpressive factors, the ears and hair should be masked as much as possible. On the other hand, the size, resolution, and gray level of all facial expression pictures were consistent through picture manager software.

Data acquisition tasks were divided into two categories: One was eye movement behavioral signals, that is, acquisition of the position and time of each gaze point when a participant views positive or negative bias tasks. Positive bias task is composed of 36 pictures of happy and neutral expressions, while negative bias task is composed of 36 pictures of sad and neutral expressions. Figure 1 shows an example of eye movements when a depressed patient views positive bias task and negative bias task. Each circle represents a gaze point, and its size indicates the length of fixation time.

The other is eye movement physiological signal (pupil diameter) acquisition, that is, the size of pupil diameter produced by each fixation point during a participant viewing positive, negative, or neutral tasks. Positive tasks consist of

two pictures of happy expressions (36 in total), negative tasks consist of two pictures of sad expressions (36 in total), and neutral tasks consist of two pictures of neutral expressions (36 in total), as shown in Figure 2.

The eye-tracking device Tobii T120 was used to collect and record the position, time, and pupil size of each fixation point at a frequency of 120 Hz. As shown in Figure 3, the data acquisition process was as follows: Firstly, a 1000 ms white “+” appeared in the center of the black screen to focus on it [10]. Then, the screen presented a pair of pictures with a time of 3500 ms, followed by a “\*” of 2000 ms, prompting them to rest for the next trial.

**2.2. Classification Feature Calculation.** Firstly, the eye movement behavioral and physiological features are calculated according to the collected behavioral data and physiological data of eye movement.

**2.2.1. Behavioral Features—Emotional Attentional Bias.** Emotional attention bias includes positive bias and negative bias [20]. Among them, positive attention bias refers to the total fixation time of happy expression subtracted from the total fixation time of neutral expression in the “positive bias” task of Figure 1; negative attention bias refers to the total fixation time of sad expression subtracted from the total gaze time of neutral expression in the “negative bias” task of Figure 1.

**2.2.2. Physiological Features—Affective Bandwidth and Change Rate of Pupil Diameter.** Affective bandwidth includes positive affective bandwidth and negative affective bandwidth. Although pupil size is affected by external emotional stimuli, the pupil size of each person is different and is also affected by light, mental load, and so on. Therefore, the use of individual pupil size cannot simply express the effect of emotion. Here, affective bandwidth [26] is used to characterize the effect of emotion on pupil size, including positive and negative affective bandwidth. Positive affective bandwidth indicates the ability of individuals to process and experience happy emotions. The larger the positive affective bandwidth, the stronger the ability of individuals to process and experience happy information. Negative affective bandwidth indicates the ability of individuals to process and experience negative emotions. The larger the negative affective bandwidth, the stronger the ability of individuals to process and experience negative information.

Positive affective bandwidth is calculated that the mean pupil diameter of all fixation points in the “neutral” task (Figure 2(a)) is subtracted from that of all fixation points in the “positive” task (Figure 2(c)). Negative affective bandwidth is calculated that the mean pupil diameter of all fixation points in the “neutral” task (Figure 2(b)) is subtracted from that of all fixation points in the “negative” task (Figure 2(c)).

At the same time, the change rate of positive pupil diameter and negative pupil diameter is calculated with the mean pupil diameter in the “neutral” task as the baseline.

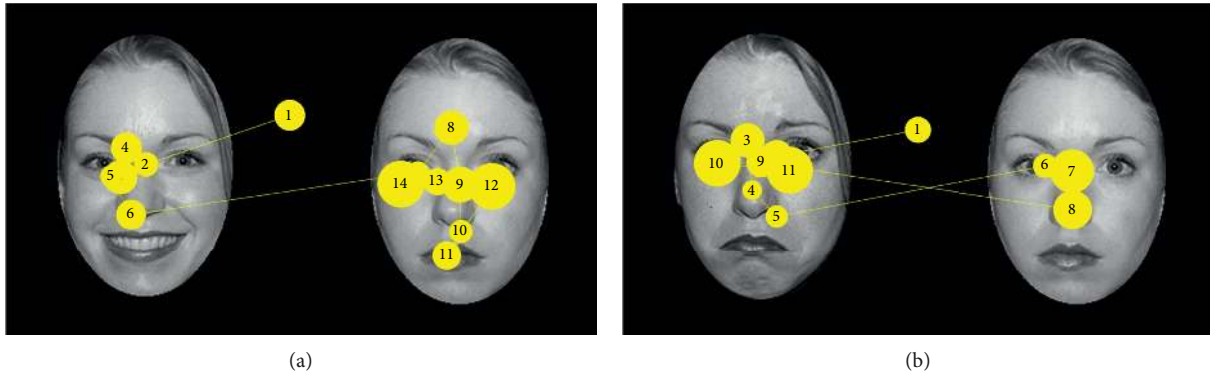


FIGURE 1: An example of eye movements when a depressed patient views (a) positive bias tasks and (b) negative bias tasks. Each circle represents a gaze point, and its size represents the length of fixation time.

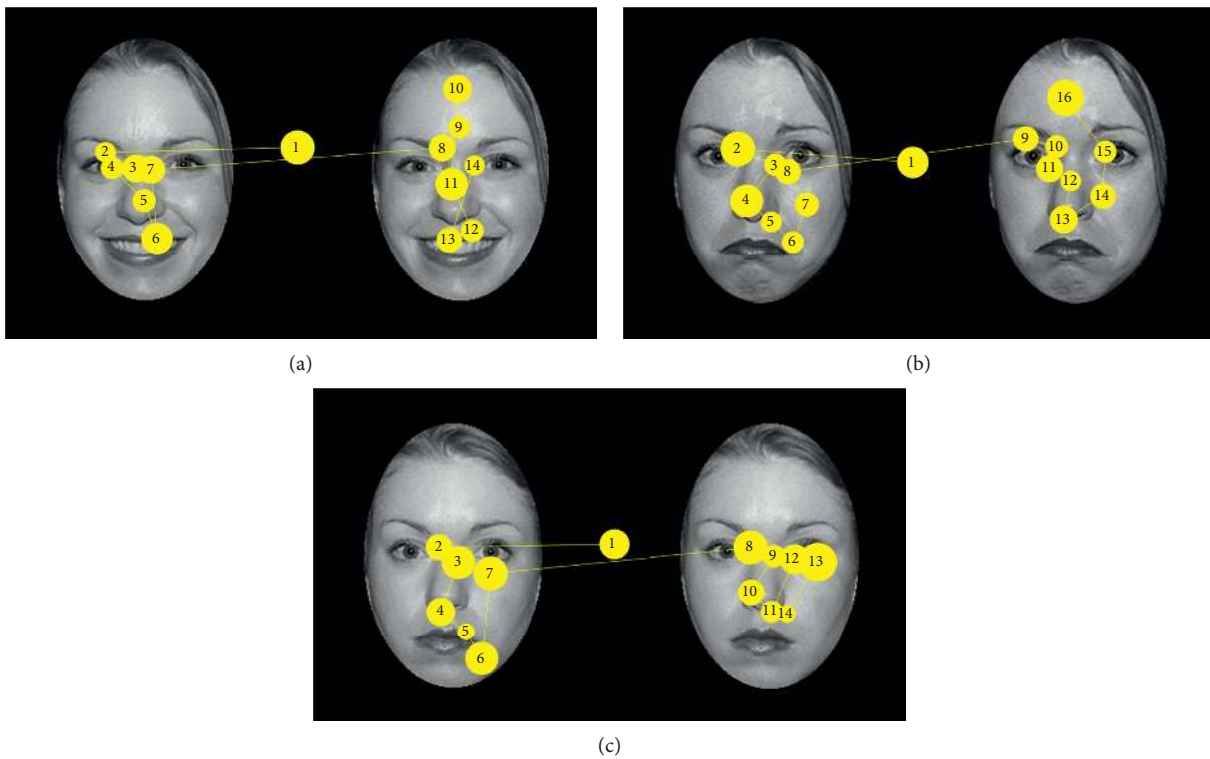


FIGURE 2: An example of eye movements when a depressed patient views (a) positive task, (b) negative task, and (c) neutral task. Each circle represents a gaze point, and its size represents the length of fixation time.

On this basis, the statistical indicators of each eye movement behavioral feature (positive attention bias and negative attention bias) and each eye movement physiological feature (positive/negative affective bandwidth, positive/negative pupil diameter change rate) are calculated, including minimum, lower quartile, median, upper quartile, maximum, mean, and standard deviation. This constitutes 42 classification features, including 14 behavioral features and 28 physiological features of eye movement, as shown in Table 1.

**2.3. Classification Feature Reduction.** Since there may be irrelevant and redundant information in eye movement

features, dimensionality reduction is carried out after normalization. Here, feature selection and feature extraction are used to complete dimensionality reduction.

**2.3.1. Feature Selection.** Feature selection is to select a subset of features that are effective for classification from all features. Here, the FDR corrected  $t$ -test was used to analyze whether there was a significant difference between the depression group and healthy control group ( $P < 0.05$ ). As shown in Table 1, only 24 of the 42 features showed significant differences among groups (including 11 behavioral and 13 physiological features).



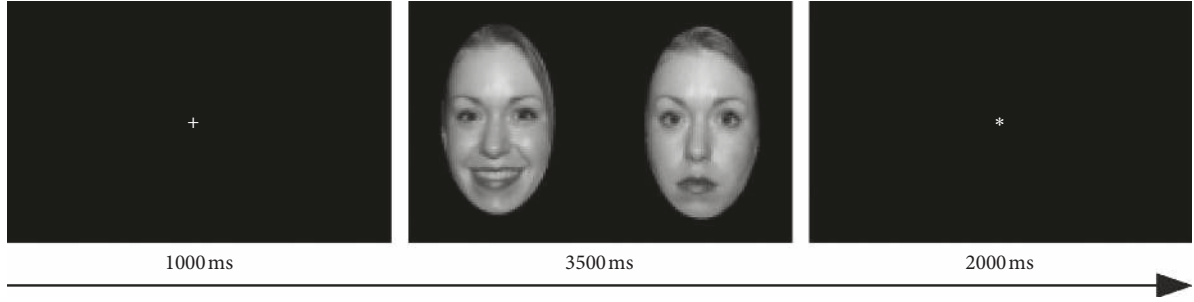


FIGURE 3: Data acquisition process.

TABLE 1: Analysis of differences between groups in 42 classification features (FDR corrected  $t$ -test).

	Minimum	Lower quartile	Median	Upper quartile	Maximum	Mean	Standard deviation
Positive bias	-1.47	1.73	5.71***	7.59***	6.58***	6.00***	6.30***
Negative bias	-5.45***	-6.99***	-8.25***	-8.3***	-8.25***	-8.15***	-0.10
Positive affective bandwidth	6.50***	6.49***	5.74***	4.68***	3.32**	5.97***	-2.06
Negative affective bandwidth	1.63	1.03	0.62	0.14	-1.02	0.50	-2.60*
Positive pupil diameter change rate	6.49***	6.95***	6.21***	5.44***	4.17***	6.70***	-1.14
Negative pupil diameter change rate	1.64	1.25	1.02	0.63	-0.24	0.92	-1.47

\* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

The positions of the 24 selected features in Table 1 are expressed as F1~F24 in the order of left to right and top to bottom. Among them, F1~F11 are the behavioral signal features and F12~F24 are the physiological signal feature.

**2.3.2. Feature Extraction.** There may be redundancy in the 24 features obtained by feature selection; that is, there may be correlation between features. Therefore, feature extraction is needed.

In order to determine whether there is redundant information between features, we use the Pearson correlation coefficient to calculate the correlation between these features. The calculation formula is as follows:

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X)^2 - E^2(X)}\sqrt{E(Y)^2 - E^2(Y)}} \quad (1)$$

where  $E(X)$  represents the mathematical expectation of variable  $X$ . When  $|\rho_{X,Y}| \geq 0.6$ , there is a strong correlation between variables  $X$  and  $Y$ . On the contrary, there is a weak correlation or no correlation between them.

The correlations between behavioral features and physiological features are shown in Figure 4. It can be seen that there is a strong correlation between some behavioral features (Figure 4(a)) and some physiological features (Figure 4(b)).

To eliminate redundant information and further reduce the dimension of features, principal component analysis (PCA) is used to extract features. The behavioral and physiological features are processed with PCA, respectively. Figure 5 shows the single variance contribution rate and accumulated variance contribution rate of behavioral features PCA (PCA1, PCA2, ..., PCA11) and physiological features PCA (PCA1, PCA2, ..., PCA13).

## 2.4. Classification Model

**2.4.1. KELM Classification Model.** The output function for extreme learning machine (ELM) is

$$f(x) = h(x)\mathbf{H}^T \left( \frac{I}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{O}, \quad (2)$$

where  $f(x)$  is the actual output,  $\mathbf{O}$  is the expected output of neural networks,  $C$  is the regularization factor,  $h(x)$  is the function of hidden layer, and  $\mathbf{H}$  is the output matrix of hidden layer.

Kernel extreme learning machine (KELM) introduces the kernel function based on ELM and solves the problem that the low-dimensional space is inseparable. The kernel function is defined as

$$\Omega_{ij} = h(x_i)h(x_j) = K(x_i, x_j), \quad (3)$$

$$\Omega = \mathbf{H}\mathbf{H}^T = \begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_N) \\ \vdots & \ddots & \vdots \\ K(x_N, x_1) & \cdots & K(x_N, x_N) \end{bmatrix}. \quad (4)$$

The final output function of KELM is

$$f(x) = h(x)\mathbf{H}^T \left( \frac{1}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{O} = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T \left( \frac{1}{C} + \Omega \right)^{-1} \mathbf{O}, \quad (5)$$

where  $N$  is the number of input samples.

Since the RBF kernel function has the advantage of strong learning ability and fewer parameters to be optimized,

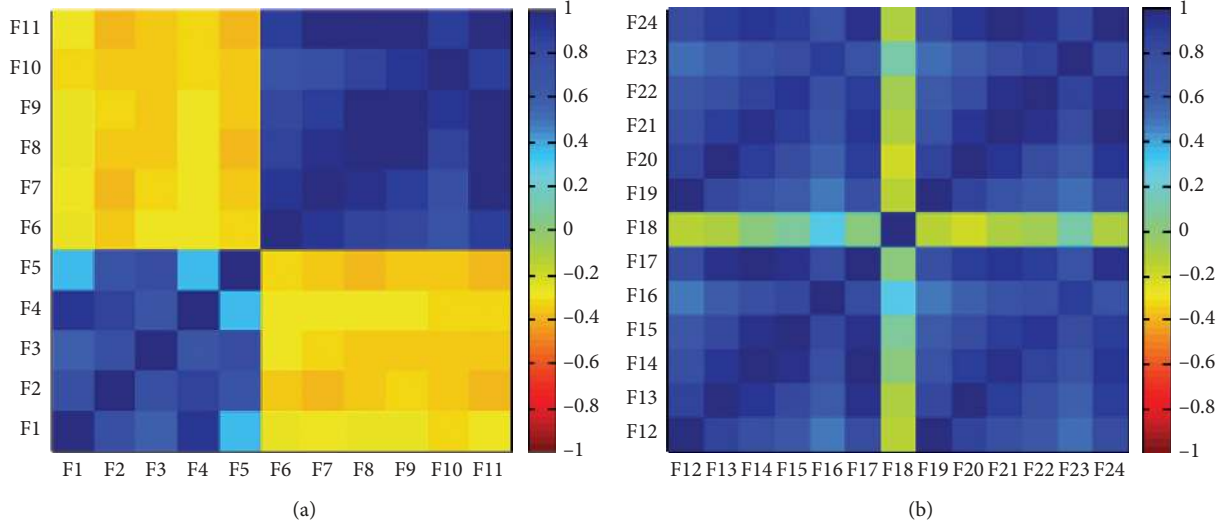


FIGURE 4: Analysis of correlations between selected features: (a) behavioral features; (b) physiological features.

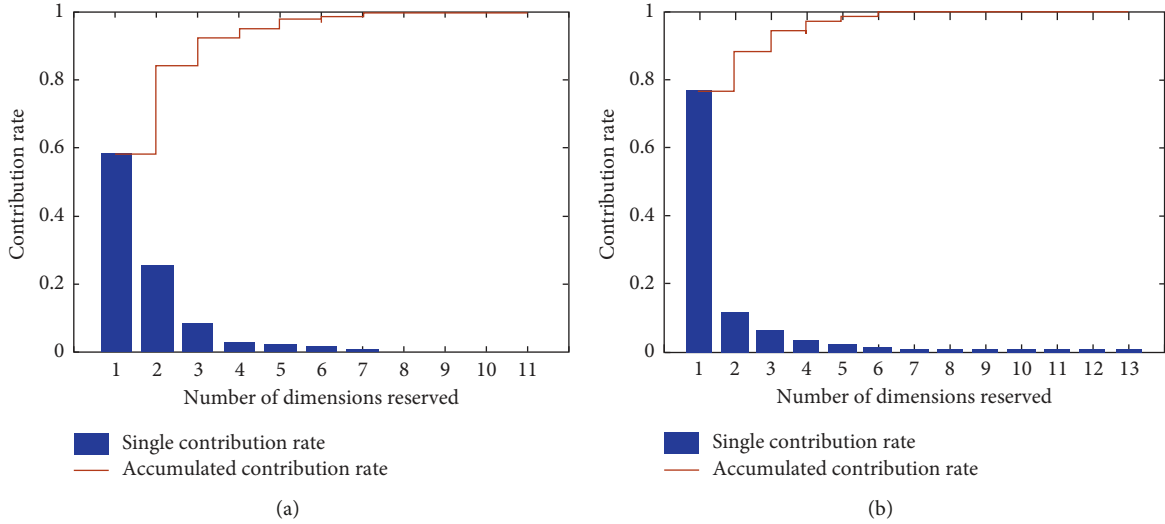


FIGURE 5: Single and accumulated variance contribution rate after PCA processing: (a) behavioral features; (b) physiological features.

RBF kernel function is adopted in this study, as shown in the following formula:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right). \quad (6)$$

**2.4.2. Model Training Strategy.** The number of participants was 96 (48 for normal subjects and 48 for depressed patients). The training process of the model uses 10-fold cross-validation. The training set is randomly divided into 10 disjoint subsets of the same size, in which 9 subsets are used as the training set and the remaining one is used as the validation set. Ten models are trained in turn, and the mean values of the classification results of the test set on 10 models are calculated.

### 3. Results and Discussion

In this section, the depressed patients group is regarded as positive class, while the normal control group is regarded as negative class. The results are analyzed from four aspects: accuracy, specificity, sensitivity, and F1 score. Among them, the accuracy reflects the ability of the model to distinguish depressed patients from normal people, the sensitivity reflects the proportion of depressed patients correctly classified, the specificity reflects the proportion of normal people correctly classified, and F1 score takes into account the precision and recall of the classification model.

**3.1. Selection of Optimal Feature Subset.** Feature selection is carried out inside the cross-validation folds to make the classification algorithm more robust. To do this, the top  $k$

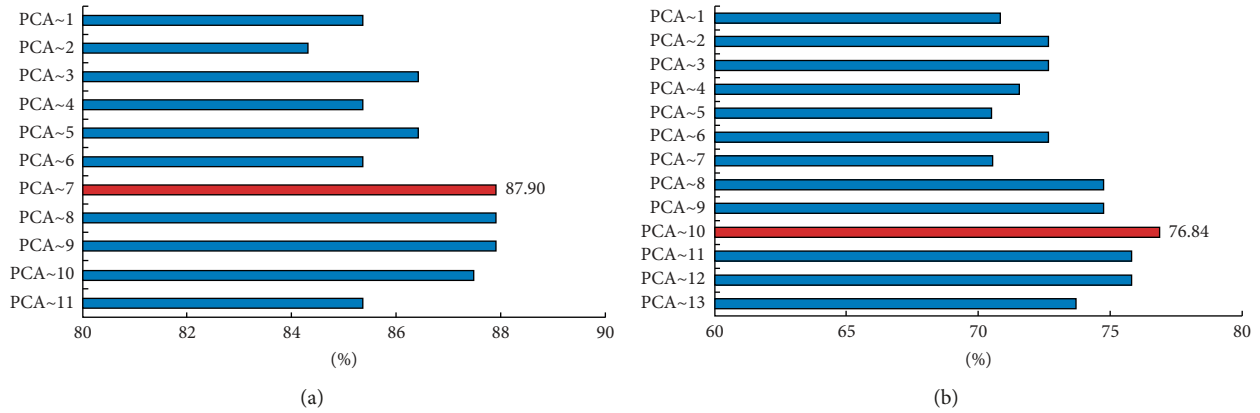


FIGURE 6: Classification accuracy of different PCA feature subsets: (a) behavioral features; (b) physiological features. PCA~m represents PCA1 + PCA2 + . . . + PCAm.

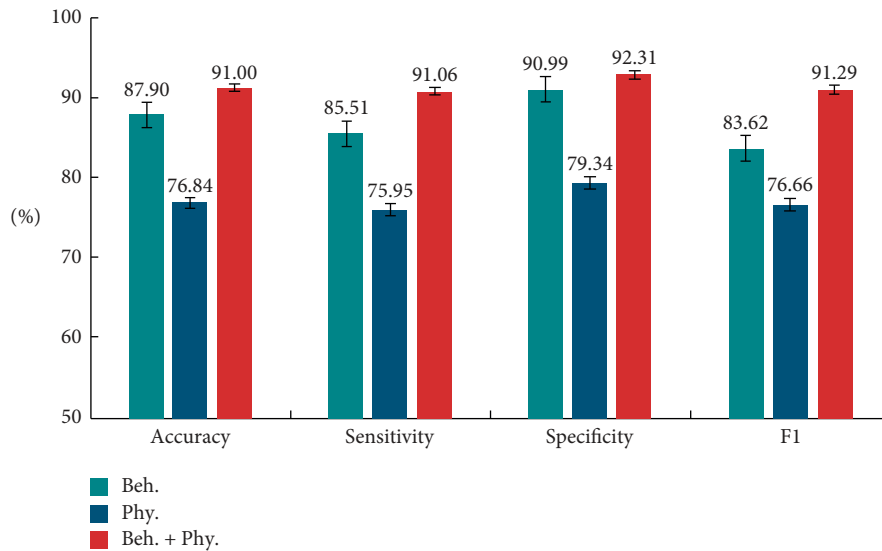


FIGURE 7: Classification results for different signal features. Beh. represents the eye movement behavioral signals; Phy. represents the eye movement physiological signals; Beh.+ Phy. represents the fusion features of eye movement behavioral and physiological signals.

PCA features ( $k$  is less than or equal to the number of PCA features) are selected iteratively as the classification features, and the 10-fold cross-validation method is used to train the model. Finally, the optimal feature subset is selected from multiple PCA feature subsets based on the classification accuracy.

Figure 6 shows the classification accuracy based on different PCA feature subsets. It can be seen that as the number of PCA features increases, the classification accuracy generally increases first and then decreases. According to the classification results, the optimal feature subset of behavioral signal PCA features is PCA1~PCA7, and the optimal feature subset of physiological signal PCA features is PCA1~PCA10.

**3.2. Comparison of Classification Results of Different Signal Features.** The classification results based on the features of eye movement behavioral signals, physiological signals, and their fusion signals are shown in Figure 7. It can be seen that

the classification accuracy, sensitivity, specificity, and F1 score of behavioral signals are much higher than those of physiological signals, which indicate that the difference of attentional function in the emotional information between depressed patients and normal people is much greater than that of pupil size changes. When behavioral and physiological signals are used together in modeling, the classification results are better than those using single eye movement behavioral signals and single physiological signals.

**3.3. Comparison with Other Methods.** Table 2 shows the comparison with other methods for depression classification based on eye movement signals. It can be seen that our proposed classification method is much better than other methods.

Table 3 shows the comparison with other methods of depression classification based on fusion signals, and our proposed classification method is also much better than other methods.

TABLE 2: Comparison with other methods of depression classification based on eye movement signals.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 (%)
Reference [17]	—	75	—	—
Reference [18]	80.1	81.1	—	—
Proposed	<b>91.00</b>	<b>91.06</b>	<b>92.31</b>	<b>91.29</b>

TABLE 3: Comparison with other methods of depression classification based on fusion signals.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 (%)
Reference [13]	80.95	72.73	90	—
Reference [14]	—	66.7	—	57.1
Reference [15]	—	83	—	77
Reference [16]	—	83.3	82.6	—
Proposed	<b>91.00</b>	<b>91.06</b>	<b>92.31</b>	<b>91.29</b>

TABLE 4: Comparison of depression classification methods based on fusion signals (behavior and physiology) and different classification models.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 (%)
SVM	88.95	90.48	88.23	87.86
KNN	90.21	90.74	90.60	89.56
Random forests	<b>91.37</b>	<b>91.32</b>	90.96	90.78
KELM	91.00	91.06	<b>92.31</b>	<b>91.29</b>

We also compare the effects of different classification models on the classification results, as shown in Table 4. It can be seen that the KELM model and random forest model have better classification results.

From the comparison of the classification results in Tables 2 and 3, we can see that the proposed classification results based on the fusion of eye movement behavioral and physiological signals are better. There are three main aspects: first, compared with the previous resting-state EEG/fMRI signals and expression/voice, we obtain emotional information of eye movement under external emotional stimulation, which reflects the current mood of depressed patients; second, we use the relative change indicators such as attention bias and affective bandwidth features as classification features and, based on this, extract statistical indicators (minimum, lower quartile, median, upper quartile, maximum, mean, and standard deviation) directly related to the data distribution, which effectively improves the difference between depressed patients and normal people; third, through statistical feature selection and PCA-based feature extraction, the irrelevant features and redundant information are effectively removed, and the classification

performance of depression is improved. In addition, compared with other methods, our proposed method also has the advantages of simple data acquisition and fewer features for classification.

## 4. Conclusion

In view of the lack of physiological and behavioral signals directly related to affective disorders in current research on depression recognition, this paper designs and extracts physiological signals based on the changes of affective bandwidth and pupil diameter and attentional bias signals based on gaze behavior. Statistical FDR corrected *t*-test and PCA are used for feature selection and feature extraction to eliminate noise and redundant information in behavioral and physiological signals of eye movement. The KELM classifier is used to classify depression based on behavioral features, physiological features, and fusion of behavior and physiological features. The results show that behavioral and physiological features are the main features of depression, and the fusion features improve the classification performance of depression.

In addition, we also use other classifiers (SVM, KNN, and random forest) to classify depression based on the fusion features of eye movement behavioral and physiological signals. The results show that the random forest model also achieved satisfactory classification results.

## Data Availability

Since these data relate to personal, private information of patients and health subjects, there are ethical restrictions. All original data referred to in the paper will be made available upon request to the corresponding author.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

Mi Li, Lei Cao, and Qian Zhai contributed equally to this work.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (61602017), the National Basic Research Programme of China (2014CB744600), "Rixin Scientist" Foundation of Beijing University of Technology (2017-RX(1)-03), the Beijing Natural Science Foundation (4164080), the Beijing Outstanding Talent Training Foundation (2014000020124G039), the National Natural Science Foundation of China (61420106005), the International Science & Technology Cooperation Program of China (2013DFA32180), the Special Fund of Beijing Municipal Science and Technology Commission (Z171100000117004), the Beijing Hospitals Authority Youth Programme (QML20181904), the Beijing Municipal Administration of Hospitals Clinical Medicine Development of Special



Funding Support (ZYLX201607), and the Beijing Municipal Administration of Hospitals' Ascent Plan (DFL20151801).

## References

- [1] L. Cao, S. Guo, Z. Xue et al., "Aberrant functional connectivity for diagnosis of major depressive disorder: a discriminant analysis," *Psychiatry and Clinical Neurosciences*, vol. 68, no. 2, pp. 110–119, 2014.
- [2] J. Qin, M. Wei, H. Liu et al., "Abnormal hubs of white matter networks in the frontal-parieto circuit contribute to depression discrimination via pattern classification," *Magnetic Resonance Imaging*, vol. 32, no. 10, pp. 1314–1320, 2014.
- [3] J. R. Sato, J. Moll, S. Green, J. F. W. Deakin, C. E. Thomaz, and R. Zahn, "Machine learning algorithm accurately detects fMRI signature of vulnerability to major depression," *Psychiatry Research: Neuroimaging*, vol. 233, no. 2, pp. 289–291, 2015.
- [4] R. Bhaumik, L. M. Jenkins, J. R. Gowins et al., "Multivariate pattern analysis strategies in detection of remitted major depressive disorder using resting state functional connectivity," *NeuroImage: Clinical*, vol. 16, pp. 390–398, 2017.
- [5] D. M. Schnyer, P. C. Clasen, C. Gonzalez, and C. G. Beevers, "Evaluating the diagnostic utility of applying a machine learning algorithm to diffusion tensor MRI measures in individuals with major depressive disorder," *Psychiatry Research: Neuroimaging*, vol. 264, pp. 1–9, 2017.
- [6] R. Ramasubbu, M. R. G. Brown, F. Cortese et al., "Accuracy of automated classification of major depressive disorder as a function of symptom severity," *NeuroImage: Clinical*, vol. 12, pp. 320–331, 2016.
- [7] S.-C. Liao, C.-T. Wu, H.-C. Huang, W.-T. Cheng, and Y.-H. Liu, "Major depression detection from EEG signals using kernel eigen-filter-bank common spatial patterns," *Sensors*, vol. 17, no. 6, p. 1385, 2017.
- [8] G. M. Bairy, O. S. Lih, Y. Hagiwara et al., "Automated diagnosis of depression electroencephalograph signals using linear prediction coding and higher order spectra features," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 8, pp. 1857–1862, 2017.
- [9] W. Mumtaz, L. Xia, S. S. A. Ali, M. A. M. Yasin, M. Hussain, and A. S. Malik, "Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD)," *Biomedical Signal Processing and Control*, vol. 31, pp. 108–115, 2017.
- [10] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, and D. P. Subha, "Automated EEG-based screening of depression using deep convolutional neural network," *Computer Methods and Programs in Biomedicine*, vol. 161, pp. 103–113, 2018.
- [11] M. Valstar, J. Gratch, B. Schuller et al., "Avec 2016: depression, mood, and emotion recognition workshop and challenge," in *Proceedings of the 6th International Workshop on Audio/visual Emotion Challenge*, pp. 3–10, ACM, Amsterdam, Netherlands, October 2016.
- [12] X. Ma, H. Yang, Q. Chen, D. Huang, and Y. Wang, "Depaudionet: an efficient deep model for audio based depression classification," in *Proceedings of the 6th International Workshop on Audio/visual Emotion Challenge*, pp. 35–42, ACM, Amsterdam, Netherlands, October 2016.
- [13] J. Zhao, G. Luo, Z. Yao, and Q. Lu, "Depression discrimination using fMRI and DTI data by wavelet based fusion scheme," *Journal of Southeast University*, vol. 28, no. 1, pp. 25–28, 2012.
- [14] Y. Le, D. Jiang, H. Lang, E. Pei, and H. Sahli, "Decision tree based depression classification from audio video and language information," in *Proceedings of the 6th International Workshop on Audio/visual Emotion Challenge*, pp. 89–96, ACM, Amsterdam, Netherlands, October 2016.
- [15] T. Al Hanai, M. M. Ghassemi, and J. R. Glass, "Detecting depression with audio/text sequence modeling of interviews," in *Proceedings of the Interspeech*, pp. 1716–1720, Hyderabad, India, September 2018.
- [16] A. Haque, M. Guo, A. S. Miner, and L. Fei-Fei, "Measuring depression symptom severity from spoken language and 3d facial expressions," 2018, <http://arxiv.org/abs/1811.08592>.
- [17] S. Alghowinem, R. Goecke, M. Wagner, G. Parker, and M. Breakspear, "Eye movement analysis for depression detection," in *Proceedings of the 2013 IEEE International Conference on Image Processing*, pp. 4220–4224, IEEE, Melbourne, Australia, September 2013.
- [18] X. Li, T. Cao, S. Sun, B. Hu, and M. Ratcliffe, "Classification study on eye movement data: towards a new approach in depression detection," in *Proceedings of the 2016 IEEE Congress on Evolutionary Computation*, pp. 1227–1232, Vancouver, BC, Canada, July 2016.
- [19] T. Armstrong and B. O. Olatunji, "Eye tracking of attention in the affective disorders: a meta-analytic review and synthesis," *Clinical Psychology Review*, vol. 32, no. 8, pp. 704–723, 2012.
- [20] A. Duque and C. Vázquez, "Double attention bias for positive and negative emotional faces in clinical depression: evidence from an eye-tracking study," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 46, pp. 107–114, 2015.
- [21] E. H. Hess and J. M. Polt, "Pupil size as related to interest value of visual stimuli," *Science*, vol. 132, no. 3423, pp. 349–350, 1960.
- [22] E. H. Hess, "Attitude and pupil size," *Scientific American*, vol. 212, no. 4, pp. 46–54, 1965.
- [23] R. A. Hicks, S. L. Williams, and F. Ferrante, "Pupillary attributions of college students to happy and angry faces," *Perceptual and Motor Skills*, vol. 48, no. 2, pp. 401–402, 1979.
- [24] M. Li, S. Lu, G. Wang, L. Feng, B. Fu, and N. Zhong, "Emotion, working memory, and cognitive control in patients with first-onset and previously untreated minor depressive disorders," *Journal of International Medical Research*, vol. 44, no. 3, pp. 529–541, 2016.
- [25] M. Li, S. Lu, L. Feng, F. Bingbing, and G. Wang, "Emotional experience and the mood-congruent working memory effect in first-onset and untreated depressive disorder patients," *Journal of Psychiatry*, vol. 19, no. 4, p. 2, 2016.
- [26] M. Li, S. Lu, G. Wang, and N. Zhong, "Affective bandwidth measurement and affective disorder determination," US Patent No. US9532711B2, 2017.
- [27] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [28] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513–529, 2011.
- [29] N. Tottenham, J. W. Tanaka, A. C. Leon et al., "The NimStim set of facial expressions: judgments from untrained research participants," *Psychiatry Research*, vol. 168, no. 3, pp. 242–249, 2009.

