Hindawi
Computational and Mathematical Methods in Medicine
Volume 2018, Article ID 6508319, 9 pages
https://doi.org/10.1155/2018/6508319

Research Article

Detecting Depression Using an Ensemble Logistic Regression Model Based on Multiple Speech Features

Haihua Jiang,[1] Bin Hu,[1] Zhenyu Liu,[2] Gang Wang,[3] Lan Zhang,[4] Xiaoyu Li,[2] and Huanyu Kang[2]

[1]*Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China*
[2]*Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China*
[3]*Beijing Anding Hospital of Capital Medical University, Beijing 100088, China*
[4]*Lanzhou University Second Hospital, Lanzhou 730030, China*

Correspondence should be addressed to Bin Hu; bh@bjut.edu.cn

Received 14 May 2018; Accepted 28 August 2018; Published 24 September 2018

Academic Editor: Raul Alcaraz

Early intervention for depression is very important to ease the disease burden, but current diagnostic methods are still limited. This study investigated automatic depressed speech classification in a sample of 170 native Chinese subjects (85 healthy controls and 85 depressed patients). The classification performances of prosodic, spectral, and glottal speech features were analyzed in recognition of depression. We proposed an ensemble logistic regression model for detecting depression (ELRDD) in speech. The logistic regression, which was superior in recognition of depression, was selected as the base classifier. This ensemble model extracted many speech features from different aspects and ensured diversity of the base classifier. ELRDD provided better classification results than the other compared classifiers. A technique for identifying depression based on ELRDD, ELRDD-E, was here suggested and tested. It offered encouraging outcomes, revealing a high accuracy level of 75.00% for females and 81.82% for males, as well as an advantageous sensitivity/specificity ratio of 79.25%/70.59% for females and 78.13%/85.29% for males.

1. Introduction

Worldwide, over 300 million people of different ages have clinical depression [1]. The rise in the prevalence of this disease has been connected to a group of important outcomes [2]. At the most extreme, patients with depression may commit suicide [3]. To halt the onset of clinical depression, advance intervention can offer a pivotal action to ease the burden of the disease. However, current depression diagnosis methods rely on self-report of patient and clinical opinion [4], which risk several subjective biases. Therefore, a convenient and objective method for detecting depression is of primary importance.

Depressed speech is distinguished invariably by clinicians as monotone, uninteresting, and spiritless [5]. The acoustic qualities of speech can be affected by the emotional state of a person with depression [6]. Therefore, depression can be detected by analyzing changes in the acoustical characteristics of speech. Several approaches have been proposed to reveal correlations between depression and acoustic features for depressed speech classification. To improve the effect of classification, many features were extracted in early studies. However, it is still unclear which acoustic features are most effective for detecting depression especially in Mandarin speech. Furthermore, an objective method based on speech is still in need.

This study investigates the classification performance of multiple speech features which were extracted from subjects to identify depression in those who spoke Mandarin language. To develop an effective objective method and improve the classification result, we propose an ensemble logistic regression model for detecting depression (ELRDD), which contributes to depression recognition based on speech in several ways. First, to make the best use of speech features, it extracts many speech features from different aspects and ensures diversity of the feature spaces and the base

classifiers. Second, to overcome the problem of dimensionality curse, the feature subspace dimensionality of each base classifier is lower than the all features space, while a feature reduction method is also used to avoid the curse of dimensionality. Third, a logistic regression model as the base classifier offers probabilities for every class, so the ensemble classifier could make the greatest use of the uncertain information to acquire the best classification outcomes.

The rest of the paper is structured as follows: Section 2 reviews the related work. Section 3 describes the speech database used for this study. Section 4 provides a detailed description of our methodology. Section 5 describes the experiments and results, and Section 6 presents the conclusions.

2. Related Work

Darby and Hollien [7] performed an introductory evaluation of patients with major depression, and they discovered that listeners could discern various distinct characteristics in depressed speech. A variety of speech features have been explored for detecting depression. Mundt et al. [4], Stassen et al. [8], and Hönig et al. [9] reported correlations between F_0 variables and depression. However, Alpert et al. [10], Cannizzaro et al. [11], and Yang et al. [12] reported no significant correlation between F_0 variables and depression. Low et al. [13], Moore et al. [14], and Ooi et al. [15, 16] evaluated classification systems with prosodic, glottal, and spectral features. Low et al. [17], Valstar et al. [18], Alghowinem et al. [19], and Jiang et al. [20] used low-level descriptors and statistical characteristics to identify depression. Cummins et al. [21, 22], Sturim et al. [23], Alghowinem et al. [24], and Joshi et al. [25] investigated mel-frequency cepstrum coefficients (MFCC) and found that the recognition performance was statistically significant for depression classification. An evaluation by Scherer et al. [26–28] revealed a tight connection between voice quality features and the degree of depression. Quatieri and Malyska [29] and Ozdas et al. [30] discovered that depressed subjects showed increased energy levels on the glottal spectrum.

The support vector machine (SVM) and the Gaussian mixture model (GMM) are the most popular classification technologies used for detecting depression in speech. Moore et al. [14] studied 15 depressed subjects and 18 healthy controls and used quadratic discriminant analysis to construct a classifier. They reported accuracies of 91% (with sensitivity to specificity 89%/93%) for males and 96% (with sensitivity to specificity 98%/94%) for females. Their analysis showed that glottal features were more discriminating than prosodic features. Cohn et al. [31] recruited 57 depressed patients and used fundamental frequency and speak-switch duration as inputs to a logistic regression (LR) classifier. They reported an accuracy of 79% (with sensitivity to specificity 88%/64%) when classifying subjects who either responded or did not respond to treatment for depression. Low et al. [13] examined 139 adolescents (71 healthy and 68 depressed) who spoke English, and they used a gender-independent GMM classifier that incorporated glottal, prosodic, and spectral features. They reported classification results of 67–69% for males and 70–75% for females. Ooi et al. [16] studied 30 participants (15 were at risk of depression and 15 were not at risk) who spoke English and presented an ensemble method using GMM classifiers that used prosodic and glottal features. They reported a classification result of 74% (with sensitivity to specificity 77%/70%). Alghowinem et al. [32] recruited 30 controls and 30 depressed patients who spoke English. They summarized low-level descriptors and statistical features and compared the following classifiers: SVM, GMM, Multilayer Perceptron Neural Network (MLP), and Hierarchical Fuzzy Signature (HFS). They concluded that SVM and GMM had better classification performance. Helfer et al. [33] studied 35 subjects whose Hamilton Depression Scale (HAMD) scores were below 7 or above 17, respectively. They used associated dynamic and the first three formant trajectories as features and reported that SVM performed better than GMM when classifying depression severity. Jiang et al. [20] studied 170 subjects and proposed a computational methodology based on SVM (STEDD). They documented accuracies of 75.96% (with sensitivity to specificity of 77.36%/74.51%) for females and 80.30% (with sensitivity to specificity of 75.00%/85.29%) for males. It should be noted that most of these previous studies were usually limited to small depressed samples and focused on participants who spoke Western languages.

To the best of our knowledge, there has been little research exploring the ensemble classifier for detecting depression based on speech. However, ensemble logistic regression has been used effectively in other research fields [34–39]. In these previous studies, two methods were used to deal with the feature spaces. In one method, all feature spaces were used in each base classifier [34–36]. In the other method, the feature spaces were randomly partitioned into several subspaces [37–39]. It should be mentioned that the feature subspace dimensionality of the previous remained higher, and the variety of the feature subspaces of the last-mentioned could not be guaranteed and the classification outcome was unsteady.

3. Speech Database

In our research, all the subjects were native Chinese speakers between the ages of 18 and 55 and had at least an elementary school education [40]. First, every participant was required to fill in a preassessment booklet that contained general information and demographic information, including health history, age, gender, educational status, and employment. Second, every participant was chosen by psychiatrists based on the *Diagnostic and Statistical Manual Of Mental Disorders* (DSM-IV) [41] rules. Finally, all the subjects were interviewed by psychiatrists to complete the patient health questionnaire-9 (PHQ-9) [42]. These subjects were then divided into two groups depending upon the PHQ-9 scores: depressed patients (PHQ-9 $\geq$ 5) and healthy controls (PHQ-9 $<$ 5). Depressed patients were diagnosed as having pure depression, and they did not experience any other mental illnesses. The controls had no previous or ongoing mental disorder and were matched to the depressed patients based on demographics.

Following the completion of the clinical evaluations, our recording experiment began and it consisted of three parts:

an interview assignment, a reading assignment, and a picture detailing assignment. The interview assignment was made up of 18 questions, and the topics were taken from the Self-Rating Depression Scale (SDS), HAMD, and DSM-IV. The following are sample questions: How do you evaluate yourself? What is the most important present you have ever been given, and how did it make you feel? What do you enjoy doing when you are not able to fall asleep? Please detail any plans you may have for an upcoming vacation. Please tell us about a friend, including their age, type of employment, personality, and pastimes. What situations could make you become desperate? The reading assignment consisted of a short story named "*The North Wind and the Sun*" [43] and three sets of words with neutral (e.g., center, since), positive (e.g., outstanding, happy), and negative (e.g., depression, wail) emotions. The picture detailing assignment involved four dissimilar pictures. Three of them, which had neutral, positive, and negative faces, were obtained from the Chinese Facial Affective Picture System (CFAPS). The last picture titled "*Crying Woman*" was chosen from the Thematic Apperception Test (TAT) [9]. In this assignment, participants were requested to openly detail the four pictures.

We collected speech recordings in a quiet, soundproof, clean laboratory. The ambient noise level in the laboratory was kept below 60 dB. The speech signals were documented with a 24-bit sampling depth and 44.1 kHz sampling rate. We segmented and labeled all these recordings manually and retained only subject voice signals. These recordings were stored in an uncompressed WAV format. The database utilized in this evaluation contained speech recordings from 85 controls (34 males and 51 females) and 85 depressed individuals (32 males and 53 females). The speech of each subject was split into 29 recordings depending on different subtasks. In all, this study utilized 4,930 speech recordings. The overall lengths of speech during the interview, picture detailing, and reading were 52,427 s, 16,203 s, and 21,425 s, respectively. The average duration of speech recording was 18.3 s.

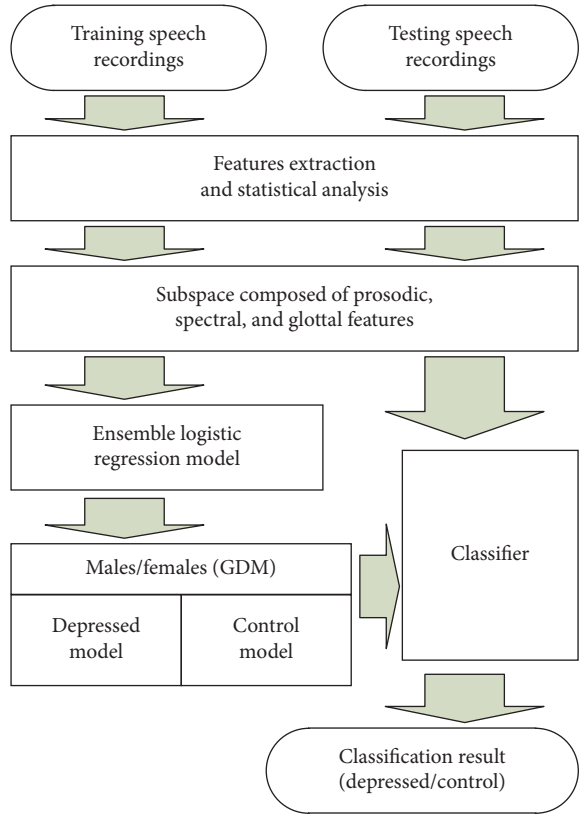

Figure 1: Block diagram of the ensemble logistic regression model for detecting depression (ELRDD).

4. Methods

In light of gender variations in depressive indications [44], there are two classification methods: gender-independent modeling (GIM) and gender-dependent modeling (GDM). Low et al. [13] discovered that GDM outperformed GIM. In our study, we used GDM, in which females and males were modeled independently. The proposed framework for the ELRDD is detailed in Figure 1. In the next sections of our paper, features extraction, features reduction, and modeling techniques are recounted.

4.1. Features Extraction and Reduction. The acoustic speech features explored in the literature can be divided into three main categories: prosodic features, spectral features, and glottal features. Each of the three categories comprises several subcategories. MFCC was one of the most frequent spectral features utilized in speech parameters, and the classification outcomes were statistically significant in identifying depression [22–25]. Therefore, MFCC was separated from spectral features as a main category. For convenience, the prosodic features are abbreviated PROS, the spectral features are abbreviated SPEC, and the glottal features are abbreviated GLOT. Table 1 presents a summary of the main speech feature categories, subcategories, the number of features, and the statistical functions. Since PROS, SPEC, MFCC, and GLOT were extracted using very different feature extraction methods, they can describe speech from diverse aspects. Thus, feature vectors were complementary to one another. Then, if the feature subspaces of the ensemble classifier were made up of a few of these feature vectors, these subspaces will have a larger diversity. We combined one or more of these four features to form 15 different feature spaces. Table 2 displays these subspaces made up of various feature vectors, in which PROS + SPEC suggests that the subspace is made up of the feature vectors of PROS and SPEC, and MFCC + PROS + SPEC + GLOT suggests that the space is made up of every one of the features. The glottal features were calculated using the TTK Aparat toolbox [45], and the prosodic and spectral features were calculated using the open-source software openSMILE [46].

Compared with the dimensionality of the whole feature space, the dimensionalities of feature subspaces had been reduced considerably, but some dimensionalities of the subspaces were still very high. We applied and compared principal component analysis (PCA), kernel PCA, Laplacian, Isomap, Landmark Isomap, and locally linear embedding (LLE) to reduce feature space dimensionality. We employed

Table 1: Summary of speech features.

Main category	Subcategory	Number of features	Functions
MFCC	MFCC (0–14)	630	Corresponding delta coefficients appended
SPEC	Flux	42	21 functions utilized
	Centroid	42	maxPos, minPos
	Entropy	42	Mean, std dev
	Roll-off	168	Skewness, kurtosis
	Band energies	84	Quartile 1/2/3
PROS	PCM loudness	42	Quartile range (2–1)/(3–2)/(3–1)
	Log mel-frequency band (0–7)	336	Linear regression error Q/A
	LSP frequency (0–7)	336	Linear regression coeff. 1/2
	F_0 envelope	42	Percentile 1/99
	Voicing probability	42	Percentile range (99–1)
	F0final, ShimmerLocal	76	19 functions by eliminating the minimum value and the
	JitterLocal, JitterDDP	76	Range functions from the 21 abovementioned functions
	Pitch onsets, duration	2	No functions
GLOT	GLT	27	Mean, max, min
	GLF	5	Mean, max, min
Total		1992	

Table 2: Subspaces composed of several different feature vectors.

No.	Subspace	No.	Subspace	No.	Subspace
1	MFCC	2	PROS	3	SPEC
4	GLOT	5	MFCC + PROS	6	MFCC + SPEC
7	MFCC + GLOT	8	PROS + SPEC	9	PROS + GLOT
10	SPEC + GLOT	11	MFCC + PROS + SPEC	12	MFCC + PROS + GLOT
13	MFCC + SPEC + GLOT	14	PROS + SPEC + GLOT	15	MFCC + PROS + SPEC + GLOT

LLE, because it outperformed other methods and preserved the local geometry of high dimensional data [47].

4.2. Ensemble Classification. Given that training data $X = \{X^{(1)}, X^{(2)}, \ldots, X^{(K)}\}$ and its label $Y = \{Y^{(1)}, Y^{(2)}, \ldots, Y^{(K)}\}$, where $X^{(k)} = \{x_1^{(k)}, x_2^{(k)}, \ldots, x_N^{(k)}\}$ is one of the feature subspaces of training points, the value of each label in $Y^{(k)} = \{y_1^{(k)}, y_2^{(k)}, \ldots, y_N^{(k)}\}$ was set to 1 for the depressed patients and 0 for the controls. Given test input data $x = \{x^{(1)}, x^{(2)}, \ldots, x^{(K)}\}$, where $x^{(k)}$ included in Table 2 is a feature subspaces of x, the outputs $P(L = 1|x)$ and $P(L = 0|x)$ providing the 1 and 0 estimated probabilities are given by

$$P(L = 1|x) = \sum_{k=1}^{K} P\left(L = 1|x^{(k)}; w^{(k)}\right) = \sum_{k=1}^{K} \frac{\exp\left(w^{(k)} \cdot x^{(k)}\right)}{1 + \exp\left(w^{(k)} \cdot x^{(k)}\right)}, \quad (1)$$

$$P(L = 0|x) = \sum_{k=1}^{K} P\left(L = 0|x^{(k)}; w^{(k)}\right) = \sum_{k=1}^{K} \frac{1}{1 + \exp\left(w^{(k)} \cdot x^{(k)}\right)}, \quad (2)$$

where $w = \{w^{(1)}, w^{(2)}, \ldots, w^{(K)}\}$ are the parameters of the ensemble logistic regression model. The log-likelihood function under this model is as follows:

$$l(w) = \sum_{k=1}^{K} l\left(w^{(k)}\right) = \sum_{k=1}^{K} \sum_{i=1}^{N} \left[y_i^{(k)}\left(w^{(k)} \cdot x_i^{(k)}\right) - \log\left(1 + \exp\left(w^{(k)} \cdot x_i^{(k)}\right)\right)\right], \quad (3)$$

where maximizing $l(w)$ produces a maximum likelihood estimator for w.

According to Section 4.1, the complete algorithm of ELRDD is outlined in Algorithm 1.

To validate ELRDD, SVM, GMM, and LR were compared as classifiers for detecting depression. SVM and GMM were usually employed for recognition of depression, while LR was taken as the base classifier for ELRDD. We utilized the expectation-maximization (EM) algorithm to approximate the GMM parameters of every Gaussian component and a radial basis function (RBF) as SVM's kernel function. Then, we looked for the most adequate parameters with a grid search utilizing five-fold cross validation on our training dataset with the LIBSVM toolbox [48].

To demonstrate that ELRDD outperforms other ensemble classifiers, three classic classifiers were compared: adaboost decision tree, bagging decision tree, and random forest. They depicted the speech recordings by the feature spaces made up of MFCC, PROS, SPEC, and GLOT. Because males and females were modeled separately, the number of base classifiers for each classifier depended on gender. These numbers were chosen from 15, 50, 100, 200,

Input: training speech recordings $\{s_1, s_2, \ldots, s_n\}$ and its label $\{y_1, y_2, \ldots, y_n\}$ and testing speech recordings $\{r_1, r_2, \ldots, r_m\}$.
Output: depressed patient or healthy control labels of $\{r_1, r_2, \ldots, r_m\}$.
//Training process
Step 1: extract MFCC, PROS, SPEC, and GLOT features for each speech recording from $\{s_1, s_2, \ldots, s_n\}$, and compute the feature statistics as listed in Table 1.
Step 2: in terms of Table 2, 15 feature subspaces are constructed $\{X^{(1)}, X^{(2)}, \ldots, X^{(15)}\}$, where $\{X(k) = x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}\}$.
For $k = 1$ to 15
Step 3: feature reduction for $X(k)$ is achieved using LLE.
End
Step 4: maximize Equation (3) to achieve the trained classifier model.
//Testing process
Step 5: extract MFCC, PROS, SPEC, and GLOT features for each speech recording from $\{r_1, r_2, \ldots, r_m\}$, and compute the feature statistics as listed in Table 1.
Step 6: in terms of Table 2, 15 feature subspaces are constructed $\{D^{(1)}, D^{(2)}, \ldots, D^{(15)}\}$, where $\{D(k) = d_1^{(k)}, d_2^{(k)}, \ldots, d_n^{(k)}\}$.
For $k = 1$ to 15
Step 7: feature reduction for $D(k)$ is achieved using LLE.
End
Step 8: based on the trained classifier model, apply Equations (1) and (2) to compute the probabilities that the testing samples belong to depressed patients $\{p_1, p_2, \ldots, p_m\}$ or healthy controls $\{q_1, q_2, \ldots, q_m\}$, and then output the category label whose probability is greater.

Algorithm 1: ELRDD.

300, 400, and 500, which yielded the best classification outcomes.

ELRDD computed the probabilities that each speech recording belonged to depressed and healthy subjects. To improve recognition performance, classifying a subject as depressed patient or healthy control could use the classification results of more than one speech recording. In our study, each participant had 29 speech recordings, and the final classification result could depend on all these speech recordings. Therefore, we proposed ELRDD-E, which can be summarized as Algorithm 2.

We examined the precise classifications of the controls and the depressed patients in sensitivity, specificity, and accuracy. The controls were distinguished as the negative cases, and the depressed patients were distinguished as the positive cases. When examining performance, each of the three parameters of a well-performing method would have high values, but if a compromise was required, it was sensible to acquire the greatest accuracy while obtaining an optimum sensitivity/specificity ratio (ideally > 1). We employed a speaker-independent split of test and train data and used a ten-fold cross validation. The one-way analysis of variance (ANOVA) and the least significant difference (LSD) tests were conducted to establish if variations in the classification outcomes were statistically significant. The level of significance was set as $p < 0.05$.

5. Experiments and Results

5.1. Experiment Using Individual and Ensemble Classifiers for Males. Table 3 reveals the classification outcomes of every individual classifier for males. It can be noted that the chosen speech features impacted the recognition performance of classifiers. For example, SVM had the best specificity and accuracy with SPEC + GLOT. In contrast, LR achieved the best specificity and accuracy with PROS + SPEC, and GMM achieved the best accuracy with MFCC + PROS + SPEC. In addition, ANOVA and LSD tests were carried out on the four speech feature subspaces (MFCC, PROS, SPEC, and GLOT) over the ten-fold cross validation outcomes utilizing SVM, GMM, and LR classifiers. The accuracy, sensitivity, and specificity significantly varied between the four feature subspaces ($p < 0.05$). The accuracy and sensitivity of GLOT were worse in comparison to MFCC, PROS, and SPEC ($p < 0.05$), and the accuracy and specificity of SPEC and PROS were greater than MFCC ($p < 0.05$). ANOVA and LSD tests were also carried out on paired classifiers over the ten-fold cross validation outcomes. The specificity and accuracy of SVM, GMM, and LR were alike ($p > 0.05$), and the sensitivity of LR and GMM was greater than SVM ($p < 0.05$).

Table 4 shows the recognition performance of ELRDD and existing ensemble classifiers for males. The number of base classifiers for adaboost decision tree, bagging decision tree, and random forest was set to 300, 500, and 400, respectively, which yielded the best classification results. From Tables 3 and 4, it was discovered that ELRDD outperformed the greatest outcome of individual classifiers in accuracy and sensitivity in the identification of depression. Following ANOVA and LSD tests being conducted on paired ensemble classifiers over the ten-fold cross validation outcomes, we discovered that ELRDD also outperformed the contrasted current ensemble classifiers for males in sensitivity and accuracy ($p < 0.05$), and specificity was alike ($p > 0.05$).

5.2. Experiment Using Individual and Ensemble Classifiers for Females. Table 5 reveals the classification outcomes of every individual classifier for females. Following ANOVA and LSD tests being carried out on paired classifiers over the

Input: training speech recordings $\{s_1, s_2, \ldots, s_{n*29}\}$ of subject $\{b_1, b_2, \ldots, b_n\}$ and its label $\{y_1, y_2, \ldots, y_{n*29}\}$ and testing speech recordings $\{r_1, r_2, \ldots, r_{29}\}$ of subject g.
Output: depressed patient or healthy control label of subject g.
Step 1: call the training process of ELRDD; the inputs are $\{s_1, s_2, \ldots, s_n\}$ and $\{y_1, y_2, \ldots, y_{n*29}\}$.
For $k = 1$ to 29
Step 2: call the testing process of ELRDD; the input is r_k of subject g and the outputs are probability pk for depressed patients and probability q_k for healthy controls.
End
Step 3: $p = p_1 + p_2 + \cdots + p_{29}, q = q_1 + q_2 + \cdots + q_{29}$, if the value of p is larger than q, subject g is classified as depressed; otherwise, g is classified as a control.

Algorithm 2: ELRDD-E.

Table 3: Classification outcomes of each individual classifier for males.

Features	SVM			GMM			LR		
	Sen. (%)	Spe. (%)	Acc. (%)	Sen. (%)	Spe. (%)	Acc. (%)	Sen. (%)	Spe. (%)	Acc. (%)
MFCC	56.14	64.91	60.66	62.72	58.22	60.40	62.50	60.75	61.60
PROS	61.96	70.39	66.30	61.75	74.14	68.13	63.15	71.10	67.24
SPEC	**63.36**	73.94	68.81	65.84	71.60	68.81	**67.35**	70.69	69.07
GLOT	36.32	60.95	49.01	47.95	54.26	51.20	44.07	54.56	49.48
MFCC + PROS	60.67	69.78	65.36	63.69	70.49	67.19	65.41	68.36	66.93
MFCC + SPEC	59.05	72.72	66.09	64.55	69.17	66.93	63.58	69.07	66.41
MFCC + GLOT	53.56	66.53	60.24	61.96	60.65	61.29	61.10	60.75	60.92
PROS + SPEC	63.25	73.83	68.70	62.72	74.14	68.60	67.13	**72.21**	**69.85**
PROS + GLOT	60.99	71.60	66.46	61.96	72.92	67.61	62.82	71.20	67.14
SPEC + GLOT	62.61	**75.15**	**69.07**	65.19	70.89	68.13	66.70	70.99	68.91
MFCC + PROS + SPEC	60.99	72.92	67.14	65.63	72.31	**69.07**	64.55	70.39	67.56
MFCC + PROS + GLOT	59.59	72.62	66.30	63.36	69.27	66.41	62.82	67.24	65.10
MFCC + SPEC + GLOT	60.24	72.82	66.72	65.84	69.98	67.97	64.66	68.66	66.72
PROS + SPEC + GLOT	60.02	73.83	67.14	62.82	**74.24**	68.70	64.12	71.91	68.13
MFCC + PROS + SPEC + GLOT	61.85	73.12	67.66	**66.27**	71.30	68.86	64.33	**72.21**	68.39

Maximum of sensitivity (sen.), specificity (spe.), and accuracy (acc.) is shown in bold.

Table 4: Recognition performance of each classifier for males.

Classifier	Number of base classifiers	Sensitivity (%)	Specificity (%)	Accuracy (%)
Adaboost decision tree	300	58.94	67.14	63.17
Bagging decision tree	500	59.48	70.28	65.05
Random forest	400	59.05	70.99	65.20
ELRDD	15	67.35	73.94	70.64

ten-fold cross validation outcomes, it can be noted that LR functioned as well as SVM and GMM ($p > 0.05$), and the greatest experimental outcome of LR outperformed SVM and GMM, which was in agreement with the outcomes for males. In addition, ANOVA and LSD tests were also conducted on the four speech feature subspaces (MFCC, PROS, SPEC, and GLOT) over the ten-fold cross validation outcomes utilizing SVM, GMM, and LR classifiers for females. The accuracy and specificity significantly varied between the four feature subspaces ($p < 0.05$). The accuracy and specificity of GLOT were worse than that of MFCC, PROS, and SPEC ($p < 0.05$), and the sensitivity of PROS was better than that of GLOT ($p < 0.05$).

Table 6 shows the recognition performances of ensemble classifiers for females. The number of base classifiers for adaboost decision tree, bagging decision tree, and random forest was set to 200, 300, and 300, respectively, which yielded the best classification results. After the LSD test, ELRDD still outperformed the other ensemble classifiers for females in terms of sensitivity ($p < 0.05$), and specificity and accuracy were similar ($p > 0.05$).

5.3. Experiment Using ELRDD-E. The classification outcomes of ELRDD-E are presented in Table 7. The outcomes of utilizing STEDD [20], which is an efficient technique according to speech types and emotions to identify depression in the identical database, are also included for comparison. From this table, it was found that ELRDD-E outperformed the results of Adaboost Decision Tree, Bagging Decision Tree, and Random Forest in terms of accuracy and sensitivity ($p < 0.05$). It also can be noted that ELRDD-E performed greater than STEDD in classification sensitivity and accuracy for males, while they had the same specificity. Further, ELRDD-E provided better sensitivity than STEDD for females, while STEDD performed minutely better in specificity and accuracy. It can be concluded that ELRDD-E provided very promising results and was effective for detecting depression.

Table 5: Classification outcomes of each individual classifier for females.

Features	SVM			GMM			LR		
	Sen. (%)	Spe. (%)	Acc. (%)	Sen. (%)	Spe. (%)	Acc. (%)	Sen. (%)	Spe. (%)	Acc. (%)
MFCC	63.24	57.27	60.31	56.47	66.06	61.17	62.79	61.80	62.30
PROS	67.21	60.65	63.99	51.72	73.29	62.30	64.35	66.73	65.52
SPEC	60.64	63.35	61.97	52.44	73.70	62.87	63.05	64.91	63.96
GLOT	56.60	42.53	49.70	51.33	50.44	50.90	52.70	46.11	49.47
MFCC + PROS	**67.53**	61.06	64.36	56.86	71.54	64.06	**64.93**	66.06	65.48
MFCC + SPEC	66.10	61.60	63.89	**57.78**	69.78	63.66	63.24	65.99	64.59
MFCC + GLOT	63.05	57.20	60.18	55.63	64.84	60.15	62.66	60.24	61.47
PROS + SPEC	64.09	**64.50**	64.29	51.01	**73.83**	62.20	63.63	**67.61**	65.58
PROS + GLOT	67.47	59.16	63.40	52.31	72.08	62.00	63.37	66.73	65.02
SPEC + GLOT	61.09	60.31	60.71	51.79	70.99	61.21	61.87	62.75	62.30
MFCC + PROS + SPEC	64.74	62.41	63.59	56.47	73.09	**64.62**	64.41	67.41	65.88
MFCC + PROS + GLOT	67.08	62.27	**64.72**	55.50	72.62	63.89	64.74	67.14	**65.92**
MFCC + SPEC + GLOT	63.37	62.68	63.03	57.71	69.37	63.43	62.39	63.42	62.90
PROS + SPEC + GLOT	64.15	63.42	63.79	51.53	73.43	62.27	63.11	67.07	65.05
MFCC + PROS + SPEC + GLOT	65.00	63.22	64.13	56.02	72.96	64.32	63.44	**67.61**	65.48

Maximum of sensitivity (sen.), specificity (spe.), and accuracy (acc.) are shown in bold.

Table 6: Recognition performance of each classifier for females.

Classifier	Number of base classifiers	Sensitivity (%)	Specificity (%)	Accuracy (%)
Adaboost decision tree	200	59.34	69.91	64.52
Bagging decision tree	300	58.75	68.56	63.56
Random forest	300	59.66	68.56	64.03
ELRDD	15	65.71	67.68	66.68

Table 7: Classification outcomes of ELRDD-E.

Gender	Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)
Male	ELRDD-E	78.13	85.29	81.82
	Adaboost decision tree	65.63	82.35	74.24
	Bagging decision tree	65.63	79.41	72.73
	Random forest	62.50	79.41	71.21
	STEDD	75.00	85.29	80.30
Female	ELRDD-E	79.25	70.59	75.00
	Adaboost decision tree	64.15	76.47	70.19
	Bagging decision tree	62.26	74.51	68.27
	Random forest	66.04	76.47	71.15
	STEDD	77.36	74.51	75.96

6. Discussion

Table 3 shows the classification outcomes of each individual classifier for males. It can be noted that the optimal features for every classifier varied. These results indicate that each feature vector could provide complementary information for the different classifiers. Moreover, it was impossible for each classifier to utilize the same feature subspace that worked best for other classifiers. This indirectly indicates that it is necessary to develop classifiers from multiple feature subsets. Results showed that SPEC and PROS features performed better than MFCC and GLOT features for males.

Table 5 reveals the classification outcomes of every individual classifier for females. It can be observed that each classifier yielded the best classification result using different feature subspaces. These outcomes suggest that every feature was complementary and offered various classifiers with different information, which was also in agreement with the discoveries for males. It was noted that utilizing SPEC, PROS, and MFCC features offered significantly better classification outcomes for females compared to utilizing GLOT features.

From Tables 3 and 5, it can be concluded that using GLOT features provided worst classification outcomes among these four feature vectors. This result is contrary to the findings of two earlier studies. Low et al. [13] and Ooi et al. [15] observed that glottal features performed better than prosodic and spectral features. The disparity may be due to the fact that previous researchers focused on participants who spoke Western languages, while all the participants in this work spoke Mandarin. The assignments used in the previous studies were also different from ours. It also can be observed that the performance of LR was no worse than that of SVM and GMM with most feature subspaces. Furthermore, the best experimental result of LR outperformed SVM and GMM. This was one of the reasons that LR was chosen as the base classifier.

Tables 4 and 6 show the recognition performances of classifiers. It can be observed that ELRDD had a better recognition effect than other classifiers. This result could be due to the fact that ELRDD could ensure the diversity of the feature subspaces and utilize more information provided by features. Moreover, compared with the other three existing ensemble classifiers, the number of base classifiers in ELRDD was much smaller.

7. Conclusion

In this evaluation, we initially contrasted the outcomes of three varying individual classifiers utilizing 15 feature

subspaces to determine the connection between speech features and the performance of classifiers. It was observed that classifier performance was sensitive to the features used for both males and females. Since each feature subspace contained different information of the speech recordings, it was reasonable to integrate suitable speech features. It was noted that utilizing SPEC and PROS features offered significantly better classification outcomes for males than utilizing MFCC and GLOT features ($p < 0.05$). It was discovered that utilizing GLOT features offered significantly worse classification outcomes for females than utilizing SPEC, PROS, and MFCC features ($p < 0.05$). It was also discovered that LR performed minutely better than SVM and GMM, which was a reason for LR being selected as the base classifier.

Second, we revealed an ensemble methodology for the classification of depression, ELRDD. It was noted that ELRDD, which was developed from multiple feature subsets, outperformed both the individual classifiers and the other ensemble classifiers including SVM, GMM, LR, adaboost decision tree, bagging decision tree, and random forest. ELRDD revealed an accuracy level of 70.64% for males and 66.68% for females, as well as a sensitivity/specificity ratio of 67.35%/73.94% for males and 65.71%/67.68% for females.

Finally, based on ELRDD, we proposed ELRDD-E, which utilized the classification results of all 29 speech recordings of each subject in our dataset. This methodology offered extremely encouraging outcomes, revealing an increased accuracy level of 81.82% for males and 75.00% for females, as well as an advantageous sensitivity/specificity ratio of 78.13%/85.29% for males and 79.25%/70.59% for females.

While the experimental outcomes are promising, a possible limitation of this research is that speech may have additional features that pertain to depression. A future direction of this study is to investigate improvements in feature extraction and selection strategy.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Basic Research Program of China (973 Program) (no. 2014CB744600).

References

[1] World Health Organization, *Depression Fact Sheet*, WHO, Geneva, Switzerland, 2018, http://www.who.int/en/news-room/fact-sheets/detail/depression.

[2] R. C. Kessler, P. Berglund, O. Demler et al., "The epidemiology of major depressive disorder: results from the national comorbidity survey replication (NCS-R)," *JAMA*, vol. 289, no. 23, pp. 3095–3105, 2003.

[3] K. Hawton, C. Casanas i Comabella, C. Haw, and K. Saunders, "Risk factors for suicide in individuals with depression: a systematic review," *Journal of Affective Disorders*, vol. 147, no. 1–3, pp. 17–28, 2013.

[4] J. C. Mundt, P. J. Snyder, M. S. Cannizzaro, K. Chappie, and D. S. Geralts, "Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology," *Journal of Neurolinguistics*, vol. 20, no. 1, pp. 50–64, 2007.

[5] C. Sobin and H. A. Sackeim, "Psychomotor symptoms of depression," *American Journal of Psychiatry*, vol. 154, pp. 4–17, 1997.

[6] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, "A review of depression and suicide risk assessment using speech analysis," *Speech Communication*, vol. 71, pp. 10–49, 2015.

[7] J. K. Darby and H. Hollien, "Vocal and speech patterns of depressive patients," *Folia Phoniatrica et Logopaedica*, vol. 29, pp. 279–291, 1977.

[8] H. H. Stassen, S. Kuny, and D. Hell, "The speech analysis approach to determining onset of improvement under antidepressants," *European Neuropsychopharmacology*, vol. 8, no. 4, pp. 303–310, 1998.

[9] F. Hönig, A. Batliner, E. Nöth, S. Schnieder, and J. Krajewski, "Automatic modelling of depressed speech: relevant features and relevance of gender," in *Proceedings of Fifteenth Annual Conference of the International Speech Communication Association*, pp. 1248–1252, ISCA, Singapore, September 2014.

[10] M. Alpert, E. R. Pouget, and R. R. Silva, "Reflections of depression in acoustic measures of the patient's speech," *Journal of Affective Disorders*, vol. 66, no. 1, pp. 59–69, 2001.

[11] M. Cannizzaro, B. Harel, N. Reilly, P. Chappell, and P. J. Snyder, "Voice acoustical measurement of the severity of major depression," *Brain and Cognition*, vol. 56, no. 1, pp. 30–35, 2004.

[12] Y. Yang, C. Fairbairn, and J. Cohn, "Detecting depression severity from vocal prosody," *IEEE Transactions on Affective Computing*, vol. 4, pp. 142–150, 2012.

[13] L. A. Low, N. C. Maddage, M. Lech, L. B. Sheeber, and N. B. Allen, "Detection of clinical depression in adolescents speech during family interactions," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 574–586, 2011.

[14] E. Moore, M. Clements, J. W. Peifer, and L. Weisser, "Critical analysis of the impact of glottal features in the classification of clinical depression in speech," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 96–107, 2008.

[15] K. E. B. Ooi, M. Lech, and N. B. Allen, "Multichannel weighted speech classification system for prediction of major depression in adolescents," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 2, pp. 497–506, 2013.

[16] K. E. B. Ooi, M. Lech, and N. B. Allen, "Prediction of major depression in adolescents using an optimized multi-channel weighted speech classification system," *Biomedical Signal Processing and Control*, vol. 14, pp. 228–239, 2014.

[17] L. A. Low, N. C. Maddage, M. Lech, L. B. Sheeber, and N. B. Allen, "Influence of acoustic low-level descriptors in the detection of clinical depression in adolescents," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5154–5157, IEEE, Dallas, TX, USA, March 2010.

[18] M. Valstar, B. Schuller, K. Smith et al., "3D dimensional affect and depression recognition challenge," in *Proceedings of 4th*

ACM International Workshop on Audio/Visual Emotion Challenge (AVEC'14), pp. 3–10, ACM, Orlando, FL, USA, November 2014.

[19] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Breakspear, and G. Parker, "Detecting depression: a comparison between spontaneous and read speech," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7547–7551, IEEE, Vancouver, Canada, 2013.

[20] H. H. Jiang, B. Hu, Z. Y. Liu et al., "Investigation of different speech types and emotions for detecting depression using different classifiers," *Speech Communication*, vol. 90, pp. 39–46, 2017.

[21] N. Cummins, J. Epps, M. Breakspear, and R. Goecke, "An investigation of depressed speech detection: features and normalization," in *Proceedings of Interspeech*, pp. 2997–3000, ISCA, Florence, Italy, August 2011.

[22] N. Cummins, J. Epps, V. Sethu, and J. Krajewski, "Variability compensation in small data: oversampled extraction of i-vectors for the classification of depressed speech," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2014*, pp. 970–974, IEEE, Florence, Italy, May 2014.

[23] D. Sturim, P. A. Torres-Carrasquillo, T. F. Quatieri, N. Malyska, and A. McCree, "Automatic detection of depression in speech using Gaussian mixture modeling with factor analysis," in *Proceedings of Interspeech*, pp. 2983–2986, ISCA, Florence, Italy, August 2011.

[24] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Breakspear, and G. Parker, "From joyous to clinically depressed: mood detection using spontaneous speech," in *Proceedings of FLAIRS Conference*, pp. 141–146, AAAI Press, Marco Island, FL, USA, May 2012.

[25] J. Joshi, R. Goecke, S. Alghowinem et al., "Multimodal assistive technologies for depression diagnosis and monitoring," *Journal on Multimodal User Interfaces*, vol. 7, no. 3, pp. 217–228, 2013.

[26] S. Scherer, G. Stratou, J. Gratch, and L. Morency, "Investigating voice quality as a speaker-independent indicator of depression and PTSD," in *Proceedings of Interspeech*, pp. 847–851, ISCA, Lyon, France, August 2013.

[27] S. Scherer, G. Stratou, M. Mahmoud, J. Boberg, and J. Gratch, "Automatic behavior descriptors for psychological disorder analysis," in *Proceedings of 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp. 1–8, IEEE, Shanghai, China, April 2013.

[28] S. Scherer, G. Stratou, and L. P. Morency, "Audiovisual behavior descriptors for depression assessment," in *Proceedings of 15th ACM on International Conference on Multimodal Interaction (ICMI)*, pp. 135–140, ACM, New York, NY, USA, 2013.

[29] T. F. Quatieri and N. Malyska, "Vocal-source biomarkers for depression: a link to psychomotor activity," in *Proceedings of Interspeech*, pp. 1059–1062, ICSA, Portland, OR, USA, September 2012.

[30] A. Ozdas, R. G. Shiavi, S. E. Silverman, M. K. Silverman, and D. M. Wilkes, "Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 9, pp. 1530–1540, 2004.

[31] J. F. Cohn, T. S. Kruez, I. Matthews et al., "Detecting depression from facial actions and vocal prosody," in *Proceedings of 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp. 1–7, IEEE, Amsterdam, Netherlands, September 2009.

[32] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, T. Gedeon, and M. Breakspear, "A comparative study of different classifiers for detecting depression from spontaneous speech," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2013*, pp. 8022–8026, IEEE, Vancouver, Canada, May 2013.

[33] B. S. Helfer, T. F. Quatieri, J. R. Williamson, D. D. Mehta, R. Horwitz, and B. Yu, "Classification of depression state based on articulatory precision," in *Proceedings of Interspeech*, pp. 2172–2176, ISCA, Lyon, France, August 2013.

[34] A. Sebti and H. Hassanpour, "Body orientation estimation with the ensemble of logistic regression classifiers," *Multimedia Tools and Applications*, vol. 76, no. 22, pp. 23589–23605, 2017.

[35] H. Wang, Q. S. Xu, and L. F. Zhou, "Large unbalanced credit scoring using lasso-logistic regression ensemble," *Plos One*, vol. 10, no. 2, Article ID e0117844, 2015.

[36] P. D. Prasad, H. N. Halahalli, J. P. John, and K. K. Majumdar, "Single-trial EEG classification using logistic regression based on ensemble synchronization," *IEEE J. Biomed. Health*, vol. 18, no. 3, pp. 1074–1080, 2014.

[37] N. Lim, H. Ahn, H. Moon, and J. J. Chen, "Classification of high-dimensional data with ensemble of logistic regression models," *Journal of Biopharmaceutical Statistics*, vol. 20, no. 1, pp. 160–171, 2010.

[38] H. Kuswanto, A. Asfihani, Y. Sarumaha, and H. Ohwada, "Logistic regression ensemble for predicting customer defection with very large sample size," in *Proceedings of Third Information Systems International Conference*, pp. 86–93, Elsevier, Amsterdam, Netherlands, December 2015.

[39] K. Lee, H. Ahn, H. Moon, R. L. Kodell, and J. J. Chen, "Multinomial logistic regression ensembles," *Journal of Biopharmaceutical Statistics*, vol. 23, no. 3, pp. 681–694, 2013.

[40] Z. Y. Liu, B. Hu, L. H. Yan et al., "Detection of depression in speech," in *Proceedings of International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 743–747, IEEE, Xi'an, China, September 2015.

[41] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Association, Washington, DC, USA, 4th edition, 1994.

[42] K. Kroencke, R. Spitzer, and J. Williams, "The phq-9: validity of a brief depression severity measure," *Journal of General Internal Medicine*, vol. 16, pp. 606–613, 2001.

[43] D. J. France, R. G. Shiavi, S. Silverman, M. Silverman, and M. Wilkes, "Acoustical properties of speech as indicators of depression and suicidal risk," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 829–837, 2000.

[44] S. Nolenhoeksema and J. S. Girgus, "The emergence of gender differences in depression during adolescence," *Psychological Bulletin*, vol. 115, no. 3, pp. 424–443, 1994.

[45] M. Airas, "TKK Aparat: an environment for voice inverse filtering and parameterization," *Logopedics Phoniatrics Vocology*, vol. 33, no. 1, pp. 49–64, 2008.

[46] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile-the Munich versatile and fast open-source audio feature extractor," in *Proceedings of 18th ACM International Conference on Multimedia*, pp. 1459–1462, ACM, Firenze, Italy, October 2010.

[47] J. Chen and Y. Liu, "Locally linear embedding: a survey," *Artificial Intelligence Review*, vol. 36, no. 1, pp. 29–48, 2011.

[48] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.