

Research Article

An Emotion Detection System Based on Multi Least Squares Twin Support Vector Machine

Divya Tomar, Divya Ojha, and Sonali Agarwal

Indian Institute of Information Technology, Allahabad, Uttar Pradesh 211012, India

Correspondence should be addressed to Divya Tomar; divyatomar26@gmail.com

Received 17 July 2014; Accepted 30 November 2014; Published 23 December 2014

Academic Editor: Ujjwal Bhattacharya

Copyright © 2014 Divya Tomar 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.

Posttraumatic stress disorder (PTSD), bipolar manic disorder (BMD), obsessive compulsive disorder (OCD), depression, and suicide are some major problems existing in civilian and military life. The change in emotion is responsible for such type of diseases. So, it is essential to develop a robust and reliable emotion detection system which is suitable for real world applications. Apart from healthcare, importance of automatically recognizing emotions from human speech has grown with the increasing role of spoken language interfaces in human-computer interaction applications. Detection of emotion in speech can be applied in a variety of situations to allocate limited human resources to clients with the highest levels of distress or need, such as in automated call centers or in a nursing home. In this paper, we used a novel multi least squares twin support vector machine classifier in order to detect seven different emotions such as anger, happiness, sadness, anxiety, disgust, panic, and neutral emotions. The experimental result indicates better performance of the proposed technique over other existing approaches. The result suggests that the proposed emotion detection system may be used for screening of mental status.

1. Introduction

Stressful situation can cause some major psychiatric problems such as depression, suicide, PTSD, BMD, and OCD in civilian as well as in military life. Earlier treatment may become useful for such type of psychiatric problems [1]. So, there is a need to develop technology for recognizing early change in human behavior. Several biomarkers are reported by the medical researchers for psychiatric diseases [1, 2]. But these biomarkers are not effective in military life as they required a big and complicated machine for detecting psychiatric diseases. On the other hand, there is a fast development in voice, speech, and emotion detection technologies in engineering field. These technologies provide human-machine interaction for emotion detection and further treatment of psychiatric problems [3–5]. Several researches measured the level of fatigue and stress from speech [6]. But the level of fatigue and stress does not lead to psychiatric disorder directly. Emotion change of a human can cause mental diseases. Mostly, clinicians recognize the mental state of a patient from his/her face and voice which represents his/her emotion. This fact leads to the possibility that emotion detection system can be used for

recognizing the mental disorder or disease in human. Early detection of disease improves the prognosis and is helpful to provide effective treatment at early stages. Emotion detection system can provide support to the clinicians to perform the task of emotion detection more efficiently. In automated call centers or in a nursing home, while nursing staff may not be available to assist everyone, automated emotion detection can be used to “triage” a patient. Automated emotion detection system is helpful to recognize whether a patient becomes angry or impatient and if so then the staff or treatment is provided to that patient as soon as possible.

Nowadays emotion detection from speech is an active research area and is useful for man-machine interaction [3–7]. Various researches have been done about automated emotion detection from facial expressions. But this task is computationally expensive and complex due to the requirement of high quality cameras for capturing face images. Apart from facial expression, emotions are also detected from speech which has been proven to be more promising modality. Since speech is the primary mode of human communication, the detection of emotion from speech is an important aspect.

Machine learning algorithms such as K -nearest neighbor (KNN), artificial neural network (ANN), and support vector machine (SVM) are widely used for emotion detection due to their excellent performance [8–13]. In this paper, the proposed emotion detection system recognizes seven different emotions which are anger, anxiety, disgust, happiness, sadness, panic, and neutral emotions. Different emotions can be seen as different classes. So, it requires a multiclassifier for emotion detection. In this paper, we proposed a novel multi least squares twin support vector machine (MLSTSVM) classifier which is the extension of binary least squares twin support vector machine (LSTSVM). So, the proposed system predicts the class or emotion for a given input. In order to check the validity of the proposed classifier, we evaluated its performance against 5 benchmark datasets.

The paper is organized as follows: introduction section includes need for emotion detection system. Section 2 provides the detail of our novel classifier which is multi least squares twin support vector machine. Proposed framework for emotion detection and dataset details are discussed in Section 3. The experimental results and conclusion of the proposed emotion detection system are presented in Sections 4 and 5, respectively.

2. Multi Least Squares Twin Support Vector Machine

Kumar and Gopal proposed LSTSVM for binary classification which solves two linear programming problems and constructs two nonparallel hyperplanes, one for each class [14]. Since real world data contains multiple classes and requires a classifier that works well for multiple classes, in this paper, we propose a novel multiclassifier termed as MLSTSVM. This classifier is an extension of the binary LSTSVM and is based on “one-versus-rest” strategy. Here, we selected and extended the binary LSTSVM because it shows better generalization ability and is faster as compared to other existing approaches [14, 15]. MLSTSVM constructs “ K ” hyperplanes, one for each class, by optimizing K -linear programming problems, where “ K ” denotes number of classes. It adopts the concept of “one-versus-rest” in which the data points of each class are trained with the data points of other classes. Consider dataset has “ n ” number of data points in training dataset: $\{x_1, y_1\}, \dots, \{x_n, y_n\}$. Here x_i is a feature vector in m -dimensional space R and $y_i \in \{1, 2, \dots, K\}$ is the label of corresponding class. “One-versus-rest” generates K binary LSTSVM classifier, each of which separates one class from the rest of the classes. The k th LSTSVM classifier assumes the data points of k th class as positive data points and the data points of other classes as negative data points. Consider the data points of k th class are indicated by the matrix $A_k \in R^{n_k \times m}$, where n_k represents number of data points in k th class. Let the data points of rest of the classes be indicated by matrix $B_k \in R^{(n-n_k) \times m}$ as

$$B_k = [(A_1)^T, (A_2)^T, \dots, (A_{k-1})^T, (A_{k+1})^T, \dots, (A_K)^T]^{-T}. \quad (1)$$

The matrix B_k includes all the data points except k th class. MLSTSVM classifier for both linear and nonlinear cases is formulated as follows.

2.1. Linear Case. The equation of k th hyperplane is obtained as

$$(w_k \cdot x) + b_k = 0, \quad (2)$$

where $w_k \in R^m$ and $b_k \in R$ represent normal vector and bias term, respectively, in real space R . The k th LSTSVM classifier optimizes the following objective function:

$$\min (w_k, b_k, \xi_k) \quad \frac{1}{2} \|A_k w_k + e_{k1} b_k\|^2 + \frac{c_k}{2} \xi_k^T \xi_k \quad (3)$$

$$\text{s.t.} \quad (B_k w_k + e_{k2} b_k) + \xi_k = e_{k2},$$

where $e_{k1} \in R^{n_k}$ and $e_{k2} \in R^{(n-n_k)}$ denote the vector of 1's and c_k and ξ_k represent the penalty parameter and the slack variable correspondingly. The first term of (3) denotes the squared sum distance of the data points of the k th class. The minimization of this term keeps the hyperplane in the close affinity of the k th class. The second term of (3) minimizes the misclassification error of the data points of rest of the $K-1$ classes. So, in this way the hyperplane is kept in the close affinity with the data points of k th class and lies as far as possible from the data points of other classes. The objective function is solved by taking its dual form. Lagrangian function of the objective function as mentioned by (3) is achieved as follows:

$$L(w_k, b_k, \xi_k, \alpha_k) = \frac{1}{2} \|A_k w_k + e_{k1} b_k\|^2 + \frac{c_k}{2} \xi_k^T \xi_k - \alpha_k^T ((B_k w_k + e_{k2} b_k) + \xi_k - e_{k2}), \quad (4)$$

where $\alpha_k > 0$ represents the Lagrangian multiplier. The optimization of Lagrangian function is achieved by differentiating it with respect to normal vector, bias, slack variable, and Lagrangian multiplier and the following Karush-Kuhn-Tucker (KKT) conditions are obtained:

$$\frac{\partial L}{\partial w_k} = A_k^T (A_k w_k + e_{k1} b_k) - Y_k^T \alpha_k = 0, \quad (5)$$

$$\frac{\partial L}{\partial b_k} = e_{k1}^T (A_k w_k + e_{k1} b_k) - e_{k2}^T \alpha_k = 0, \quad (6)$$

$$\frac{\partial L}{\partial \xi_k} = c_k \xi_k - \alpha_k = 0, \quad (7)$$

$$\frac{\partial L}{\partial \alpha_k} = (B_k w_k + e_{k2} b_k) + \xi_k - e_{k2} = 0.$$

By combining (5) and (6), the following equation is obtained:

$$\begin{bmatrix} A_k^T \\ e_{k1}^T \end{bmatrix} [A_k \ e_{k1}] \begin{bmatrix} w_k \\ b_k \end{bmatrix} - \begin{bmatrix} B_k^T \\ e_{k2}^T \end{bmatrix} \alpha_k = 0. \quad (8)$$

Consider $G_k = [A_k \ e_{k1}]$ and $H_k = [B_k \ e_{k2}]$ and $u_k = \begin{bmatrix} w_k \\ b_k \end{bmatrix}$. After putting these values in (8), it may be reformulated as

$$G_k^T G_k u_k - H_k^T \alpha_k = 0, \quad (9)$$

$$u_k = (G_k^T G_k)^{-1} H_k^T \alpha_k. \quad (10)$$

The above equation requires the inverse of $G_k^T G_k$. Sometimes a matrix may be singular or ill-conditioned due to which it is not possible to obtain its inverse. The situation may be avoided by adding a regularization term δI to the matrix and the above equation is reformulated as

$$u_k = (G_k^T G_k + \delta I)^{-1} H_k^T \alpha_k, \quad (11)$$

where δ is a very small nonnegative integer and I is an identity matrix of suitable size. Lagrangian multiplier is obtained from (7) as

$$\alpha_k = c_k (e_{k2} - H_k u_k). \quad (12)$$

After substituting the value of α_k in (10), we obtain the normal vector and bias for k th classifier as follows:

$$\begin{bmatrix} w_k \\ b_k \end{bmatrix} = u_k = \left(H_k^T H_k + \frac{1}{c_k} G_k^T G_k \right)^{-1} H_k^T e_{k2}, \quad (13)$$

$$f(x) = \arg \min_{k=1, \dots, K} \frac{|w_k \cdot x + b_k|}{\|w_k\|}. \quad (14)$$

For a new data point or test data sample, its perpendicular distance is measured from each hyperplane and the data sample is assigned to that class depending upon which of the planes lies at minimum distance from it.

Algorithm 1. (1) For $j = 1$ to K , where K is total number of classes,

(i) obtain two matrices G_j and H_j as

$$G_j = [A_j \ e_{j1}], \quad H_j = [B_j \ e_{j2}], \quad (15)$$

where A_j and B_j denote the data points of j th class and the rest of the classes, respectively;

(ii) use validation process to obtain penalty parameters;

(iii) calculate weight and bias for each class by using (13).

(2) Achieve decision function by using (14). Use this function to assign the class to new data points.

2.2. Nonlinear Case. Mostly, the real dataset is nonlinear in nature; that is, the classes are separable by nonlinear class boundaries. So, it is essential for a classifier that it works well both for linear and for nonlinear separable data points. In this section, we proposed the formulation of the MLSTSVM classifier for nonlinear cases. Firstly, kernel functions are used for mapping the input data points into higher-dimensional feature space and then the data points are classified by constructing nonlinear or kernel surfaces in this space. In higher-dimensional space, the equation of k th kernel surface or nonlinear surface for any kernel function is obtained as

$$\text{Ker}(x, Z^T) \mu_k + \gamma_k = 0, \quad \text{where } k = 1, \dots, K, \quad (16)$$

where $Z = [A_k \ B_k]^T$ and Ker is any suitable kernel function. The optimization problem of MLSTSVM for nonlinear cases is formulated as

$$\begin{aligned} \min (\mu_k, \gamma_k, \xi_k) \quad & \frac{1}{2} \|\text{Ker}(A_k, Z^T) \mu_k + e_{k1} \gamma_k\|^2 + \frac{c_k}{2} \xi_k^T \xi_k \\ \text{s.t.} \quad & (\text{Ker}(B_k, Z^T) \mu_k + e_{k2} \gamma_k) + \xi_k = e_{k2}. \end{aligned} \quad (17)$$

Lagrangian function of the above-mentioned equation is achieved as

$$\begin{aligned} L(\mu_k, \gamma_k, \xi_k, \alpha_k) \\ = \frac{1}{2} \|\text{Ker}(A_k, Z^T) \mu_k + e_{k1} \gamma_k\|^2 + \frac{c_k}{2} \xi_k^T \xi_k \\ - \alpha_k^T ((\text{Ker}(B_k, Z^T) \mu_k + e_{k2} \gamma_k) + \xi_k - e_{k2}). \end{aligned} \quad (18)$$

KKT conditions for nonlinear MLSTSVM are

$$\frac{\partial L}{\partial \mu_k} = \text{Ker}(A_k, Z^T)^T (\text{Ker}(A_k, Z^T) \mu_k + e_{k1} \gamma_k) - \text{Ker}(B_k, Z^T)^T \alpha_k = 0, \quad (19)$$

$$\frac{\partial L}{\partial \gamma_k} = e_{k1}^T (\text{Ker}(A_k, Z^T) \mu_k + e_{k1} \gamma_k) - e_{k2}^T \alpha_k = 0, \quad (20)$$

$$\frac{\partial L}{\partial \xi_k} = c_k \xi_k - \alpha_k = 0, \quad (21)$$

$$\frac{\partial L}{\partial \alpha_k} = (\text{Ker}(B_k, Z^T) \mu_k + e_{k2} \gamma_k) + \xi_k - e_{k2} = 0.$$

Combining (19) and (20), we get

$$\begin{aligned} \begin{bmatrix} \text{Ker}(A_k, Z^T)^T \\ e_{k1}^T \end{bmatrix} [\text{Ker}(A_k, Z^T) \ e_{k1}] \begin{bmatrix} \mu_k \\ \gamma_k \end{bmatrix} \\ - \begin{bmatrix} \text{Ker}(B_k, Z^T)^T \\ e_{k2}^T \end{bmatrix} \alpha_k = 0. \end{aligned} \quad (22)$$

Let $S_k = [\text{Ker}(A_k, Z^T) \ e_{k1}]$ and $T_k = [\text{Ker}(B_k, Z^T) \ e_{k2}]$. Then (22) can be rewritten as

$$S_k^T S_k \begin{bmatrix} \mu_k \\ \gamma_k \end{bmatrix} - T_k^T \alpha_k = 0. \quad (23)$$

The value of normal vector and bias is achieved by solving (21) and (22) as

$$\begin{bmatrix} \mu_k \\ \gamma_k \end{bmatrix} = \left(T_k^T T_k + \frac{1}{c_k} S_k^T S_k \right)^{-1} T_k^T e_{k2}. \quad (24)$$

For a new data point, its perpendicular distance is measured from each nonlinear surface and it is assigned to that class depending upon which of the planes lies at minimum distance from it. The values of weight and bias are used to construct kernel surfaces for each class. The decision function for nonlinear MLSTSVM is obtained as

$$f(x) = \arg \min_{k=1, \dots, K} \frac{|\mu_k \cdot \text{Ker}(x, Z^T) + \gamma_k|}{\|\mu_k\|}. \quad (25)$$

Gaussian and polynomial kernel functions for two input vectors x_i and x_j are obtained as

$$K_G(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad (26)$$

$$K_P(x_i, x_j) = (1 + x_i^T x_j)^d.$$

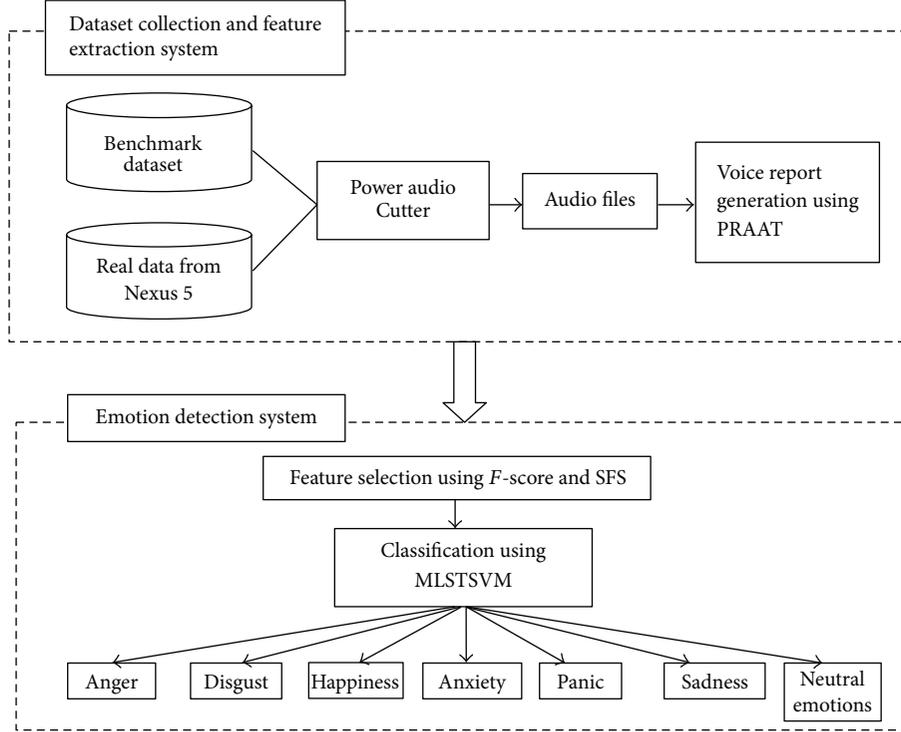


FIGURE 1: Block diagram of emotion detection system.

Algorithm 2. (1) Choose kernel function.

(2) For $j = 1$ to K , where K is total number of classes,

(i) obtain two matrices S_j and T_j as

$$\begin{aligned} S_j &= [\text{Ker}(A_j, Z^T) \ e_{j1}], \\ T_j &= [\text{Ker}(B_j, Z^T) \ e_{j2}], \end{aligned} \quad (27)$$

where A_j and B_j denote the data points of j th class and the rest of the classes correspondingly;

(ii) use validation process to obtain penalty parameters;

(iii) calculate weight and bias for each class by using (24).

(3) Obtain decision function by using (25) and assign the class to new data points by using this decision function.

In order to prove the validity of the proposed MLSTSVM, we performed experiment on five benchmark datasets. All the datasets are taken from UCI machine learning database [16]. Table 1 shows the accuracy comparison of the proposed MLSTSVM classifier with other exiting classifiers. Accuracy refers to the correct classification rate and is calculated by taking the average of testing accuracies. It is clear from the table that the proposed classifier has achieved better accuracy for Wine, Glass, Vehicle, and Teaching Evaluation datasets as compared to KNN, ANN, and multi-SVM, while for Iris dataset MLSTSVM obtained 97.75% accuracy which is better than ANN and multi-SVM and comparable with KNN.

TABLE 1: Accuracy comparison.

Dataset	KNN	ANN	Multi-SVM	MLSTSVM
Wine ($178 \times 13 \times 3$)	95.50%	98.70%	99.43%	99.54%
Glass ($214 \times 13 \times 6$)	72.00%	69.88%	66.35%	78.23%
Iris ($150 \times 4 \times 3$)	98.00%	97.30%	97.33%	97.75%
Vehicle ($846 \times 18 \times 6$)	82.23%	83.45%	85.02%	88.61%
Teaching Evaluation ($151 \times 5 \times 3$)	66.67%	68.07%	69.11%	72.37%

The experiment is performed using 10-fold cross validation method.

3. Description of Dataset and Proposed Model

3.1. Dataset Description. Emotion detection system from human speech is divided into two parts: dataset collection and feature extraction system and emotion detection system as shown in Figure 1. Emotion dataset, used in this research work, is taken from two sources: a benchmark dataset from source <http://www.eecs.qmul.ac.uk/mmv/datasets/deap/download.html> and real dataset (audio file) collected using Nexus 5 smart phone. The benchmark dataset contains the voice recordings of expert reciting numbers and dates with different emotions and real dataset contains the audio recording of 10 male persons where each person voice is recorded minimum 2-3 times with different sets of emotions.

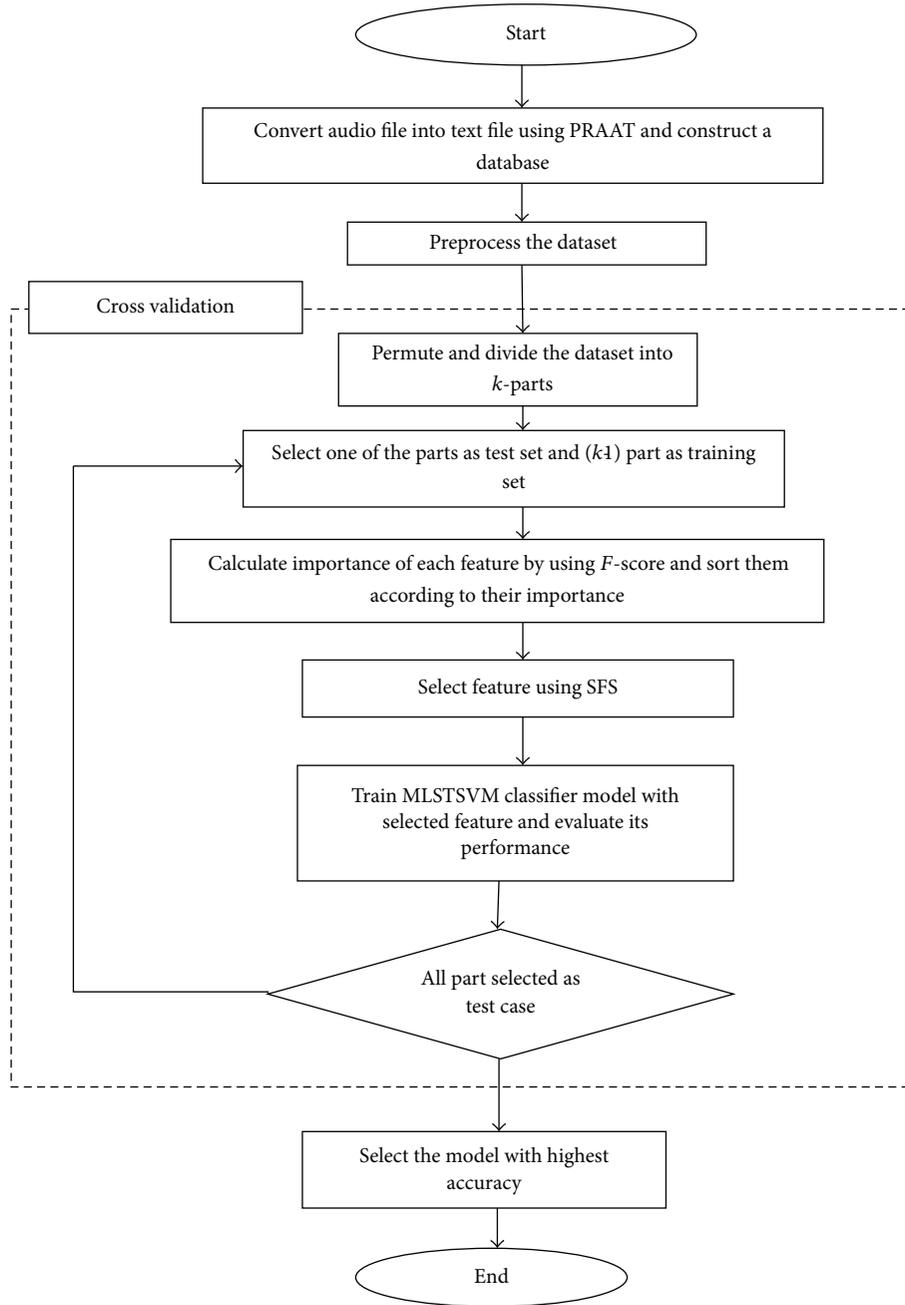


FIGURE 2: Proposed framework.

The dataset includes seven different emotions such as anger, anxiety, happiness, sadness, disgust, neutral emotions, and panic for emotion detection. Audio file of a person or a patient is given for emotion detection. Power Audio Cutter is used to cut the audio files and the duration of each audio file is 2 seconds. This research used PRAAT scripting tool for feature extraction which is a freeware and flexible tool developed by Paul Boersma and David Weenink of the University of Amsterdam for speech analysis. It performs spectral, formant, pitch, intensity, jitter, and shimmer analyses [17, 18].

This tool generates voice report from audio files and converts the audio files into text files. The voice report of these audio files contains different features of voice like pitch, intensity, shimmer, jitter, and so forth. Pitch, also known as vibration rate of the vocal folds, is one of the most important and essential parts of the human voice. The sound of the voice varies according to the vibration rate. High pitch refers to high vibration rate which further increases the sound of the voice while low pitch corresponds to the lower sound. Vibration rate is dependent on the duration and thickness of vocal cords. Relaxation and tightening of

the muscles around vocal cords also affect the vibration rate. Emotion or mood of a person also has an effect on his/her pitch. During excitement or fright, the muscles put strain on vocal cords which further produce high pitch voice. The tone of a person describes the way a statement is presented and can convey the emotion, psychological arousal, and mood of that person. Usually, softer pitch and tone are seen as nonaggressive and indicate the friendly behavior of a person. Jitter and shimmer are another important attribute of a voice. Jitter and shimmer measure the irregularity percentage in the pitch and in the amplitude of the vocal note correspondingly. Voice quality and signal-to-noise ratio can be estimated from harmonicity. The second part includes significant feature selection from voice report generated by PRAAT and classification of emotions using MLSTSVM. Feature selection (FS) is the process of selecting relevant and important attributes from a dataset and plays a significant role in the construction of a classification system [19–21]. FS is also termed as attribute selection process which reduces the number of input attributes by selecting only important attributes for a classifier in order to enhance its performance. In this paper, we used the combination of *F*-score and sequential forward selection approaches for feature selection.

F-Score. It is a significant FS approach and mostly used in machine learning. Initially, *F*-score calculates the discrimination between two class sets of real numbers. Later, it is extended efficiently for calculating the discrimination between more than two class sets of real numbers [19–21]. Let the dataset contain “*K*” classes and each class contain n_k data points, where $k = 1, 2, \dots, K$. The formulation of *F*-score for *j*th feature is obtained as [21, 22]

$$F_j = \frac{\sum_{k=1}^K (\bar{X}_j^k - \bar{X}_j)^2}{\sum_{k=1}^K (1/(n_k - 1)) \sum_{i=1}^{n_k} (X_{i,j}^k - \bar{X}_j^k)^2}, \quad (28)$$

where \bar{X}_j is average of the *j*th feature of the whole dataset, \bar{X}_j^k is average of the *j*th feature of the *k*th class, and $X_{i,j}^k$ is *j*th feature of the *i*th instance in the *k*th class. The value of *F*-score for a feature represents its discriminative ability and larger value of it indicates that the corresponding feature has better separation ability.

Sequential Forward Selection. It works in bottom to top fashion and begins with an empty set of features. It starts with adding single best feature to the empty set. At each subsequent step, the best one feature of the remaining original features is added to the previous feature set.

The result obtained from feature selection is given to the proposed classifier and usability of feature selection is compared while observing the behavior of the same classifier without feature selection. The detailed description of MLSTSVM is explained in Section 2.

3.2. Proposed Framework. Figure 2 shows the proposed framework for emotion detection from audio files by using

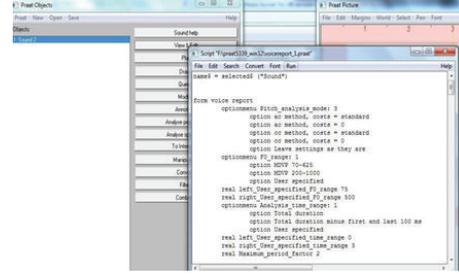


FIGURE 3: Selecting audio file and running PRAAT script for producing features.

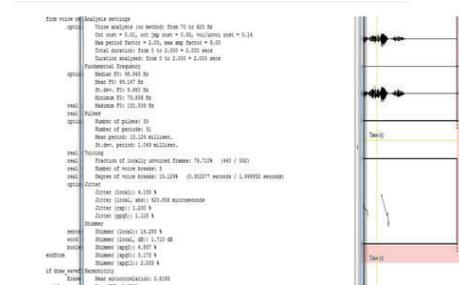


FIGURE 4: Voice report of an audio file.

a novel MLSTSVM classifier. The proposed system contains the following steps.

Step 1. Record the voice in audio format of different emotions.

Step 2. Convert the audio files into text files using PRAAT scripting tool.

Step 3. Generate a database containing different emotions.

Step 4. Preprocess the dataset.

Step 5. Partition the dataset into training and testing datasets by using *k*-fold cross validation.

Step 6. Select the relevant features with the help of *F*-score and sequential forward selection approaches.

Step 7. Train and test the model with selected features and evaluate its performance with different features.

Step 8. Select the model with highest classification accuracy.

4. Experimental Results and Discussion

The benchmark and real data exist in the form of audio files. The experiment is performed on the time span of 2 seconds for each audio file for which a tool Power Audio Cutter is used to cut the audio files as per required duration. The feature of audio files is extracted by using PRAAT scripting tool. Figure 3 shows the browsing of audio file for running PRAAT script in order to extract features from it. The voice report of an audio file generated by using PRAAT is shown in Figure 4.

TABLE 2: Extracted speech features after using PRAAT.

Attributes	Subattributes	Description
Fundamental frequency	MDVP: Fo (Hz)	Median vocal fundamental frequency
	MDVP: Fhi (Hz)	Maximum vocal fundamental frequency
	MDVP: Flo (Hz)	Minimum vocal fundamental frequency
	St.dev.: F0 (Hz)	Standard deviation vocal fundamental frequency
	Mean: Fm0 (Hz)	Mean vocal fundamental frequency
	Pulses	Npul
Nper		Number of periods
Mper (ms)		Mean period
St.dev. per (ms)		Standard deviation period
Voicing	Fofl (%)	Fraction of locally unvoiced frames
	NofV	Number of voice breaks
	DofV	Degree of voice breaks
Jitter	JL (%)	Jitter (local)
	JLA (ms)	Jitter (local, abs)
	JPP (%)	Jitter (ppq5)
	JR (%)	Jitter (rap)
Shimmer	SL (%)	Shimmer (local)
	SLD (dB)	Shimmer (local, dB)
	SA3 (%)	Shimmer (apq3)
	SA5 (%)	Shimmer (apq5)
	SA11 (%)	Shimmer (apq11)
Harmonicity	MA	Mean autocorrelation
	MNHR	Mean NHR
	MHNR (dB)	Mean HNR

TABLE 3: Dataset details.

Emotions	Number of instances	Number of attributes
Anger	54	24
Anxiety	30	24
Happiness	60	24
Sadness	46	24
Disgust	36	24
Neutral emotions	32	24
Panic	32	24

TABLE 4: Average value of F -score using 10-fold cross validation.

Features	F -score	Features	F -score
F1	3.6977	F13	0.0090
F2	1.6892	F14	0.1872
F3	0.0966	F15	0.0074
F4	0.2402	F16	0.0088
F5	0.1119	F17	0.0171
F6	0.1250	F18	0.0070
F7	0.1143	F19	0.0108
F8	0.0187	F20	0.0163
F9	0.0095	F21	0.0012
F10	0.0588	F22	0.0039
F11	0.0106	F23	0.0168
F12	0.0405	F24	0.0010

Figure 5 shows the snapshot of the emotion detection dataset. In this snapshot 35 instances, 5 instances of each emotion, have been taken to generate a complete view of range of various attributes for corresponding class. The first attribute of the snapshot denotes emotions. Here, “1,” “2,” “3,” “4,” “5,” “6,” and “7” are used for anger, anxiety, disgust, happiness, neutral emotions, panic, and sadness, respectively.

Since the range of attributes varies from each other, normalization of each attribute value is performed to take them within the specified range. Two feature selection techniques, F -score and SFS, are used for selecting significant features. Table 4 shows the average value of F -score for each attribute or feature by using 10-fold cross validation. After calculating the F -score of each attribute, SFS is used for obtaining 24 feature subsets or models. The importance or F -score values of each feature from high to low are 1, 2, 4, 14, 6, 7, 5, 3, 10, 12, 8, 17, 23, 20, 19, 11, 9, 13, 16, 15, 18, 22, 21, and 24. Table 5 shows the twenty-four feature subsets or models on the basis of SFS. For each feature subset, a MLSTSVM classifier is constructed and its predictive accuracy is checked using 10-fold cross validation method. The proposed MLSTSVM classifier is implemented using MATLAB R2012a.

Appropriate choice of parameters such as penalty parameters c_1 and c_2 , σ , and degree can enhance the performance of the MLSTSVM. This study used grid search approach for this purpose. The parameters are selected from the following range: $c_1, c_2 \in \{2^{-10}, \dots, 2^{10}\}$, $\sigma \in \{2^{-5}, \dots, 2^{10}\}$, and degree $\in \{2^{-2}, \dots, 2^5\}$ by using 10-fold cross validation.

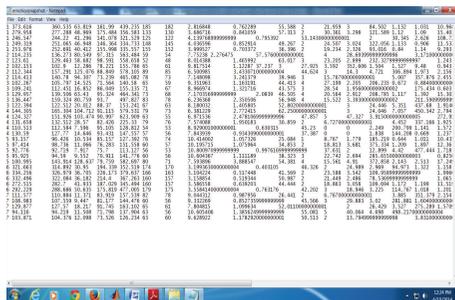


FIGURE 5: Snapshot of the dataset.

Table 2 shows the extracted speech features by using PRAAT. PRAAT extracted six major attributes and 24 sub-attributes from the audio files.

The emotion detection benchmark dataset used in this research work contains 290 voice recordings with 24 features and 7 class labels are given in Table 3.

TABLE 5: Twenty-four feature subsets based on SFS.

Model	Selected feature
1	F1
2	F1, F2
3	F1, F2, F4
4	F1, F2, F4, F14
5	F1, F2, F4, F14, F6
6	F1, F2, F4, F14, F6, F7
7	F1, F2, F4, F14, F6, F7, F5
8	F1, F2, F4, F14, F6, F7, F5, F3
9	F1, F2, F4, F14, F6, F7, F5, F3, F10
10	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12
11	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8
12	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8, F17
13	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8, F17, F23
14	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8, F17, F23, F20
15	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8, F17, F23, F20, F19
16	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8, F17, F23, F20, F19, F11
17	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8, F17, F23, F20, F19, F11, F9
18	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8, F17, F23, F20, F19, F11, F9, F13
19	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8, F17, F23, F20, F19, F11, F9, F13, F16
20	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8, F17, F23, F20, F19, F11, F9, F13, F16, F15
21	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8, F17, F23, F20, F19, F11, F9, F13, F16, F15, F18
22	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8, F17, F23, F20, F19, F11, F9, F13, F16, F15, F18, F22
23	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8, F17, F23, F20, F19, F11, F9, F13, F16, F15, F18, F22, F21
24	F1, F2, F4, F14, F6, F7, F5, F3, F10, F12, F8, F17, F23, F20, F19, F11, F9, F13, F16, F15, F18, F22, F21, F24

Table 6 indicates the average predictive accuracies of the proposed linear and nonlinear MLSTSVM classifier on each feature subset. In this research, we used Gaussian and polynomial kernel functions. Among 24 models, model 16 obtained highest predictive accuracy, 87.28% for linear MLSTSVM, 92.89% for Gaussian MLSTSVM, and 88.87% for polynomial MLSTSVM classifiers. Therefore, model 16 is considered as best feature subset.

We have compared the performance of the proposed classifier with KNN, ANN, and multi-SVM on emotion detection dataset as shown in Table 7. All these machine learning algorithms are also implemented with MATLAB R2012a on Windows 7 operating system. It is found that the proposed feature selection based MLSTSVM model with Gaussian kernel function has obtained highest accuracy, 92.89%, as compared to the other approaches. Table 8 indicates the comparison of emotion detection by using various classifiers. MLSTSVM correctly detects neutral, happy, and panic emotions. While for anger emotion, it detects 90% anger and 10% sadness. For anxiety, it detects 80% anxiety, 10% disgust, and 10% sadness. For disgust, it detects 90% disgust and 10% anxiety. For sadness, it detects 90% sadness and 10% disgust.

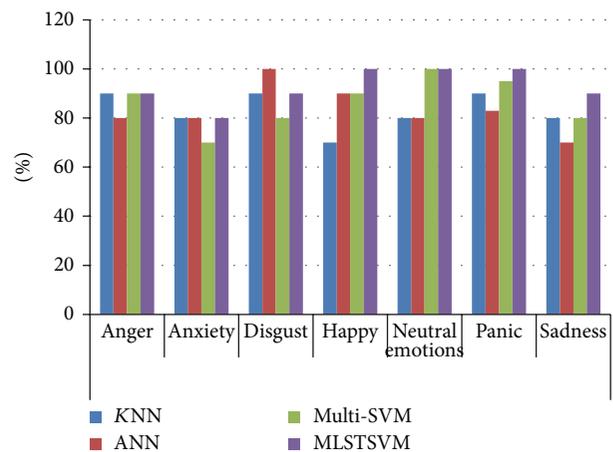


FIGURE 6: Accuracy comparison of various classifiers for different emotions.

Figure 6 indicates the accuracy comparison of KNN, ANN, multi-SVM, and MLSTSVM for different emotions. It is observed from the figure that, for anger emotion, the accuracy of KNN, multi-SVM, and MLSTSVM is 90%, while

TABLE 6: Average accuracy of MLSTSVM classifier on different models.

S. number	Model	MLSTSVM		
		Linear	Gaussian	Polynomial
1	M1	39.33%	47.40%	40.65%
2	M2	36.49%	49.15%	45.54%
3	M3	37.12%	49.20%	50.40%
4	M4	40.16%	58.90%	50.98%
5	M5	44.50%	60.00%	55.23%
6	M6	40.28%	60.32%	60.50%
7	M7	49.27%	61.20%	60.32%
8	M8	50.59%	65.10%	65.60%
9	M9	54.05%	65.60%	61.92%
10	M10	52.74%	67.80%	63.57%
11	M11	50.98%	66.66%	66.92%
12	M12	61.31%	67.80%	74.63%
13	M13	78.85%	81.36%	79.55%
14	M14	80.95%	86.51%	82.05%
15	M15	84.18%	89.83%	85.36%
16	M16	87.28%	92.89%	88.87%
17	M17	86.19%	88.15%	88.01%
18	M18	83.18%	85.37%	84.54%
19	M19	83.06%	84.83%	83.20%
20	M20	80.50%	81.10%	82.07%
21	M21	82.18%	83.72%	83.50%
22	M22	81.26%	82.58%	83.49%
23	M23	82.10%	83.08%	80.30%
24	M24	83.91%	86.83%	84.28%

TABLE 7: Performance comparison.

Classifier method	Accuracy
KNN	82.85%
ANN	83.35%
Multi-SVM	86.67%
Linear MLSTSVM	83.91%
Polynomial MLSTSVM	84.28%
Gaussian MLSTSVM	86.83%
FS+ linear MLSTSVM	87.28%
FS+ polynomial MLSTSVM	88.87%
FS+ gaussian MLSTSVM	92.89%

the accuracy of ANN is 80%. The accuracy of the proposed MLSTSVM is 100% for happy, neutral, and panic emotions. For disgust emotion, the accuracy of ANN is highest as compared to other approaches. So, it is concluded that the accuracy of MLSTSVM is lower only for disgust emotion as compared to ANN, while for other emotions, its accuracy is either equal or higher with respect to other approaches.

We also have checked the performance of the proposed system on real data. For this purpose, we have recorded the voice of 10 male persons by using Nexus 5 smart phone. Nexus 5 voice recorder has dual membrane MEMS microphones to record audio. Each membrane serves a specific purpose, one

focused on high-sensitivity while the other focused on high sound pressure. After the two signals have been combined, background noise is greatly reduced, which increases the overall sound quality of the recording. The real data contains 160 voice recordings of seven different emotions. The performance comparison of various classifiers such as KNN, ANN, multi-SVM, and MLSTSVM on real dataset is shown in Table 9. So, the proposed MLSTSVM also works well for real data as compared to the other existing approaches and obtains highest accuracy, 86.18%.

5. Conclusion

The proposed emotion detection system can be used in automated call centres or in a nursing home where resources or nursing staff may not be available to aid everyone. Automated emotion detection system can be useful to identify the emotion change of patients and to trigger the alarm according to their emotion change so that effective treatment or facility can be provided to patients as soon as possible. This system can assist the clinician to perform the task of emotion detection more efficiently. The proposed emotion detection system may serve as an important tool because change in emotion is responsible for several diseases such as PTSD, BMD, OCD, depression, and suicide. In this paper, emotions are detected by using a novel classifier, named MLSTSVM, and its performance is validated on five benchmark datasets. PRAAT scripting tool is used for feature extraction and extracted 24 features from voice recording. The combination of *F*-score and SFS is used for selecting important features from emotion detection dataset. It is found that MLSTSVM classifier based emotion detection system with sixteen features has achieved better predictive accuracy, 87.28% for linear MLSTSVM, 92.89% for Gaussian MLSTSVM, and 88.87% for polynomial MLSTSVM classifier. The performance of proposed system is compared with KNN, ANN, and multi-SVM approaches. Experimental results indicate that our proposed novel classifier based emotion detection performs well as compared to the other existing approaches. The results of proposed classifier are also verified by using real dataset containing the voice of 10 persons with different emotions and obtained 86.18% accuracy with Gaussian MLSTSVM. The whole system may be adopted and extended as an intelligent personal assistant application for helping disable, autistic children, psychic patients, and elderly people. Apart from healthcare, importance of automatically recognizing emotions from human speech as achieved here in this proposed system may also be used as a part of human computer interaction applications such as robotics, games, and intelligent tutoring system. We have developed the emotion recognition system using MATLAB on Windows operating system. The system has certain limitations; for example, it does not deal with the background noise and is trained for male voices only. Hence few enhancements are possible in the future. A better performance could be guaranteed by optimizing the values of certain parameters like sigma (for Gaussian kernel function) and cost parameters by using genetic algorithm, particle swarm optimization, or any other optimization approaches.

TABLE 8: Comparison of emotion detection of various classifiers.

Original emotion	Detected emotion			
	KNN	Multi-SVM	ANN	MLSTSVM
Anger	90% anger, 10% happiness	90% anger, 10% happiness	80% anger, 20% sadness	90% anger, 10% sadness
Anxiety	80% anxiety, 20% sadness	70% anxiety, 20% happiness, and 10% sadness	80% anxiety, 20% happiness	80% anxiety, 10% disgust, and 10% sadness
Disgust	90% disgust, 10% anxiety	80% disgust, 10% anxiety, and 10% neutral	100% disgust	90% disgust, 10% anxiety
Happiness	70% happiness, 10% anger, and 20% sadness	90% happiness, 10% panic	90% happiness, 10% anger	100% happiness
Neutral emotions	80% neutral, 20% anxiety	100% neutral	80% neutral, 20% anxiety	100% neutral
Panic	90% panic, 10% disgust	95% panic, 5% sadness	83% panic, 7% sadness, and 10% neutral	100% panic
Sadness	80% sadness, 20% anxiety	80% sadness, 20% anxiety	70% sadness, 20% anxiety, and 10% disgust	90% sadness, 10% disgust

TABLE 9: Performance comparison on real dataset.

Classifier	Accuracy
KNN	80.37%
ANN	78.61%
Multi-SVM	83.23%
MLSTSVM	86.18%

Conflict of Interests

The authors declare that they have no conflict of interests regarding the publication of this paper.

References

- [1] S. Tokuno, G. Tsumatori, S. Shono et al., "Usage of emotion recognition in military health care," in *Proceedings of the Defense Science Research Conference and Expo (DSR '11)*, pp. 1–5, Singapore, August 2011.
- [2] A. H. Miller, V. Maletic, and C. L. Raison, "Inflammation and its discontents: the role of cytokines in the pathophysiology of major depression," *Biological Psychiatry*, vol. 65, no. 9, pp. 732–741, 2009.
- [3] C.-N. Anagnostopoulos, T. Iliou, and I. Giannoukos, "Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011," *Artificial Intelligence Review*, 2012.
- [4] S. G. Koolagudi and K. S. Rao, "Emotion recognition from speech: a review," *International Journal of Speech Technology*, vol. 15, no. 2, pp. 99–117, 2012.
- [5] S. Ramakrishnan and I. M. M. El Emary, "Speech emotion recognition approaches in human computer interaction," *Telecommunication Systems*, vol. 52, no. 3, pp. 1467–1478, 2013.
- [6] K. Shiomi, "Voice processing technique for human cerebral activity measurement," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '08)*, pp. 3343–3347, October 2008.
- [7] T. Pfister, "Emotion Detection from Speech," 2010.
- [8] H. Sato, Y. Mitsukura, M. Fukumi, and N. Akamatsu, "Emotional speech classification with prosodic parameters by using neural networks," in *Proceedings of the 7th Australian and New Zealand Intelligent Information Systems Conference*, pp. 395–398, IEEE, 2001.
- [9] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011.
- [10] H. Altun and G. Polat, "Boosting selection of speech related features to improve performance of multi-class SVMs in emotion detection," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8197–8203, 2009.
- [11] D. Tomar and S. Agarwal, "A survey on data mining approaches for healthcare," *International Journal of Bio-Science and Bio-Technology*, vol. 5, no. 5, pp. 241–266, 2013.
- [12] D.-Y. Huang, Z. Zhang, and S. S. Ge, "Speaker state classification based on fusion of asymmetric simple partial least squares (SIMPLS) and support vector machines," *Computer Speech and Language*, vol. 28, no. 2, pp. 392–414, 2014.
- [13] M. Sheikhan, M. Bejani, and D. Gharavian, "Modular neural-SVM scheme for speech emotion recognition using ANOVA feature selection method," *Neural Computing and Applications*, vol. 23, no. 1, pp. 215–227, 2013.
- [14] M. A. Kumar and M. Gopal, "Least squares twin support vector machines for pattern classification," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7535–7543, 2009.
- [15] D. Tomar and S. Agarwal, "Twin support vector machine: a review from 2007 to 2014," *Egyptian Informatics Journal*. In press.
- [16] 2014, <https://archive.ics.uci.edu/ml/datasets.html>.
- [17] 2014, <http://www.fon.hum.uva.nl/praat>.
- [18] http://www.stanford.edu/dept/linguistics/corpora/material/PRAAT_workshop_manual_v421.pdf.
- [19] D. Tomar and S. Agarwal, "Feature selection based least square twin support vector machine for diagnosis of heart disease," *International Journal of Bio-Science and Bio-Technology*, vol. 6, no. 2, pp. 69–82, 2014.
- [20] S. Agarwal and D. Tomar, "A feature selection based model for software defect prediction," *International Journal of Advanced Science and Technology*, vol. 65, pp. 39–58, 2014.

- [21] D. Tomar and S. Agarwal, "A survey on pre-processing and post-processing techniques in data mining," *International Journal of Database Theory & Application*, vol. 7, no. 4, 2014.
- [22] J. Xie and C. Wang, "Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5809–5815, 2011.




Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

