Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2022.030406
ISSN 2790-0983 Vol. 3, Issue 4: 44-52

Speech Emotion Recognition Based on Fuzzy Support Vector Machine

Hui Ma[*]

College of Foreign Language, Chongqing University of Technology, Chongqing, China

65559003@qq.com

[*]*corresponding author*

Keywords: Support Vector Machine, Speech Emotion, Emotion Recognition, FSVM Method

Abstract: Language is an important part of everyday information communication, and automatic speech emotion information recognition technology has broad application prospects in education, services and medicine. The purpose of this paper is to study speech emotion recognition by specific fuzzy support vector machines. The applicability of fuzzy theory for solving speech emotion states is investigated and the theory of support vector machines, fuzzy support vector machines, is introduced to analyse speech emotion feature extraction in terms of both short-time energy and resonance peaks. The performance of SVM and FSVM methods are compared using Emo-DB corpus tests and it is found that the FSVM method has a higher recognition rate than the SVM method.

1. Introduction

The transmission of information through speech signals is an important way for humans to communicate their thoughts and feelings, and since its inception, computers have been expected to communicate smoothly between humans and machines [1]. The gradual development of speech signal processing technology has enabled human-computer interaction to transition from a single keyboard and screen input and output method to a more humane voice communication method. With the support of speech sensors and other electronic technologies, various speech electronic terminals have become an essential tool in daily office, transportation, business and medical activities. However, in traditional speech signal processing techniques the importance of the emotional information contained in the speech signal is severely underestimated and is even considered as noise that is eliminated by various regularised pattern anomalies processing techniques, resulting in biased perception by the listener [2].

Speech emotion recognition (SER) is a challenging task and its performance often depends on the effectiveness of its feature classification. Therefore, Bharat Richhariya proposed a multiscale convolutional recurrent neural network (AMCRNN) for a more comprehensive representation of multiscale features. Furthermore, we introduce prior knowledge to guide their model discriminative learning and evaluate the proposed model, and the results show that their approach can achieve comparable performance [3]. Idowu Sunday Oyetade proposed a feature extraction technique based on histogram of gradients (HOG) oriented and fuzzy support vector machine (FSVM). In the FSVM model, a multiplane-based fuzzy support vector machine is constructed using a new membership function, which improves the classification function, reduces noise point interference and increases the classification efficiency. Extensive experiments on various benchmarks have shown that this method outperforms other methods [4].Kicheol Jeong proposed a class of fuzzy support vector machines based on OC-FSVM-RCH to simplify the convex hull of large noisy data classification. The ordinary class data is encapsulated in a minimal superdomain to maximise the boundaries between anomaly class data and superdomain data [5]. How to make computers intelligent so that they can capture the emotion signals of users for analysis and processing, and then further provide a more friendly communication environment for the users while minimising the hindrance of emotion conversion between the operator and the machine has become an urgent problem in computing [6-7].

This paper introduces the research background and importance of speech emotion recognition, details the development and application of speech emotion recognition, analyses the current state of research at home and abroad, and also analyses the current stage of the construction of the membership function of the general support vector machine, the classification principle membership function construction problem. The problem of non-linear, small sample and large size pattern recognition is solved using fuzzy support vector machines. In order to fully illustrate the advantages of the algorithm, data comparisons are designed and the results are illustrated.

2. A Study of Speech Emotion Recognition Based on Fuzzy Support Vector Machines

2.1. Support Vector Machines

The SVM (SVM) is a structural risk minimization classifier based on statistical theory that allows us to better solve small sample learning problems [8-9]. The central idea is that for an input space linear problem, the small linear problem is transformed into a large linear problem by assigning spatial sampling points to a large feature space through an appropriate kernel function. For the following T samples :

$$\begin{cases}(x1,y1),(x2,y2),...,(x_i,y_i),(x_T,y_T)\\ y_i\in\{-1,+1\};x_i\in R^d;n=1,2,...,T\end{cases} \tag{1}$$

Finding the optimal classification plane that maximizes the classification interval is the basic idea of a support vector machine.

2.2. Fuzzy Support Vector Machines

Each sample trained by a fuzzy support vector machine adds an affiliation term in addition to the sample features and class identification [10]. Let the training sample set be $\{x_i,y_i\}_{i=1}^{n}\in R^m*\{+1,-1\}$ and the anonymous mapping of the kernel function be $\phi(x)$, the training samples become $(\varphi(x_i),y_i)$, and then a fuzzy factor $s_i(0<s_i\le 1)i=1,2,,,n$ is introduced to represent the difference in weights assigned to different sample points. The optimisation problem for solving the ideal

hyperplane is :

$$\begin{aligned}&\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} s_i\xi_i\\&s.t. y_i[w\cdot\varphi(x_i)+b]-1+\xi_i \geq 0\\&\xi_i \geq 0, i=1,2,3...,n\end{aligned} \tag{2}$$

Where c is a constant, w is the weight, and b is the offset. The Lagrangian dual problem of Eq. (2) is shown in Eq. (3):

$$\begin{aligned}&\max w(\alpha)=\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j k(x_i\cdot x_j)\\&\sum_{i=1}^{n}\alpha_i y_i = 0, 0\leq\alpha_i\leq s_i C, i=1,2,...,n\end{aligned} \tag{3}$$

The above equation $k(x_i, x_j)=\varphi(x_i)\varphi(x_i)^T$ is a kernel function [11-12].

2.3. Speech Emotion Feature Extraction

(1) Short-time energy

Different emotional types of speech signals have different amplitudes, and different emotions can be distinguished by the difference in amplitude, and the short-time energy of speech is a function of the change in amplitude of the speech signal. the tone will be high, i.e. the corresponding short term energy is high, and when low, e.g. sad, the tone will be more subdued, i.e. the corresponding short term energy is low [13-14].

(2) Resonance peaks

The fact that people can still recognise and understand sounds in noisy environments is closely related to the resonance peak feature that carries the discriminative properties of the sound, in addition to the robustness of humans to noise [15-16]. Treating the vocal tract as a resonant cavity for articulation, the vocal tract will oscillate back and forth to its maximum amplitude when the excitation frequency is equal to the resonant frequency of the vocal tract, i.e. resonance is generated, which causes the resonant cavity to vibrate to produce resonant frequencies, and the amplified resonant frequencies will peak up one after another, these resonant frequencies are the resonant frequencies and the peaks generated are the resonant peaks [17-18].

3. Experimental Study of Speech Emotion Recognition Based on Fuzzy Support Vector Machine

In the experiments of this paper, four categories of emotion speech samples, namely happy, angry, sad and calm, were selected from the Emo-DB corpus, of which 40 utterances from each speech category were selected as training samples. Ten test utterances of each of the four emotion categories, happy, angry, sad and calm, were then selected from the Emo-DB corpus for recognition.

The structure of the speech emotion recognition system for the fuzzy support vector machine classification method is shown in Figure 1. The test and training utterances are endpoint detected with two thresholds combining the short-time average over-zero rate and the short-time average energy, pre-emphasised with a pre-emphasis factor of 0.8256, and then selected with a frame length of 30ms and a frame shift of 15ms for framing and truncated with a Hamming window for utility pre-processing. In each training utterance sample and test utterance sample, 80 frames of short-time signal were selected, and then short-time average energy, short-time average over-zero rate, first resonance peak, second resonance peak, third resonance peak, fundamental tone, 26-dimensional MFCC and other sentiment feature parameters were extracted.

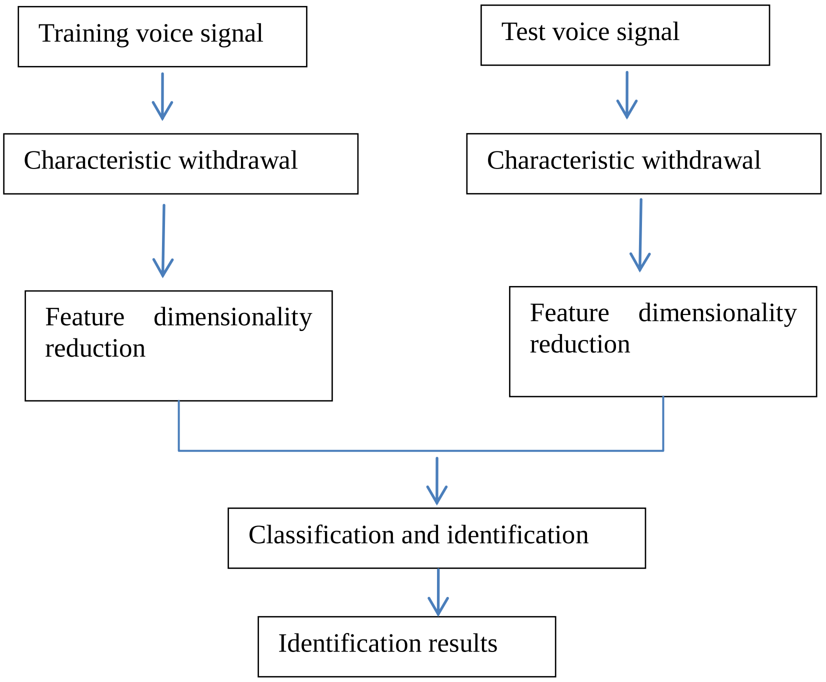
Training voice signal
Test voice signal
Characteristic withdrawal
Characteristic withdrawal
Feature dimensionality reduction
Feature dimensionality reduction
Classification and identification
Identification results

Figure 1. Speech emotion recognition system diagram under fuzzy support vector machine classification method

In using FSVM, it is necessary to first number the emotional categories, numbered as happy → 1, angry → 2, sad → 3 and calm → 4. The training feature projections of the four emotional feature parameters and the numbers of the emotional categories are used as the input xi and output yi of FSVM, respectively. The affiliation degree si can be found with the K-mean algorithm, and here the MATLAB function can be called to find si, and the training sample So of FSVM is obtained according to xi, yi, si. The test feature projection is used as the input Xo of the FSVM to be classified, and finally the classification recognition results are counted.

4. Analysis and Research of Speech Emotion Recognition Based on Fuzzy Support Vector Machine

4.1. Comparison of SVM Classification Method and FSVM Method

The feature parameters here are short-time average energy, short-time average over-zero rate, first resonance peak, second resonance peak and third resonance peak, fundamental tone, 26-dimensional MFCC and other sentiment feature parameters of a total of 32 dimensions as the object of processing. The 32-dimensional data were processed by KPCA to reduce the dimensionality and redundancy, and then the SVM and FSVM methods in this paper were used for comparison experiments. Here the penalty factors C and C0 are taken as 1 for both SVM and FSVM, and the kernel function is chosen as Gaussian kernel function with δ taken as 1.

Table 1. Comparison results of SVM classification method and FSVM method

Emotional category	SVM	FSVM
Happy	80%	88%
Get angry	79%	83%
Sadness	77%	82%
Calm	70%	80%

From Table 1 above, it can be seen that under the comparison of FSVM and SVM methods, the recognition rate of speech emotion is improved by both FSVM methods because FSVM methods are more adaptable to fuzzy variables like emotion, where the recognition rate of happy is increased by 8%, angry by 4%, sad by 5% and calm by 10%, and the recognition rate of calm is increased the most, while the recognition rate of happy is the highest, indicating that the feature parameters extracted by happy emotion are more discriminative, as shown in Figure 2.

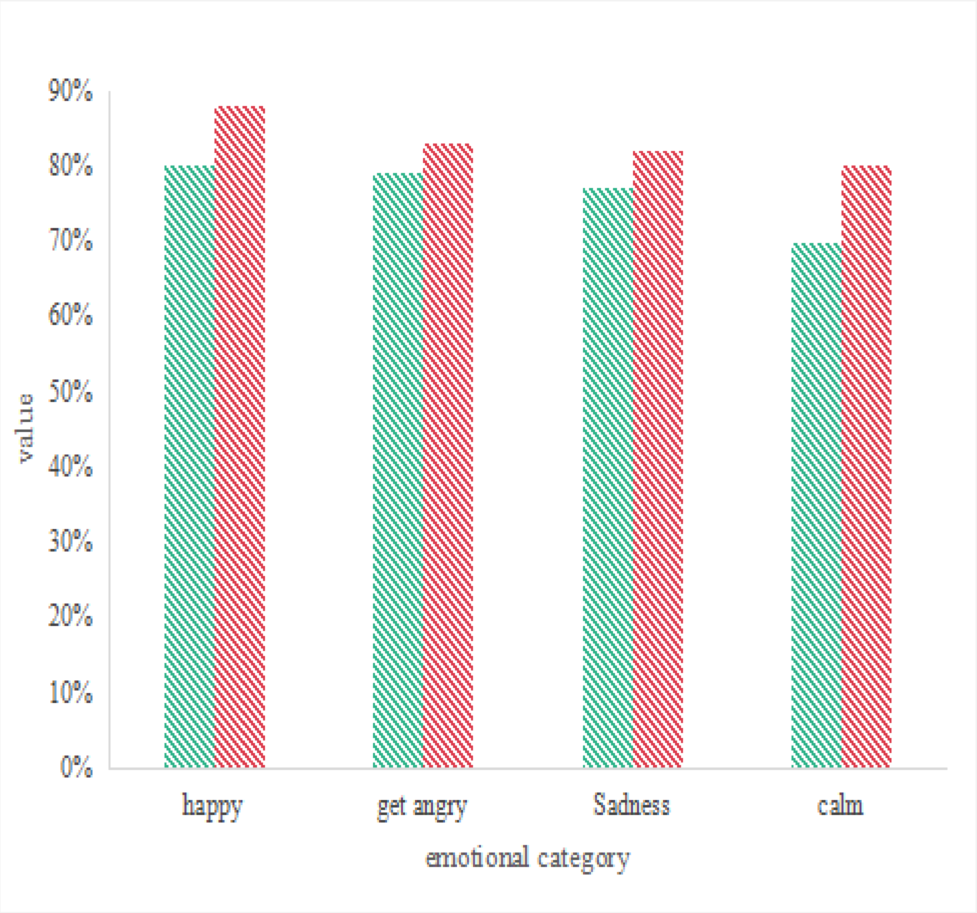

Figure 2. Algorithm comparison

4.2. Comparative Recognition Experiment of Different Single Speech Emotion Features under FSVM Method

Under the FSVM method, a single voice emotion feature is used as the parameter input for recognition, and all feature parameters are 32 dimensions in total as the recognition feature input. Then KPCA is used to reduce the dimension of the 32 dimension feature and remove redundancy as the recognition input feature of FSVM. Here, the penalty factor C0 of FSVM is taken as 1, the kernel function is chosen as Gaussian kernel function, and δ is taken as 1. The comparative experimental results are shown in Table 2.

Table 2. Recognition results of single speech emotion feature under FSVM method

Category	Average energy	First formant	Second formant	Third formant	Pitch	32 dimensional feature	32 dimensional feature+ Kpca
Happy	77	75	80	77	76	70	81
Get angry	87	85	85	84	85	76	88
Sadness	79	81	77	76	73	71	86
Calm	85	88	88	84	86	73	89

It can be seen that the recognition rate for the four emotion categories is the highest after dimensionality reduction of the 32-dimensional features and then using the FSVM method, while among the individual emotion feature parameters, only MFCC has the highest recognition rate under the FSVM method, and among the individual emotion categories, only Calm has the lowest recognition rate. The average recognition rate for calm was around 85%, the average recognition rate for angry was also around 85%, and the recognition rate for happy was around 75%.

4.3. Comparison of FSVM and SVM Methods for Recognition under Different Sex-to-noise Ratios

In order to illustrate the anti-noise performance of FSVM, different noises were added to the experimental speech signal with a noise ratio (SNR) of 20dB, 15dB, 10dB, 8dB and 0dB of Gaussian white noise. The SVM and FSVM methods were used to compare the recognition of 32-dimensional features proposed from the glad emotion category, and the results are shown in Fig. 3. Here the penalty factors C and C0 were taken as 1 for both SVM and FSVM, and the Gaussian kernel function was chosen with δ taken as 1.

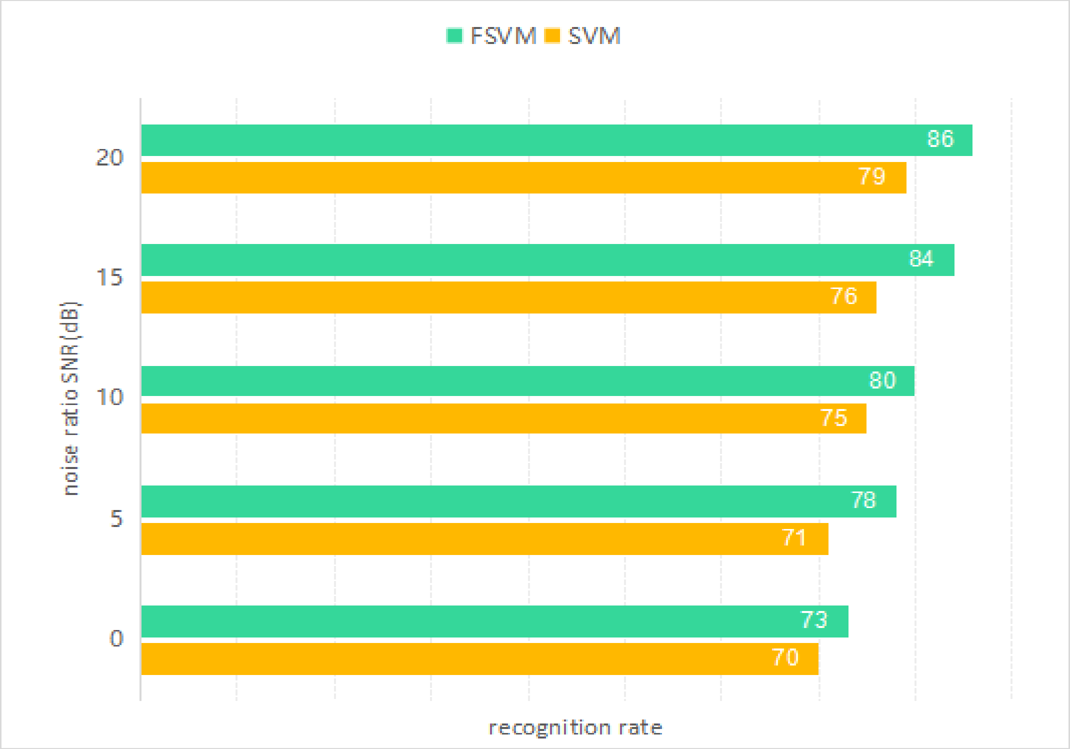

Figure 3. Comparative experimental results of FSVM and SVM methods under different property to noise ratios

Table 3. Comparison of anti noise performance

Noise ratio SNR(dB)	SVM	FSVM
0	70	73
5	71	78
10	75	80
15	76	84
20	79	86

Table 3 shows that when the SNR is 20dB, the recognition rate of FSVM reaches 86% and is 7% higher than that of SVM. When the SNR is 0dB, the recognition rate of FSVM is also above 70%, indicating that FSVM has better noise immunity.

5. Conclusion

Along with social information, the level of intelligence is getting higher and higher, and people are doing more and more research in the field of computers. Computer personalisation to make you feel the attitude and emotion of the operator is an important direction of current research. This paper discusses speech emotion recognition, choosing speech emotion as the object of study and examining some key techniques in speech emotion recognition. Although this paper has achieved some research results, there are some shortcomings. (1) It is recommended to start with complex emotion speech recognition. This paper only contains some limited emotions, but in real life, human emotions are complex and variable, and need not be characterised by certain personal emotions. (2) Speech emotion recognition models can be further investigated. There are various models for speech emotion recognition. This paper uses a support vector machine. Different speech emotion recognition models can also be compared and improved, or different models can be combined, so that a more reasonable and effective speech emotion recognition method can be selected.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Parashjyoti Borah, Deepak Gupta: Affinity and transformed class probability-based fuzzy least squares support vector machines. Fuzzy Sets Syst. 443(Part): 203-235 (2022) https://doi.org/10.1016/j.fss.2022.03.009

[2] Deepak Gupta, Parashjyoti Borah, Usha Mary Sharma, Mukesh Prasad: Data-driven mechanism based on fuzzy Lagrangian twin parametric-margin support vector machine for biomedical data analysis. Neural Comput. Appl. 34(14): 11335-11345 (2022) https://doi.org/10.1007/s00521-021-05866-2

[3] Bharat Richhariya, Muhammad Tanveer: A fuzzy universum least squares twin support vector machine (FULSTSVM). Neural Comput. Appl. 34(14): 11411-11422 (2022) https://doi.org/10.1007/s00521-021-05721-4

[4] Idowu Sunday Oyetade, Joshua Ojo Ayeni, Adewale Opeoluwa Ogunde, Bosede Oyenike Oguntunde, Toluwase Ayobami Olowookere: Hybridized Deep Convolutional Neural Network and Fuzzy Support Vector Machines for Breast Cancer Detection. SN Comput. Sci. 3(1): 58 (2022) https://doi.org/10.1007/s42979-021-00882-4

[5] Kicheol Jeong, Seibum B. Choi: Takagi-Sugeno Fuzzy Observer-Based Magnetorheological Damper Fault Diagnosis Using a Support Vector Machine. IEEE Trans. Control. Syst. Technol. 30(4): 1723-1735 (2022) https://doi.org/10.1109/TCST.2021.3123611

[6] Umesh Gupta, Deepak Gupta: Kernel-Target Alignment Based Fuzzy Lagrangian Twin Bounded Support Vector Machine. Int. J. Uncertain. Fuzziness Knowl. Based Syst. 29(5): 677-707 (2021) https://doi.org/10.1142/S021848852150029X

[7] R. R. Thirrunavukkarasu, T. Meera Devi: Empirical Mode Decomposition with Fuzzy Weight Beetle Swarm Optimization (EMD-FWBSO) Denoising and Enhanced Kernel Support Vector Machine (EKSVM) Classifier for Arrhythmia in Electrocardiogram Recordings. J. Medical Imaging Health Informatics 11(11): 2778-2789 (2021) https://doi.org/10.1166/jmihi.2021.3870

[8] Scindhiya Laxmi, Shiv Kumar Gupta, Sumit Kumar: Intuitionistic fuzzy proximal support vector machine for multicategory classification problems. Soft Comput. 25(22): 14039-14057 (2021) https://doi.org/10.1007/s00500-021-06193-3

[9] Somaye Moslemnejad, Javad Hamidzadeh: Weighted support vector machine using fuzzy rough set theory. Soft Comput. 25(13): 8461-8481 (2021) https://doi.org/10.1007/s00500-021-05773-7

[10] R. Jeen Retna Kumar, M. Sundaram, N. Arumugam: Facial emotion recognition using subband selective multilevel stationary wavelet gradient transform and fuzzy support vector machine. Vis. Comput. 37(8): 2315-2329 (2021) https://doi.org/10.1007/s00371-020-01988-1

[11] Anastasia Iskhakova, Daniyar Volf, Roman V. Meshcheryakov: Method for Reducing the Feature Space Dimension in Speech Emotion Recognition Using Convolutional Neural Networks. Autom. Remote. Control. 83(6): 857-868 (2022) https://doi.org/10.1134/S0005117922060042

[12] Md. Shah Fahad, Ashish Ranjan, Akshay Deepak, Gayadhar Pradhan: Speaker Adversarial Neural Network (SANN) for Speaker-independent Speech Emotion Recognition. Circuits Syst. Signal Process. 41(11): 6113-6135 (2022) https://doi.org/10.1007/s00034-022-02068-6

[13] Rajasekhar B, M. Kamaraju, Sumalatha V: Glowworm swarm based fuzzy classifier with dual features for speech emotion recognition. Evol. Intell. 15(2): 939-953 (2022) https://doi.org/10.1007/s12065-019-00262-1

[14] Arul Valiyavalappil Haridas, Ramalatha Marimuthu, Vaazi Gangadharan Sivakumar, Basabi Chakraborty: Emotion recognition of speech signal using Taylor series and deep belief network based classification. Evol. Intell. 15(2): 1145-1158 (2022) https://doi.org/10.1007/s12065-019-00333-3

[15] Tulika Jha, Ramisetty Kavya, J. Jabez Christopher, Vasan Arunachalam: Machine learning techniques for speech emotion recognition using paralinguistic acoustic features. Int. J. Speech Technol. 25(3): 707-725 (2022) https://doi.org/10.1007/s10772-022-09985-6

[16] Pradeep Tiwari, Anand D. Darji: Pertinent feature selection techniques for automatic emotion recognition in stressed speech. Int. J. Speech Technol. 25(2): 511-526 (2022) https://doi.org/10.1007/s10772-022-09978-5

[17] Kasiprasad Mannepalli, Panyam Narahari Sastry, Maloji Suman: Emotion recognition in speech signals using optimization based multi-SVNN classifier. J. King Saud Univ. Comput. Inf. Sci. 34(2): 384-397 (2022) https://doi.org/10.1016/j.jksuci.2018.11.012

[18] Musatafa Abbas Abbood Albadr, Sabrina Tiun, Masri Ayob, Fahad Taha AL-Dhief, Khairuddin Omar, Mhd Khaled Maen: Speech emotion recognition using optimized genetic algorithm-extreme learning machine. Multim. Tools Appl. 81(17): 23963-23989 (2022) https://doi.org/10.1007/s11042-022-12747-w