Published in AJSE, Vol:22, Issue: 2
Received on 23rd February 2023
Revised on 9th May 2023
Accepted on 30th June 2023

AIUB JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 1608 – 3679 (print) 2520 – 4890 (Online)

Comparison of Various Feature Selection Algorithms in Speech Emotion Recognition

Kamaldeep Kaur and Parminder Singh

***Abstract*—Speech Emotion Recognition (SER) plays a predominant role in human-machine interaction. SER is a challenging task because of number of complexities involved in it. For an accurate emotion classification system, feature extraction is the first and important step carried out on speech signals. And after the features are extracted, it is very important to select the best features out of all and reject the redundant and least important features. Feature selection methods play an important role in SER performance. The classifier gets the selected features, so as to reduce the unnecessary overload and perform better to classify the emotions. In this study, a good combination of features is selected from Punjabi Emotional Speech Database. Then a number of feature selection algorithms are explored and experimented upon, to select the best features. 1D-CNN is used for classification purpose. The results are shown and compared on the basis of number of performance metrics. LASSO has shown the best performance results as compared to other feature selection methods.**

***Index Terms*— Speech, features, feature extraction, feature selection, CNN, Punjabi**

I. Introduction

Human beings use speech as the natural and most common mode of interaction among themselves. The researchers have been motivated to develop methods in human-machine interaction, using speech signals as their field of research [1][2]. A lot of information is embedded in the speech signals, in the form of emotions. There should be a machine which is intelligent enough to recognize emotions from a human's speech. The researchers are exploring the human audio speech to judge the emotions in context with many medical and other applications [3].

People in various different regions throughout the world, exhibit distinct languages, styles of speaking, speaking rates, ethnic and societal backgrounds. This difference in the cultural perspectives, makes the process of SER very challenging[2].

Kamaldeep Kaur is a Research Scholar at IKGPTU, Kapurthala and Assistant Professor, Dept. of CSE, Guru Nanak Dev Engineering College, Ludhiana (corresponding author, phone: +91 9815452427; e-mail: kamal.gndec@gmail.com)

Parminder Singh is Professor and Head of Department, Dept. of CSE, Guru Nanak Dev Engineering College, Ludhiana. (e-mail: parminder2u@gmail.com).

Preparing/ choosing an emotional speech database is the foremost and first step in SER. A predominant ingredient of SER is to extract the speech features that represent the emotions hidden within the speech effectively. The accomplishment of SER system is dependent on the dataset, standard of features representing the emotions, and classifiers used for training the system [4][5]. But the feature extraction algorithms may produce a large number of features, all of which may not be effective for the classification process. There is another process, known as feature selection, that has an important place between feature extraction and classification processes. In other words, feature selection is a procedure, where a subset of appropriate features is selected, that is used to identify a dataset with a purpose to improve the performance of an algorithm for some intended work. The extracted features may consist of unrequired, redundant and irrelevant data that do not put up to the model's performance or may even decrease the accuracy. So, the selection of relevant features has become necessary. Accuracy of classification may get improved by feature selection, and it may minimize the complexity of the algorithms. There are some factors which also suggest that some feature selection algorithms may not improve the SER performance, due to their inefficiency to cope up with the type of data and feature correlations [4][5][6][7][8].

In this research paper, the focus is on the effect of various feature selection algorithms on SER for Punjabi language. Punjabi speech emotion dataset is prepared for implementing emotion recognition from Punjabi audio speech signals[9]. A number of different sets of features are explored and a set of 16 features is selected for this research [10]. So, a number of features are then extracted using various methods of extracting features from speech signals. A number of feature selection methods such as Least Absolute Shrinkage and Selection Operator (LASSO), ANOVA, Recursive Feature Elimination (RFE), t-statistics, are then applied on set of extracted features separately, to experiment their effect on the classification process. Then 1-D Convolutional Neural Network (CNN) is used as classifier to recognize emotions from speech.

The prime contribution of this research work is summarized as follows:

- A Punjabi audio speech emotional database is prepared, with 15 non-professional speakers, including 8 females and 7 males, with age ranging between 20-45 years. 6 basic emotions, namely, anger, happy, fear, neutral, sad and surprise, are chosen. 10 Punjabi sentences are

selected for recording purpose, which are neutral in nature. So, 10 sentences are recorded per emotion per speaker, yielding a total of 900 utterances, with 150 utterances per emotion[9].

- A technique is proposed to extract a number of features from Punjabi emotional speech dataset. Different types of feature extraction methods used in this research work involve - Mel spectrogram, MFCC, contrast, pitch, chroma, zero crossing rate, LPCC, tonnetz, formant, jitter, shimmer, entropy, duration, harmonic, PLP and energy[10].
- Four feature selection methods are implemented, including LASSO, ANOVA, RFE, t-statistics, to select relevant features from the total features extracted.
- A CNN model is evolved to give the selected features as input to train it, and then test it, to recognize emotions from Punjabi speech.
- The overall experiments involve the SER process on Punjabi language sentences, with different feature selection algorithms, so as to measure the achievement of these feature selection methods and focusing on their importance in the complete process of SER.
- Performance metrics chosen to compute the results for different feature selection algorithms involve - accuracy, recall, precision and F1-score. The results obtained for different algorithms are compared based on these metrics.

The rest of the paper is organized as follows: section 2 describes in detail the proposed SER process, including database, feature extraction method, feature selection methods and the classifier model. Section 3 elaborates the experiments, their results and analysis in detail. The next section discusses and concludes the work along with its future scope.

II. SPEECH EMOTION RECOGNITION PROCESS

SER has been an emerging field since few years. In literature, various foreign languages are explored by researchers, including Chinese SER with Deep Belief Network [11], Mandarin [12], Persian[13] using Hidden Markov Model with 79.50% performance accuracy, for Polish[14] using k-nearest neighbor, Linear Discriminant Analysis, with performance of 80%. Emotion recognition has been improved by combining feature selection approaches, ranking models, and Fourier parameter models, as well as validating the models against standardized existing speech datasets including CASIA, EMODB, EESDB, FAU Aibo and LDC [15]–[17]. On Berlin EmoDB of speaker-dependent and speaker-independent tests, 2D CNN LSTM network achieves recognition accuracies of 95.33 percent and 95.89 percent, respectively. This contrasts favorably with the accuracy of 91.6% and 92.9% attained by conventional methods [18]. Eight emotional classes from the Ryerson Recordings-Visual Database of Emotional Speech and Song audio are used to train three proposed models (RAVDESS). The proposed models outperformed state-of-the-art models utilizing the same dataset, achieving overall precision between 81.5 percent and 85.5 percent[19].

For Indian languages, [20] has shown work for Assamese using Gaussian Mixture Model, with performance of 74% and highest mean classification score as 76.5 and [21] has shown work for Odia language using prosodic features and Support Vector Machine, with 75% performance. Tamil has been explored by [22] with 71.3%, Bengali by [23] with 74.26%, Malayalam by [24] with Support Vector Machine and Artificial Neural Network with 88.4% performance, Telugu by [22], [25] with performance of 81% and Hindi with 74%[26]. Using an adaptive artificial neural network, emotion recognition from a Marathi speech database has also been accomplished. The experimental research showed that the proposed model is 10.85% more accurate than the standard models[27]. Figure 1 illustrates the SER process.

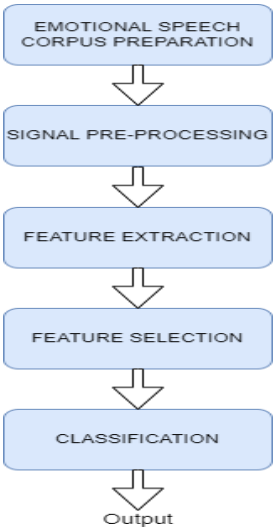

Fig. 1: SER Process

A. *Emotional Speech Corpus Preparation*

As there is no standard about the properties of a particular database, and there was no database that existed for Punjabi language, so for this research work, the authors have prepared Punjabi emotional speech database (PESD), with 15 non-professional speakers, including 8 females and 7 males, with age ranging between 20-45 years. The reason behind taking this mid age group is the voice clarity. The voice clarity and consistency of speech is low in a child and an old person as compared to a mid-age person. Also, gender has an impact on speech features such as pitch, intensity, etc. That is why, a balance is maintained while gender selection, so that the features are not more biased towards particular gender.

6 basic emotions, namely, anger, happy, fear, neutral, sad and surprise, are chosen. The evolutionary theory defines the big six model, which is the most fundamental set of emotions including anger, happiness, neutral, sadness, fear, and surprise. So, these standard emotions are taken into consideration for both databases. There is acoustic difference and similarities between different emotions. According to studies, the acoustic characteristics of happy/angry and sad/neutral are comparable. Longer utterances, shorter inter-word pauses, higher pitch, and energy values with a larger range are characteristics of speech that is connected with anger and enjoyment. Sad speech has a slightly higher tone and a wider range than

neutral speech. Speaking rate, Root Mean Square (RMS) energy, and inter-word are helpful in separating grief from other emotions. There is weak acoustic separability between anger and happiness. There is also similarity between fear/ sadness and happiness/ surprise. 10 Punjabi sentences are selected for recording purpose, which are neutral in nature. The sentences selected are not specific to any emotion. This is done so that the sentence does not sound biased towards a particular emotion only, and reflecting other emotions in it does not seem difficult. So, 10 sentences were recorded per emotion per speaker, yielding a total of 900 utterances, with 150 utterances per emotion [9].

The design of Punjabi emotional speech database is summarized in table I as follows:

TABLE I
DESIGN OF DATABASE

Specification	Value
Speakers	15 (7 males and 8 females)
Age	20-45 years
Emotions	6 (happy, sad, neutral, angry, surprise, fear)
Sentences	10 sentences × 6 emotions × 15 speakers × 1 session = 900 total (150 sentences per emotion)

All the recordings were done at a studio, trying to be done without obstacles in the recording path. The recordings were done using Sennheiser e835 microphone and audacity 2.2.2 software. The sampling rate is taken as 16 KHz, represented as 16-bit numbers, with mono-channel recording.

The whole recording process was monitored, with a single session, along with feedback and instructions throughout the process. The speakers were asked to recall an actual circumstance from their past when they had experienced this emotion. Following this method, the speakers tried to re-capture the emotions by developing the same physiological effects as in the real situation. The speakers produced each of the sentences as many times as they liked with several variants of a sentence, and out of those, one was selected for further process.

A listening test was undertaken after the recording session to examine if typical listeners could recognize the type or class of emotion that the recorded statements belonged to.

B. Signal Pre-processing

It is always necessary to pre-process the signal to remove noise and echo, and for signal smoothening. For this research work, the audio files in the database are converted in '.wav' format, with mono-channel, 16 KHz sampling rate and 16-bit rate, and then pre-processed using various standard filters, such as Kalman filter [28], Normalized Least Mean Square (NLMS) with FIR filter [29], Wiener filter [30][31]. Kalman filter is used for noise reduction, NLMS for echo cancellation and Wiener filter for smoothening of speech signal.

The evaluation of speech quality is measured with metrics such as Signal to Noise ratio (SNR), Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Perceptual Evaluation of Speech Quality (PESQ).

Table II summarizes the pre-processing results as follows:

TABLE II
OUTPUT OF PRE-PROCESSING

Emotion	Input Speech Signal					Output Speech Signal				
	SNR	MSE	RMSE	MAE	PESQ	SNR	MSE	RMSE	MAE	PESQ
Happy	0.43	3.04	1.74	0.22	4.00	11.93	2.40	1.54	0.10	4.40
Sad	5.25	1.49	1.22	1.97	4.08	6.41	0.97	0.98	1.13	4.23
Neutral	0.78	4.27	2.06	3.62	1.71	3.70	2.26	1.50	1.98	2.64
Surprise	0.24	3.90	1.97	0.72	2.53	4.25	2.21	1.48	0.39	3.78
Fear	0.36	3.24	1.8	1.14	3.61	10.88	2.01	1.41	0.65	4.12
Anger	0.22	2.23	1.49	2.51	3.25	3.37	1.51	1.22	1.89	4.05

C. Feature Extraction

After the speech signal is pre-processed, a number of features are extracted from the signal, which are now the representatives for the human speech uttered. These emotions embedded in the input signal, are maximally represented by these features [32].

There are different types of features, including prosodic features such as energy, pitch, zero crossing rate, duration, and their derivatives, and spectral features such as Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC) and Linear Prediction Cepstral Coefficients (LPCC) [33].

It has been proved in many researches that fusion of different features results in better accuracy as compared to using a single feature for the work. For this research work, the main challenge is to decide that which features would work best for this system out of the several feature extraction methods available for the voice stream. As it is demonstrated in numerous studies, combining a variety of features improves accuracy as compared to employing only one feature for the task. This is also true when it comes to emotion recognition. So, a combination of number of various features are tested with different experiments and a set of 16 features is selected. A number of prosodic and spectral features are fused together in hybrid form, and extracted to enhance the system performance. This impact of feature extraction on SER is also shown by authors in [10]. In total, 523 features are extracted with details listed in table III.

TABLE III
LIST OF FEATURES EXTRACTED FROM PESD

Feature	Number of attributes
Pitch	181
Mel Spectrogram	128
MFCC (Mel Frequency Cepstral Coefficient)	120
LPCC (Linear Predictive Cepstral Coefficients)	19
PLP (Perceptual Linear Prediction)	19
Chroma	12
Formant	9
Contrast	7

Tonnetz	6
Shimmer	6
Jitter	5
Energy	4
Entropy	4
Duration	1
Harmonic	1
ZCR (Zero Crossing Rate)	1

D. Feature Selection

Feature selection is an important step to be followed after feature extraction. Feature selection methods select a subset of features from the complete set of features extracted. It may remove the duplicate features, or the features which are of least importance and don't tend to contribute to the system performance. Analysis of large number of features, may cause problems, and slow down the learning of SER system. Feature selection methods are used to solve this problem. The feature selection algorithms may improve the classification accuracy, and may reduce the complexity of the system. The prime goal of feature selection is to discover the variables, which can lead to the best performance of classification from the determined feature's subset. The selection of features attempts to strengthen the accuracy and performance of classification, by reducing the size of feature data set. There is ample number of methods found in the literature for feature selection, but not all of them tend to improve the success of SER. Some algorithms may also reduce the success of SER after reducing the size of feature set. Also, there are some methods which are not specific to SER, but applicable to some other area of study, which may reduce SER performance [5][7][8][34].

The feature selection algorithms can be broadly categorized into three types as shown in figure 2 and described briefly.

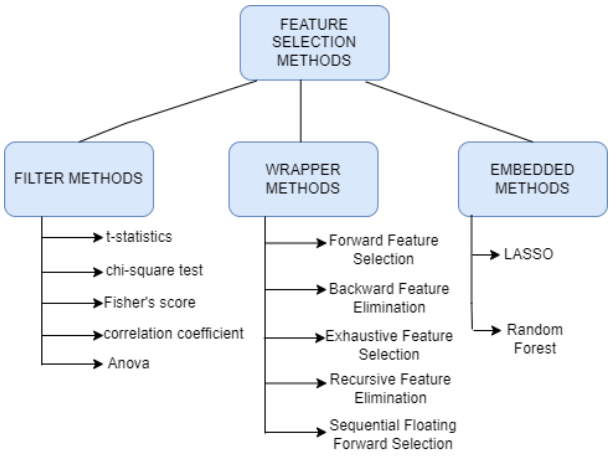

Fig. 2: Types of Feature Selection Methods

i) Filter methods: They use a measure to determine whether a particular feature is useful or not. They take less time to compute and do not try to overfit the data. They are blind to any interactions or correlations between characteristics, which is a downside. The various algorithms include ANOVA, t-statistics, chi-square test, F-test, correlation coefficient, etc.

ii) Wrapper methods: They create models using a particular selection of features and rank each feature's significance. Once the ideal subset is found, they repeat the process and test a different subset of traits. When there are few data points, they have a long computation time and have a tendency to overfit the model. They are more accurate in making predictions than filtering techniques. Exhaustive Feature Selection, Recursive Feature Elimination, Sequential Floating Forward Selection, and Forward Feature Selection are few of the several algorithms.

iii) Embedded methods: By including feature interactions, they combine the advantages of wrapper and filter approaches, yet retaining an acceptable computational cost. It is incremental and carefully selects the features that are most helpful in training for a given iteration. The different algorithms include Ridge Regression, LASSO and others.

In our research, we have chosen four feature selection methods for experiments as we wanted to test at least one method out of each category. They are briefly described below:

a) t-statistics: A *t*-test uses the t-statistics, in order to find the rejection or acceptance of null hypothesis. It is employed when the standard deviation of population is not known, or when the size of sample is small. For example, the t-statistic estimates the mean of population from a sample distribution if the population standard deviation is unknown [35].

b) ANOVA: ANOVA stands for Analysis of variance. As the name suggests, it uses variance as its parameter to compare multiple independent groups. ANOVA can be one-way ANOVA or two-way ANOVA. One-way ANOVA is applied if there are more than two independent groups of a variable [36].

c) RFE: RFE is more widely used, for the reason that it targets those features which are the most relevant with respect to the target variable. It is much effective, and easier than other methods to configure and use [37].

d) LASSO: The statistical methods may have some prediction errors, which are significantly minimized by LASSO. LASSO selects every non-zero feature, then works on regularising the model parameters, shrinks the coefficients of regression, which may lead to some of the coefficients as zero. Then it proceeds towards the feature selection phase. LASSO gives an upper bound to the sum of a model, to act as a constraint, to include and exclude some specific parameters [38].

E. Classification

A number of classifiers are reported in literature, including the conventional ones such as Support Vector Machine [39], Gaussian Mixture Model [40], Hidden Markov Model [41], k-NN [42], Decision tree, etc. In recent years, many neural network models have been designed so as to enhance the system performance. Some of these models include Convolutional Neural Network [43], [44], Deep Belief Network [45], Long Short-Term Memory [46], Recurrent Neural Network [47], etc.

In our experimental work, 1-D Convolutional Neural Network (CNN) is used. The model designed consists of input layer, then 1D-CNN layer, followed by 4 more 1D-CNN layers with Batch Normalization, Max Pooling and ReLU activation function. There is flatten layer after these, followed by two fully connected dense layers with Softmax

activation function. The Batch Normalization layer is used to avoid overfitting. When applying batch normalization, an input feature is always in conjunction with other features in each batch. Therefore, each normalized feature produced by Batch Normalization layer is no longer a deterministic value for each input feature. This effect facilitates the generalization of the deep networks, speeds up the training and reduces overfitting in experiments. The Max Pooling layer makes the features robust against noise and distortion. It divides the input into a set of non-overlapping regions and outputs the maximum value of each such sub-region. The top layer of this architecture is Softmax classifier, which is utilized to recognize the emotion according to the learned features [18]. The model gets the input in the form of training data and produces predicted emotion as output with the test data.

III. EXPERIMENTATION

Five experiments have been carried out. For all the experiments, there are two modules of CNN, that is, training and testing, in which the total data samples are divided. In this proposed methodology, 720 speech signals are used for training phase, and 180 speech signals are used for testing phase. Features as listed in table III are extracted during the feature extraction phase. After this, the relevant features are selected by feature selection algorithm. Four feature selection algorithms are implemented on the set of extracted features shown in table III. Table IV gives the details of experiments done.

TABLE IV
EXPERIMENTS PERFORMED

Experiment	Feature Selection Algorithm Used
1	NIL (Without Feature Selection Algorithm)
2	t-statistics
3	ANOVA
4	RFE
5	LASSO

The first experiment doesn't incorporate any feature selection algorithm. It involves feature extraction phase, and then classification process. In this experiment, there are maximum number of features, as listed in table III, which are fed to CNN. All other experiments involve the use of a feature selection algorithm as listed in table IV. t-statistics, ANOVA, RFE and LASSO are used in experiments 2,3,4 and 5 respectively. All of these 4 methods select 360 features, out of 523, to be further used in SER process. They remove the redundant features from the set, and create a subset, which they feel is the most suitable one. Table V elaborates the features selected and their number of attributes by t-statistics, ANOVA, RFE and LASSO respectively.

TABLE V
FEATURES SELECTED BY FEATURE SELECTION ALGORITHMS

Feature	Number of Attributes			
	t-statistics	ANOVA	RFE	LASSO
Pitch	180	91	150	135
Mel Spectrogram	71	128	70	78
MFCC (Mel Frequency Cepstral Coefficient)	60	79	110	112
LPCC (Linear Predictive Cepstral Coefficients)	11	8	3	4
PLP (Perceptual Linear Prediction)	2	8	2	4
Chroma	12	11	3	4
Formant	7	6	3	2
Contrast	7	5	7	7
Tonnetz	0	4	4	5
Shimmer	2	6	1	2
Jitter	0	5	1	1
Energy	1	4	1	2
Entropy	4	2	2	1
Duration	1	1	1	1
Harmonic	1	1	1	1
ZCR (Zero Crossing Rate)	1	1	1	1

The statistical measurements of accuracy, precision, recall, F1-score are used to examine efficacy of the proposed model. This SER is developed to recognize the six emotion classes, such as anger, happy, sad, surprise, fear and neutral. The experimental results on PESD by proposed model for various feature selection algorithms are shown in table VI.

TABLE VI
EXPERIMENTAL RESULTS

Experiment	Feature Selection Algorithm Used	Precision	Recall	F1-Score	Accuracy
1	NIL	79.6	79.3	79.0	79.3
2	T-STATISTICS	71.4	71.2	71.2	71.26
3	ANOVA	74.5	73.8	74.3	74.1
4	RFE	81	80.2	80	80.5
5	LASSO	82	81	81.29	81

The results can be visualized in graphical form also as shown in figure 3. It can be observed from the results that the system performance is certainly improved by RFE and LASSO methods. In experiment 1, there is no feature selection method used and the classifier has worked on complete set of 523 features, generating an accuracy of 79.3%. In experiments 2 and 3, filter methods are used, namely, t-statistics and ANOVA. But they do not improve the SER performance, with accuracy of 71.26% and 74.1% respectively. The reason for reduction of SER performance is the drawback of filter methods to overlook the correlations between characteristics, and being traditional methods, they are not more specific to SER as compared to more advantageous wrapper and embedded methods. The experiments 4 and 5 have used wrapper method RFE and embedded method LASSO. They perform better than the filter-based methods and have also improved the recognition rate of system with accuracy of 80.5% for RFE and 81% for LASSO.

Fig. 3: Experimental Results

IV. CONCLUSION & FUTURE SCOPE

Feature extraction and selection are important steps in SER. We have focused on the importance of feature selection algorithm in SER methodology. Experiments are conducted to illustrate the impact of feature selection on SER. It is found that the total number of features extracted by the system without feature selection method is 523. Each of the four feature selection methods used in experiments, have selected 360 features. The distribution of features before and after feature selection is shown in tables III and V respectively. It can be seen in table V that the selection of features is completely dependent upon the algorithm used. Some method is giving more importance to one function and other algorithm is considering some other function as most important. The precision, recall, f1-score and accuracy values are given in table VI and also illustrated in the form of graph in figure 3. Maximum accuracy is found in case of embedded method LASSO, which outperforms the wrapper method RFE. These two methods improve the SER results and filter methods ANOVA and t-statistics have no impact on improvement of recognition rate of the system due to their traditional approach.

This research work can be further extended by exploring some more or hybrid feature selection technique, so as to improve the system performance more. The classifier model can be further improved by using some hybrid model with LSTM and CNN, or implementing 2D-CNN.

V. CONFLICT OF INTEREST

There are no conflicts to declare.

VI. ACKNOWLEDGEMENT

This work was supported by Guru Nanak Dev Engineering College, Ludhiana, Punjab (India) and IKG Punjab Technical University, Kapurthala, Punjab (India). The authors are thankful to these organizations for their support in this research work.

REFERENCES

[1] M. Farooq, F. Hussain, N. K. Baloch, F. R. Raja, H. Yu, and Y. Bin Zikria, "Impact of feature selection algorithm on speech emotion recognition using deep convolutional neural network," *Sensors (Switzerland)*, vol. 20, no. 21. pp. 1–18, 2020, doi: 10.3390/s20216008.

[2] 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, doi: 10.1016/j.patcog.2010.09.020.

[3] J. Nicholson, K. Takahashi, and R. Nakatsu, "Emotion Recognition in Speech Using Neural Networks," *Neural Computing & Applications*, vol. 9, no. 4, pp. 290–296, 2000, doi: 10.1007/s005210070006.

[4] T. Özseven, "A novel feature selection method for speech emotion recognition," *Applied Acoustics*, vol. 146, pp. 320–326, 2019, doi: 10.1016/j.apacoust.2018.11.028.

[5] L. Kerkeni, Y. Serrestou, K. Raoof, M. Mbarki, M. A. Mahjoub, and C. Cleder, "Automatic speech emotion recognition using an optimal combination of features based on EMD-TKEO," *Speech Communication*, vol. 114, pp. 22–35, 2019, doi: 10.1016/j.specom.2019.09.002.

[6] S. Kuchibhotla, H. Deepthi, V. Koteswara, and R. Anne, "An optimal two stage feature selection for speech emotion recognition using acoustic features," *International Journal of Speech Technology*, vol. 19, no. 4, pp. 657–667, 2016, doi: 10.1007/s10772-016-9358-0.

[7] L. Bankert, "Feature Selection for Case-Based Classification of Cloud Types: An Empirical Comparison," 1994.

[8] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 4, pp. 1060–1073, 2022, doi: https://doi.org/10.1016/j.jksuci.2019.06.012.

[9] K. Kaur and P. Singh, "Punjabi Emotional Speech Database: Design, Recording and Verification," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 9, no. 4, Dec. 2021, doi: 10.18201/ijisae.2021473641.

[10] K. Kaur and P. Singh, "Impact of Feature Extraction and Feature Selection Algorithms on Punjabi Speech Emotion Recognition Using Convolutional Neural Network," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 21, no. 5, Apr. 2022, doi: 10.1145/3511888.

[11] B. Chen, Q. Yin, and P. Guo, "A study of deep belief network based Chinese speech emotion recognition," *Proceedings - 2014 10th International Conference on Computational Intelligence and Security, CIS 2014*, pp. 180–184, 2014, doi: 10.1109/CIS.2014.148.

[12] A. Milton and S. Tamil Selvi, "Class-specific multiple classifiers scheme to recognize emotions from speech signals," *Computer Speech and Language*, vol. 28, no. 3, pp. 727–742, 2014, doi: 10.1016/j.csl.2013.08.004.

[13] M. Savargiv and A. Bastanfard, "Persian speech emotion recognition," in *2015 7th Conference on Information and Knowledge Technology (IKT)*, 2015, pp. 1–5, doi: 10.1109/IKT.2015.7288756.

[14] A. Majkowski, M. Kołodziej, R. J. Rak, and R. Korczynski, "Classification of emotions from speech signal," *Signal Processing - Algorithms, Architectures, Arrangements, and Applications Conference Proceedings, SPA*, pp. 276–281, 2016, doi: 10.1109/SPA.2016.7763627.

[15] H. Cao, R. Verma, and A. Nenkova, "Speaker-sensitive emotion recognition via ranking: Studies on acted and spontaneous speech," *Computer Speech and Language*, vol. 29, no. 1, pp. 186–202, 2015, doi: 10.1016/j.csl.2014.01.003.

[16] K. Wang, N. An, B. N. Li, Y. Zhang, and L. Li, "Speech emotion recognition using Fourier parameters," *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 69–75, 2015.

[17] H. K. Palo, M. N. Mohanty, and M. Chandra, "Efficient feature combination techniques for emotional speech classification," *International Journal of Speech Technology*, vol. 19, no. 1, pp. 135–150, 2016, doi: 10.1007/s10772-016-9333-9.

[18] J. Zhao, X. Mao, and L. Chen, "Speech emotion recognition using deep 1D & 2D CNN LSTM networks," *Biomedical Signal Processing and Control*, vol. 47, pp. 312–323, 2019, doi: 10.1016/j.bspc.2018.08.035.

[19] M. Ezz-Eldin, A. A. M. Khalaf, H. F. A. Hamed, and A. I. Hussein, "Efficient Feature-Aware Hybrid Model of Deep Learning Architectures for Speech Emotion Recognition," *IEEE Access*, vol. 9, pp. 1–1, 2021, doi: 10.1109/access.2021.3054345.

[20] A. B. Kandali, A. Routray, and T. K. Basu, "Emotion recognition from Assamese speeches using MFCC features and GMM classifier," *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 2008, doi: 10.1109/TENCON.2008.4766487.

[21] M. Swain, A. Routray, P. Kabisatpathy, and J. N. Kundu, "Study of prosodic feature extraction for multidialectal Odia speech emotion recognition," *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, pp. 1644–1649, 2017, doi: 10.1109/TENCON.2016.7848296.

[22] S. R. Krothapalli and S. G. Koolagudi, "Characterization and recognition of emotions from speech using excitation source information," *International Journal of Speech Technology*, vol. 16, no. 2, pp. 181–201, 2013, doi: 10.1007/s10772-012-9175-z.

[23] A. Mohanta and U. Sharma, "Bengali Speech Emotion Recognition," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 2812–2814.

[24] T. M. Rajisha, A. P. Sunija, and K. S. Riyas, "Performance Analysis of Malayalam Language Speech Emotion Recognition System Using ANN/SVM," *Procedia Technology*, vol. 24, pp. 1097–1104, 2016, doi: 10.1016/j.protcy.2016.05.242.

[25] S. G. Koolagudi and K. S. Rao, "Emotion recognition from speech using source, system, and prosodic features," *International Journal of Speech Technology*, vol. 15, no. 2, pp. 265–289, 2012, doi: 10.1007/s10772-012-9139-3.

[26] S. Bansal and A. Dev, "Emotional Hindi speech: Feature extraction and classification," *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, vol. 03, pp. 1865–1868, 2015.

[27] R. V. Darekar and A. P. Dhande, "Emotion recognition from Marathi speech database using adaptive artificial neural

network," *Biologically Inspired Cognitive Architectures*, vol. 23, no. January, pp. 35–42, 2018, doi: 10.1016/j.bica.2018.01.002.
[28] S. Bhattacharyya *et al*., "Speech Background Noise Removal Using Different Linear Filtering Techniques," in *Lecture Notes in Electrical Engineering*, 2018, vol. 475, pp. 297–307, doi: 10.1007/978-981-10-8240-5.
[29] M. Ma, M. Wang, and J. Hu, "Research on adaptive acoustic echo cancellation algorithm in digital hearing AIDS," *AIP Conference Proceedings*, vol. 1864, 2017, doi: 10.1063/1.4992990.
[30] S. Vihari, A. S. Murthy, P. Soni, and D. C. Naik, "Comparison of Speech Enhancement Algorithms," *Procedia Computer Science*, vol. 89, pp. 666–676, 2016, doi: 10.1016/j.procs.2016.06.032.
[31] T. M., A. Adeel, and A. Hussain, "A Survey on Techniques for Enhancing Speech," *International Journal of Computer Applications*, vol. 179, no. 17, pp. 1–14, 2018, doi: 10.5120/ijca2018916290.
[32] J. Rong, G. Li, and Y. P. P. Chen, "Acoustic feature selection for automatic emotion recognition from speech," *Information Processing and Management*, vol. 45, no. 3, pp. 315–328, 2009, doi: 10.1016/j.ipm.2008.09.003.
[33] K. S. Rao and B. Yegnanarayana, "Prosody modification using instants of significant excitation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 3, pp. 972–980, 2006, doi: 10.1109/TSA.2005.858051.
[34] 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*, vol. 43, no. 2, pp. 155–177, 2012, doi: 10.1007/s10462-012-9368-5.
[35] F. Pyrczak, D. M. Oh, F. Pyrczak, and D. M. Oh, "Introduction to the t Test," 2019. doi: 10.4324/9781315179803-28.
[36] 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, doi: 10.1007/s00521-012-0814-8.
[37] X. Chen and J. C. Jeong, "Enhanced recursive feature elimination," in *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*, Jan. 2007, pp. 429–435, doi: 10.1109/ICMLA.2007.35.
[38] V. Fonti, "Feature Selection using LASSO," *VU Amsterdam*, pp. 1–26, 2017.
[39] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998, doi: 10.1023/A:1009715923555.
[40] Y. Chen and J. Xie, "Emotional speech recognition based on SVM with GMM supervector," *Journal of Electronics (China)*, vol. 29, 2012, doi: 10.1007/s11767-012-0871-2.
[41] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989, doi: 10.1109/5.18626.
[42] R. Manjunath, "Dimensionality reduction and classification of color features data using svm and knn," *International Journal of Image Processing and Visual Communication*, vol. 1, pp. 16–21, Jan. 2013.
[43] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," *ArXiv e-prints*, Nov. 01, 2015.
[44] J. Wu, "Introduction to Convolutional Neural Networks," in *Introduction to Convolutional Neural Networks*, 2017, pp. 1–31.
[45] A. Khan and muhammad Islam, *Deep Belief Networks*. 2016.
[46] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[47] A. Graves, A. Mohamed, and G. Hinton, "Speech Recognition With Deep Recurrent Neural Networks," *Icassp*, no. 3, pp. 6645–6649, 2013, doi: 10.1109/ICASSP.2013.6638947.

Kamaldeep Kaur is presently working as Assistant Professor in Department of Computer Science and Engineering, at Guru Nanak Dev Engineering College, Ludhiana (India). She completed her B.Tech. in Computer Science & Engineering from GNDEC, Ludhiana. Then completed her Masters in Engineering from Panjab University, Chandigarh, in Computer Science & Engineering. She was a gold medalist at graduation and post-graduation levels. Her field of research is Natural Language Processing. She had developed the first Topic Tracking System for Punjabi language. She is currently doing her research on Recognition of emotions from Punjabi Speech, and has published various research papers in reputed journals.

Dr. Parminder Singh is presently working as Professor in Department of Computer Science and Engineering at Guru Nanak Dev Engineering College, Ludhiana (India). He is B.Tech., M.Tech. and Ph.D. in Computer Science and Engineering. He has more than 23 years of teaching experience. His field of interest is Natural Language Processing, particularly Text-to-Speech Synthesis. He has developed first Text-to-Speech synthesis system for the Punjabi language. He has guided more than 51 post-graduate students for research work and has published about 75 research papers in international and national journals and conferences. He is reviewer of different reputed journals and conferences.