Artificial Neural Network based Emotion Classification and Recognition from Speech

Mudasser Iqbal[1]*, Syed Ali Raza[2]
Department of Computer Science
Institute of Southern Punjab (ISP) Multan, Pakistan[1]
Government College University (GCU)
Lahore, Pakistan[2]

Furqan Majeed[4]
Department of Computer Science
Institute of Southern Punjab (ISP)
Multan, Pakistan

Muhammad Abid[3]
Department of Computer Science
Govt. Postgraduate College
Layyah, Pakistan

Ans Ali Hussain[5]
Department of Computer Science
University of Agriculture
Faisalabad, Pakistan

***Abstract*—Emotion recognition from speech signals is still a challenging task. Hence, proposing an efficient and accurate technique for speech-based emotion recognition is also an important task. This study is focused on four basic human emotions (sad, angry, happy, and normal) recognition using an artificial neural network that can be detected through vocal expressions resulting in more efficient and productive machine behaviors. An effective model based on a Bayesian regularized artificial neural network (BRANN) is proposed in this study for speech-based emotion recognition. The experiments are conducted on a well-known Berlin database having 1470 speech samples carrying basic emotions with 500 samples of angry emotions, 300 samples of happy emotions, 350 samples of a neutral state, and 320 samples of sad emotions. The four features Frequency, Pitch, Amplitude, and formant of speech is used to recognize four basic emotions from speech. The performance of the proposed methodology is compared with the performance of state-of-the-art methodologies used for emotion recognition from speech. The proposed methodology achieved 95% accuracy of emotion recognition which is highest as compared to other states of the art techniques in the relevant domain.**

***Keywords*—Emotion States; ANN; BR; BRANN; emotion classifier; speech emotion recognition**

I. Introduction

In the modern age of technology, Emotion recognition from speech is a hot research topic in the field of speech signal processing [1]. The objective of emotion recognition from the speech is to make Human-Computer Interaction (HCI) more natural and human friendly [2] [3]. However, there is a gap between humans and computers, that computers act logically and humans act logically as well as emotionally. This gap makes the computers less compatible with humans. To reduce this gap, to make the interface easier to use, and to develop more understanding between humans and machines, it is necessary for the machines to understand and act according to human emotions.

There are some well-known emotions like anger, happy, sad, and neutral [4] that can affect speech signals. There are different features of sound like frequency, amplitude, pitch, and format that has been in use for emotion recognition from speech signal [5]. Researchers have been used different approaches such as wavelet-based feature, Mel frequency cepstral coefficient (MFCC), and linear prediction cepstral coefficient (LPCC) [6]. However, Mel-Frequency cepstral coefficients (MFCC) is the most used feature for emotion recognition from speech [7]. The objective of this research is to review emotion recognition techniques with recognition accuracy and to produce an emotion recognition system to recognize four basic emotions anger, sad, happy, and neutral using speech signals.

Emotions recognition can be elaborated as the extraction of emotion from speech signals to make human-computer interaction (HCI) more efficient and convenient [8]. Various techniques are in use for emotion recognition that includes feature selection and extraction and then applying classifier. Hidden Markov model (HMM), Gaussian mixture model (GMM), Support vector machine (SVM), and Artificial neural network (ANN) are the classifiers that can be used for emotion recognition [9], however, the success rate of emotion recognition depends upon features and training algorithm. Training and testing are the two phases in the supervised machine learning approach. During the training phase, the neural network is created and trained by providing a large number of training data and training algorithms with a random number of neurons [10]. During the testing phase, the test data set is provided and its features are extracted and matched with the trained model [10]. Accurate speech recognition depends upon the training algorithm and machine learning accuracy.

II. Related Work

Spectral feature-based emotions classification using Gaussian Mixture Models (GMMs) a technique of acoustic-phonetic approach was proposed by [11]. A private database for emotional speech is used where specific content was recorded by a male and a female actor voices with four basic emotions anger, happy, sad, and neutral in acted speech corpus. For the real speech corpus, samples of single male

voices were taken. The four emotions: sad, happy, anger, and neutral were selected for the classification of real speech corpus. 75% of samples were used for training and the rest 25% were used for testing. Gaussian Mixture Model (GMM) was used for the emotions classification because it was considered better for performing clustering. It produced 85% average correct classification for the real speech samples in case of separate gender voices. However, it was observed that when the voices of two males or two females or male and female were mixed, the performance of the purposed system decreased which was the sign that the system was speaker-dependent as well as gender-dependent.

Emotion recognition from the human speech is still a complex task because of the ambiguity in classifying the natural and acted emotion in a human speech explained by [12]. They [12] applied several machine learning techniques including K-nearest neighbor (KNN) and artificial neural network (ANN) to create recognition agents. The agent was able to recognize five emotional states including normal, happiness, anger, sadness, and fear with the accuracy of 55-75%, 60-70%, 70-80%, 75-85%, and 35-55%, respectively. The author in [12] proposed a modified MFCC approach by separating male and female data into separate frames to increase the accuracy of emotion recognition. By applying this modified MFCC approach, before breaking down the speech sample into frames, classification of the speech sample of male and female was done and then compared with the database. So, the overall success rate of the standard approach was 54.54% whereas the modified MFCC approach produced a 63.63% overall successive rate.

Assigning human-like properties e.g. watching, interpreting, and producing effective features is referred to as affective computing which enables the machine to recognize human emotion and to react accordingly. To obtain this goal, a study is conducted by [13]. They have divided the emotion recognition system into feature extraction and feature classification phases. An acted database which is the Berlin Database of emotional speech of 5 male and 5 female actors is used. Three categories of the audio file which were sad, happiness, and neutral states of person are used to reveal this study. At the feature selection phase, a combination of energy, skewness, and MFFC is used. They have used a total of 354 instances with 121 instances of happiness, 117 instances of sadness, and the rest 116 instances of neutral states to conduct this study. A variant number of neurons at the hidden layer of vanilla Artificial Neural Network is used to classify the selected features. The experiment results showed 72% accuracy with 45 neurons at the hidden layer.

Many classification methods have been applied for emotion recognition from speech such as Support Vector Machine (SVM), Hidden Markov Model (HMM), K-Nearest Neighbors (KNN), and Artificial Neural Network (ANN) [14].

A classification technique (optimized Support Vector Machine) for emotion recognition from the speech was proposed by [15]. In that work, they used four emotions: anger, happiness, sadness, and surprise. For the experimental purpose, the voice of a group of non-professionals was recorded having four basic emotions. The experiment was speaker and gender independent, therefore, 5 males and 9 female speakers with age group about 20 took part in recording the emotional speeches. Fast Fourier Transform (FFT) was used to transform every frame of N samples into a frequency spectrum. After that, an optimized support vector machine (SVM) was established to perform two-class pattern classification. Four kernel functions which were Linear, Polynomial, Radial basis function (RBF), and Sigmoid function to evaluate the performance of optimized support vector machine were used for emotion recognition from speech. According to the [15] research, the Radial basis function (RBF) produced 71.89% accuracy at the training set and 88.75% accuracy at the testing set.

A study on emotion recognition from speech was directed by [16] to design and propose an efficient classifier for emotion recognition from speech. Their work for the recognition of emotions from the speech was based on a discrete emotion classification system. They constructed a model for emotion recognition from speech based on support vector machine (SVM) and artificial neural network (ANN) respectively. The effect of emotion recognition of feature reduction was compared respectively for both support vector machine (SVM) and artificial neural network (ANN). According to the [16], experimental results showed that artificial neural networks (ANN) produced 75 % testing accuracy and support vector machine (SVM) produced 85% accuracy. It was proposed that the support vector machine (SVM) is slightly better than the artificial neural network (ANN).

III. Proposed Methodology

The importance of emotion recognition from an audio signal is increasing to make human-machine interaction more efficient. It is based on depth analysis of speech signal, feature extraction that contains emotional information, and taking appropriate pattern recognition techniques to recognize the emotional state. The first step is to extract features like pitch, frequency, timbre, and amplitude from the audio signal to train a neural network. The second step is to give an audio input to a machine to compare with the store information obtained at the training time [17]. Speech emotion recognition system is shown in below-given Fig. 1 which includes a sensor module to receive speech data as input, a pre-processing module to remove noise from the data, and a feature extraction module to extract features (frequency, formant, pitch, and amplitude) from the processed data and a training module (Neural Network) with training function.

A. Preprocessing of Speech Signal

Before converting the auditory signals into numerical values, there exist some unnecessary data that get mixed into the original data and affect the original values. The noise of the environment is one of the major factors that get the mix into the actual voice signals and changes their characteristics [18]. Therefore, it is necessary to remove the noise of the given data. This feature extraction module is further subdivided into two modules are Noise removal system and Feature extraction. The job of this module is to remove the noise from the actual data. Because removal of noise is important to get the correct results [11].

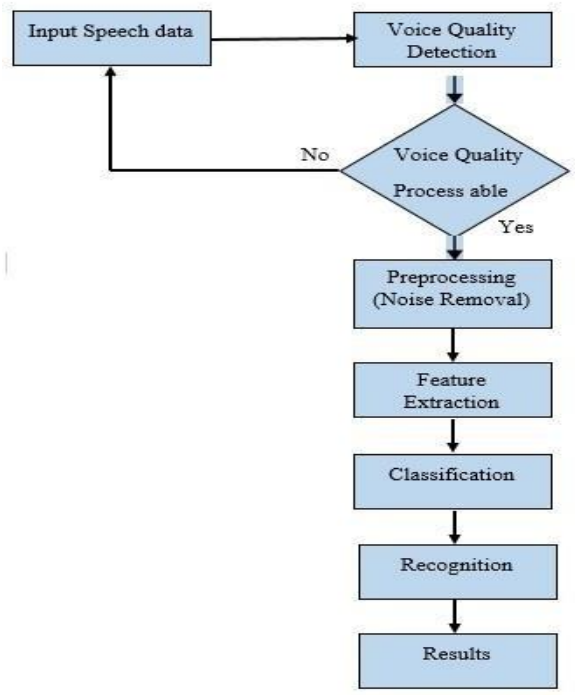

Fig. 1. Flow of Proposed Emotion Recognition System.

B. Feature Extraction

After removing the noise, the task is to extract the features from the data. A sound signal has a lot of features but after the literature review, it came to know that the four features (frequency, amplitude, pitch, and formant) can play a vital role in emotion detection. These four features are considered for emotion detection. The characteristics of these features are given below.

1) FREQUENCY: Frequency, additionally called wave recurrence, is an estimation of the absolute number of vibrations or motions made inside a specific unit of time [19]. As the number of waves per unit time increases the frequency increases which can be written as:

$$F = 1/T \quad (1)$$

The Time interval in which a wave completes its one cycle is called the time period which is denoted by T and the number of cycles in one second is called frequency which is the reciprocal of the time period as described in equation (1).

2) FORMANT: A formant plays a very vital role in emotion recognition from speech and is a widely used feature for emotion recognition [19]. There are many formants and each formant has a different frequency, roughly assumed one in each 1000Hz band. In the vocal set, every formant corresponds to a specific character. Formants are displayed as dark bands and can be seen very clearly in a wideband spectrogram. As much as the formant is darker, it can be reproduced in the spectrogram and has more energy (stronger it is or the more audible it is). The equation for formant is:

$$L = C/4F \quad (2)$$

Where "C" is the speech of sound (340.29 m/s) and "F" is the first formant of the frequency.

3) PITCH: Pitch is also a widely used feature in emotion recognition and it is a perceptual property of sounds that permits their ordering on a Frequency related scale or pitch is the nature of frequency that makes it conceivable to acknowledge sounds as" higher" and" lower". Pitch is considered as the hear-able quality of sound recurrence as per which sounds can be ordered on a scale from low to high [20] [21]. The term" high" pitch is considered as exceptionally quick wavering, and the term" low" pitch portrays the more slow swaying.

$$\text{Pitch} \propto 1/T \quad (3)$$

As described in equation (3), pitch directly depends upon frequency. This shows that the greater is the frequency of a sound, the greater will be its pitch.

4) AMPLITUDE: Amplitudes can be defined either as instantaneous values or mostly as peak values. Amplitude is referred to as the fluctuation or displacement of a wave. With sound waves, the amplitude is the loudness of the sound and the extent to which air particles are displaced [22]. Also analyzed the feature of Amplitude, Pitch, and Formant in their paper.

$$X = X_m \sin(\omega t+\varphi) \quad (4)$$

In equation (4), X is the instantaneous value of displacement of a wave from the mean position and X_m is amplitude where $(\omega t+\varphi)$ is called the phase of the motion.

C. Bayesian Regularization

Bayesian Regularization (BR) typically consumes less memory and more time but it has a good generalization to process difficult and noisy datasets. Training of Bayesian Regularization algorithm automatically stops according to the adaptive weight minimization regularization. Bayesian Regularization algorithm is more accurate as compare to Levenberg Marquardt (LM) and Scale Conjugate Gradient (SCG) [23]. It prevents overtraining and provides an efficient criterion for stopping the training process. This efficiency of the BR algorithm makes it a more adaptive algorithm for training a neural network to perform emotion recognition from speech. We can write Bayesian theorem as:

$$P\left(\frac{X}{Y}\right) = P\left(\frac{X}{Y}\right)\left(\frac{P(X)}{P(Y)}\right) \quad (5)$$

Bayesian regularized artificial neural network (BRANNs) is stronger than standard back-propagation networks and can decrease or expel the requirement for long cross-validation. It changes over a nonlinear regression into a well-structured linear regression as:

$$Y = B0X / (B1 + X) \quad \text{Nonlinear regression} \quad (6)$$

$$Y = a + bX + u \quad \text{Linear regression} \quad (7)$$

Where Y is the variable that is to be predicted and X is used to predict Y as well as a is the intercept, b is the slope and u is the regression residual. Therefore, Bayesian

regularized artificial neural network (BRANN) can be expressed in the equation as:

$$D(Z) = \beta \sum_{x=1}^{N_C} [Y_i - f(B_i)]^2 + \alpha \sum_{j=1}^{N_Z} Z_j^2 \quad (8)$$

Where Nz is the number of weights. Given starting values of the hyperparameters α and β the cost function, D (Z), is reduced regarding the weights Z. A re-estimate of α and β is made by amplifying the proof. The initial probability over the weight, Z, can be composed as:

$$P(Z \mid \alpha, H) = \frac{1}{C_Z} exp(-\alpha E_Z) \quad (9)$$

$$E_Z = \sum_{x=1}^{N_Z} Z_j^2 \; being\ the\ error\ of\ the\ weights. \quad (10)$$

With

Similarly, the probability of errors can be described as

$$P(E)|Z, \beta, H) = \frac{1}{W_E(\beta)} exp(-\beta E_D) \quad (11)$$

With

$$E_Z = \sum_{j=1}^{N_Z} [y_i - f(X_j)]^2 \; being\ the\ error\ of\ the\ data \quad (12)$$

The most common function used for Bayesian estimation is the mean square error (MSE) which can be defined as:

$$\text{MSE} = E[\left(\hat{\theta}\ (x) - \theta\right)^2] \quad (13)$$

Where the expectation is taken over the joint distribution of θ and x.

D. Artificial Neural Network

Artificial Neural Network (ANN) is a scientific model, which works similar to the human neural system [24] [25]. A two-layer feed-forward network is trained with a Bayesian Regularization algorithm. A feed-forward neural network, which consists of multiple numbers of neurons as a processing unit and is a biologically inspired classification algorithm [26]. Each processing unit in a layer is connected with all processing units of the previous layer. Speech-based emotion classification is examined in three main steps: Data preprocessing and noise removal, feature extraction, emotion classification and recognition, and then output.

Artificial Neural Network (ANN) is trained using emotional speech data taken from the Berlin Database of Emotional Speech of the Technical University of Berlin, Institute of Speech and Communication, department of communication science [27] [28]. This Database Consists of 3300 utterances spoken by the different actors in happy, angry, sad, disgust, boredom, fear, and normal ways. As this work aims at 4 emotions States as Angry, Sad, Normal, and Happy, therefore, four utterances are filtered and used the Feature Extraction Module to generate the Dataset to train Bayesian Regularized Artificial Neural Network (BRANN). A total of 1470 samples of Emotions (500 samples of angry emotion, 300 samples for happiness, 350 samples for natural, and 320 samples for sorrow) are taken to train the proposed ANN. Emotion recognition from the speech is done in three main phases. In the First Phase, Speech data is received and noise is removed to get better results and more accurate feature extraction. In the second Phase, features (Frequency, Pitch, Amplitude, and Formant) are extracted. After extracting the features from auditory data, these features are fed to the neural network. In the third Phase, emotional speech is classified and passed to the recognition phase that produced the final output. As mentioned earlier network consists of seventy-eight layers including seventy hidden layers 4 input layers and 4 output layers. The output of the ith layer becomes the input for the (i+1)the layer. Every time biases are added to the input. After every iteration, the input values change for every neuron because new weights are adjusted after every iteration. This neural network contains seventy-eight layers of neurons including 4 input layers, 4 output layers, and seventy hidden fully connected layers as shown in below Fig. 2.

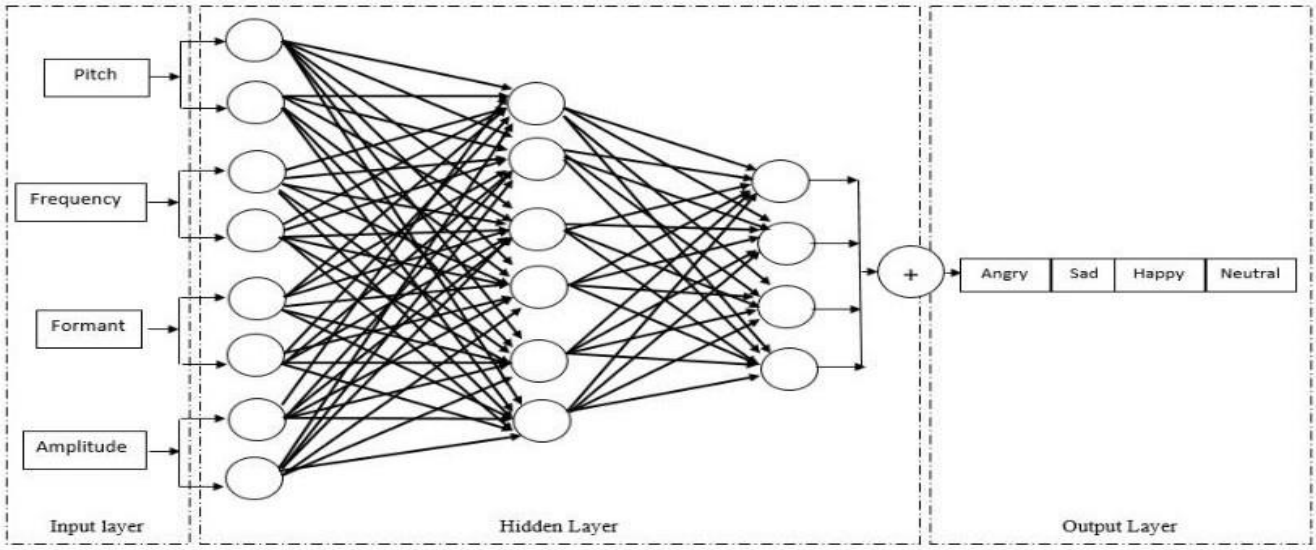

Fig. 2. Artificial Neural Network Framework for Speech-based Emotion Recognition.

IV. Results and Experiments

It has been observed that the emotions of a human change the features of its voice with a specific pattern. For example, when a human gets angry his intensity of voice increases so the frequency increases and in most of the cases, he goes louder to show his anger so the amplitude of his voice signal also increases. Similarly, when someone becomes sad his intensity and amplitude of voice signal also get low. The experiments were executed in Matlab 2017r2, Dell i7 with properties of 8 GB RAM, Windows 10 operating system. The proposed system is trained and tested with 1470 speech datasets taken from the Berlin database carrying four basic emotions. After training the network, testing is done by providing emotional speeches to the network, which are successfully recognized. The following figures show the testing result of the proposed network based on four features (frequency, pitch, formant, and amplitude. The experimental results of speech-based emotion recognition through Bayesian Regularized Artificial Neural Network on the Berlin database of emotional datasets have produced an efficient performance as compared to the state-of-the-art techniques. The proposed methodology is executed in three phases which are briefly explained in Section III.

In the first phase, preprocessing like noise removal is performed on the audio dataset and passed to the next phase. In the second phase, which is also an important phase of the proposed system, features are extracted from the speech data. After the feature (frequency, pitch, amplitude, and formant) extraction the speech data is passed to the third phase, and classification is done using ANN. We have used a 1470 emotional speech dataset to train the Artificial Neural Network with the Bayesian Regularized algorithm due to its less memory and time consumption and then tested with 700 emotional datasets. Results are shown in below-given Fig. 3.

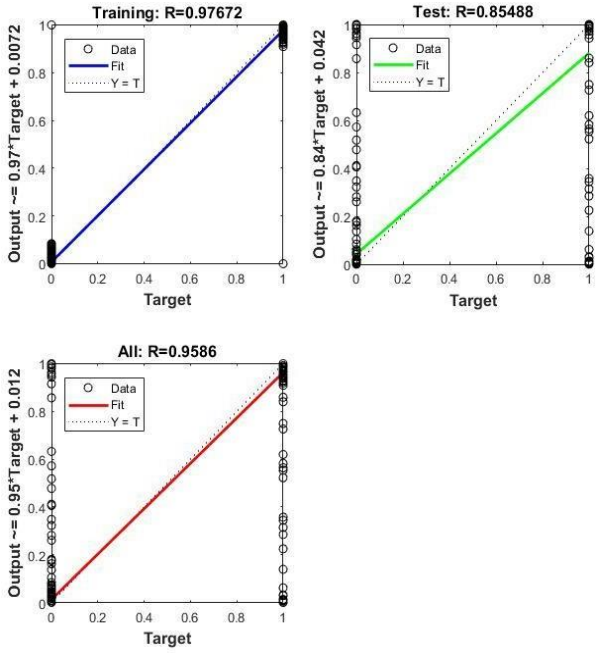

Fig. 3. Regression Plot for Training and Testing of Artificial Neural Network having 70 Neurons at Hidden Layer with BR Algorithm.

As shown in Fig. 3, optimum results are obtained with 70 neurons at the hidden layer. The proposed technique produced 97.6 % training accuracy, 85.4% testing accuracy, and 95.8% overall accuracy of emotion recognition from the speech which is superior when performing a comparison with the state-of-the-art techniques of emotion recognition from speech. Table I and Fig. 4 shows the confusion matrix of the proposed system. The proposed system performance is also compared with the state-of-the-art techniques in "Table II" which shows that the proposed system produced more accurate results than existing techniques.

TABLE I. Performance Comparison of the Proposed System with a Variant Number of Neurons at the Hidden Layer of BRANN

Training Algorithm: **Bayesian Regularization**				
Number of Neurons	**Training Rate**	**Testing Rate**	**Overall Accuracy**	**Best Performance (Error Rate %)**
10	88%	84%	87%	0.04
20	93%	82%	91%	0.02
30	95%	82%	93%	0.01
40	96%	80%	94%	0.01
70	97%	85%	95%	0.007
80	97%	72%	93%	0.02

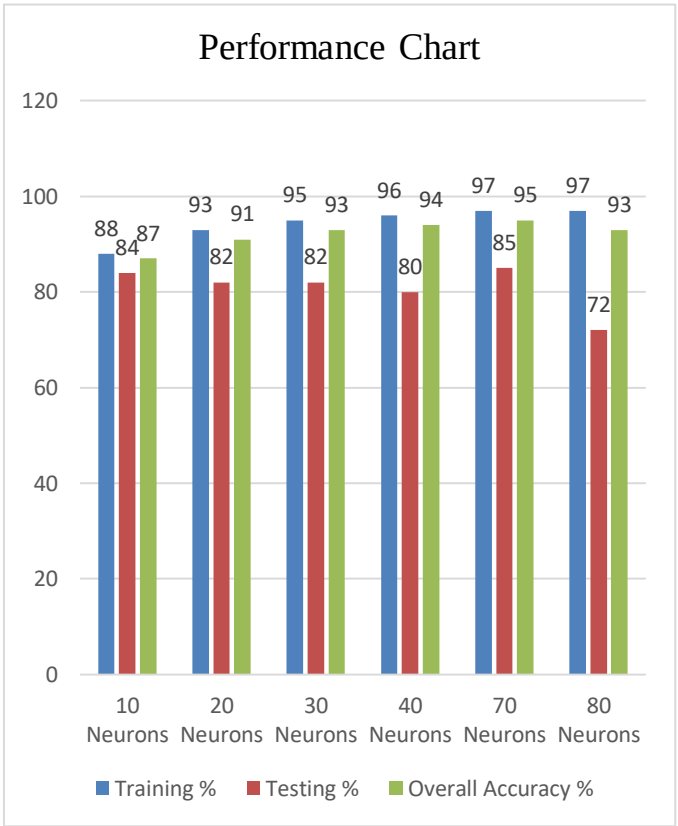

Fig. 4. Performance Comparison of Proposed Technique.

As shown in Table I, and Fig. 4, accuracy in results increased with increasing the number of neurons at the hidden layer but with 80 neurons at the hidden layer, results are

significantly decreased therefore the training process is stopped. Optimum results are obtained with 70 number of neurons at the hidden layer. The performance of the proposed system is also compared with state-of-the-art methodologies and it is demonstrated that the performance of the proposed technique is comparatively better and efficient. The following figures show the results of testing the proposed technique by providing an emotional dataset.

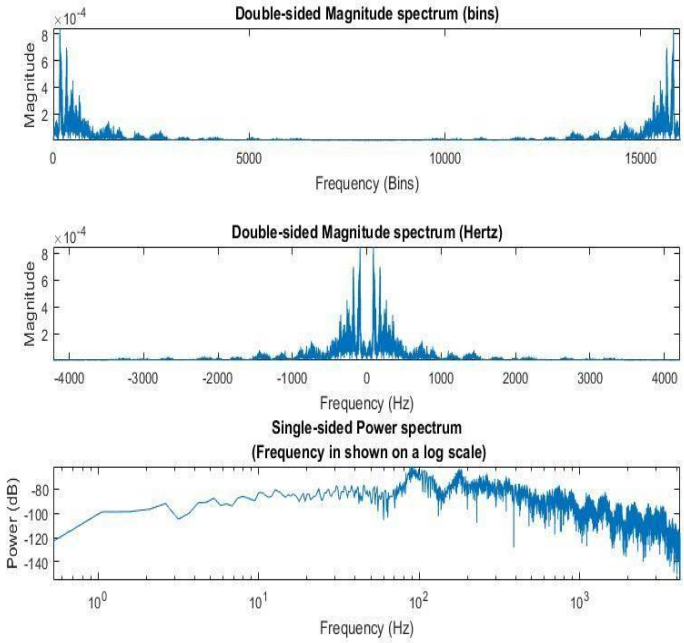

Fig. 5. The Plot of Frequency Feature for Speech Data Containing Angry Emotion.

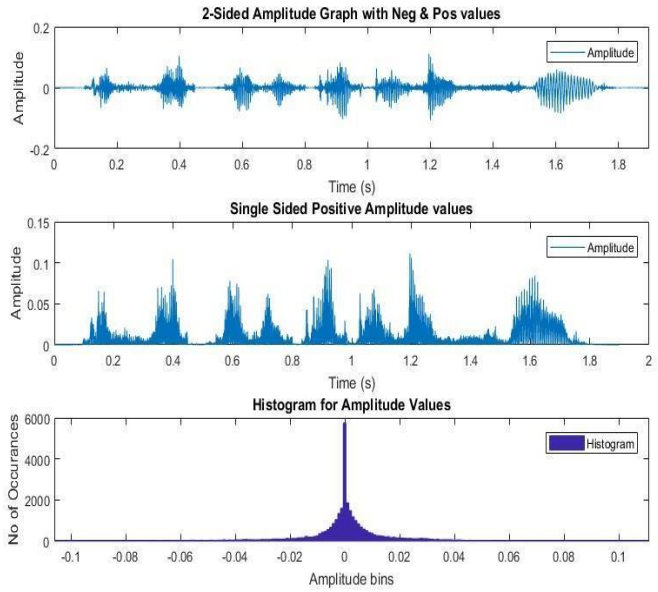

Fig. 6. The Plot of Amplitude Feature for Speech Data Containing Angry Emotion.

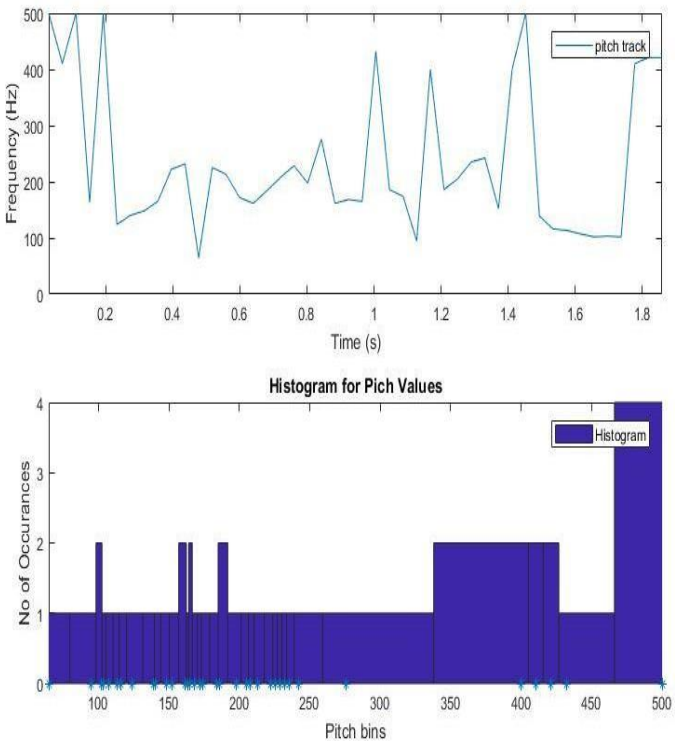

Fig. 7. The Plot of Pitch Feature for Speech Data Containing Angry Emotion.

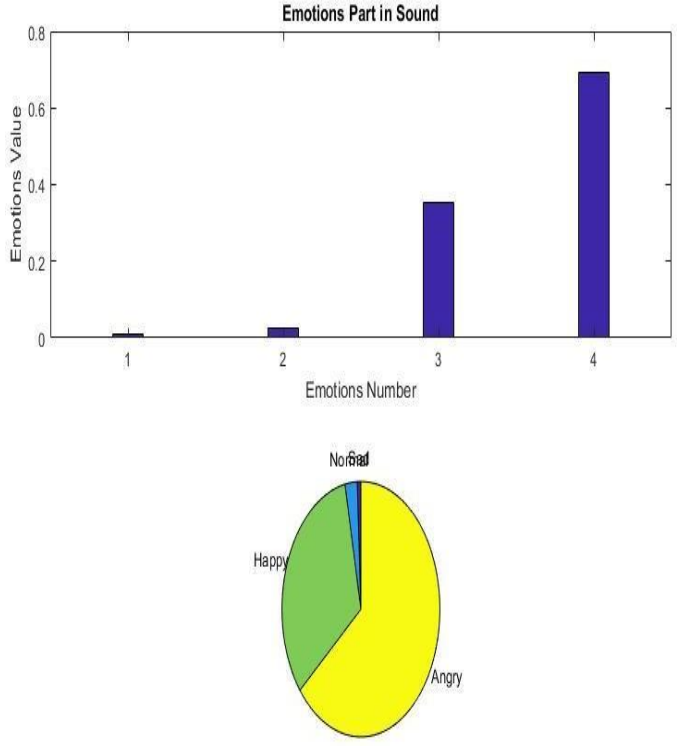

Fig. 8. Angry Emotion Recognized Successfully from Speech Dataset.

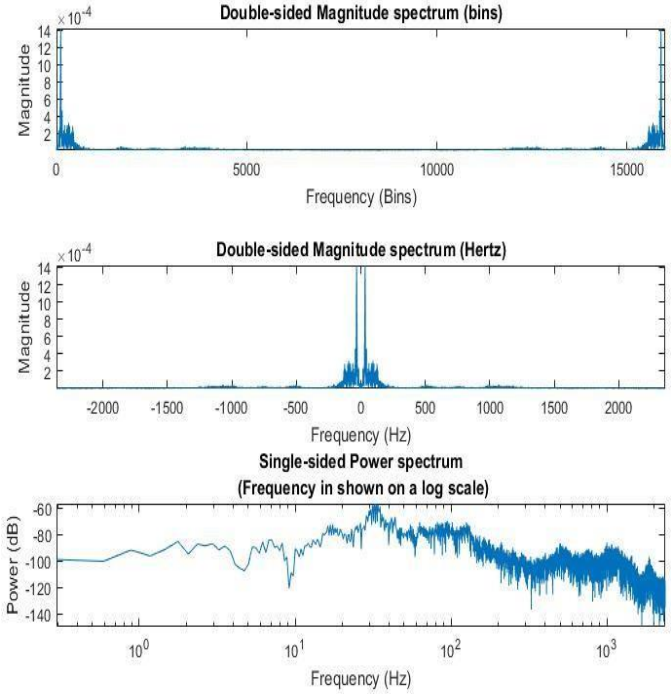

Fig. 9. The Plot of Frequency Feature for Recognizing Happy Emotion.

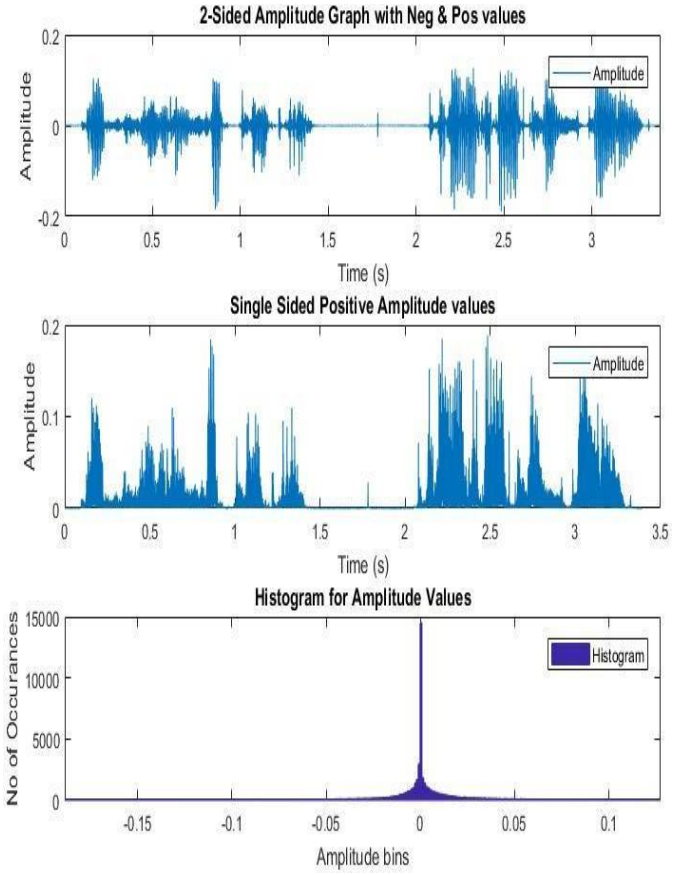

Fig. 10. The Plot of Amplitude Feature for Speech Data Containing Happy Emotion.

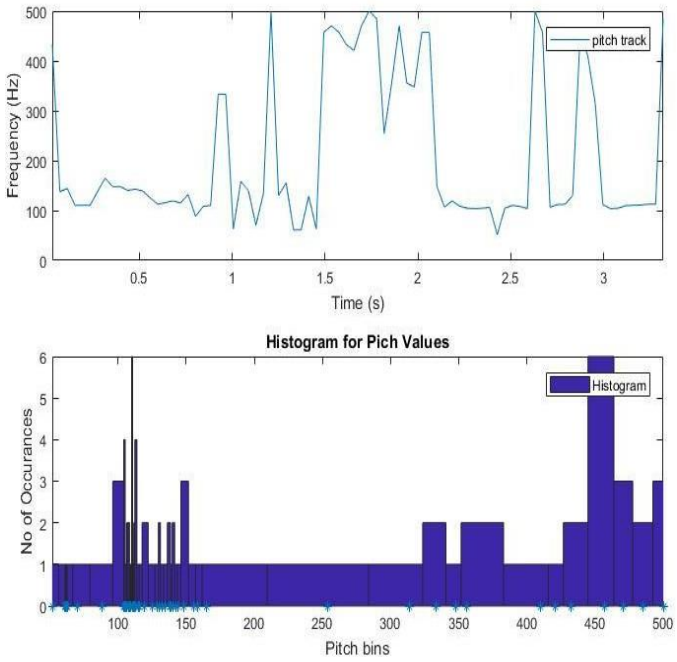

Fig. 11. The Plot of Pitch Feature for Speech Data Containing Happy Emotion.

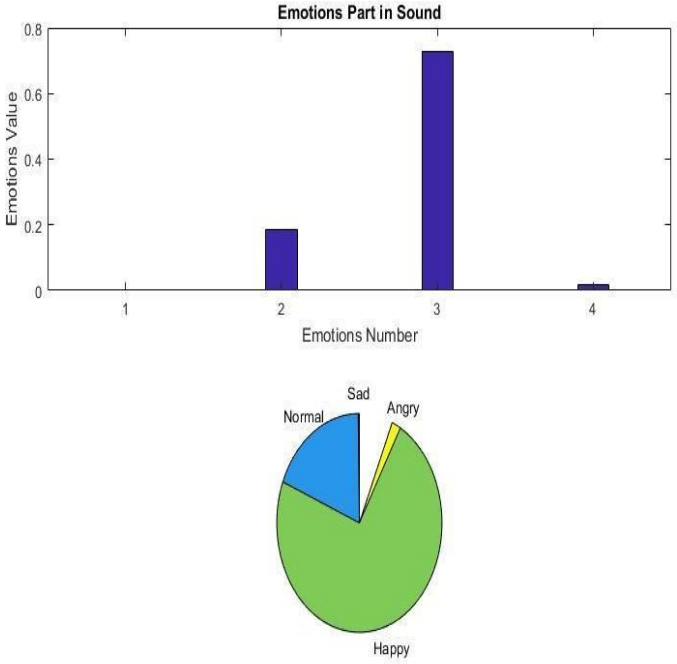

Fig. 12. Happy Emotion Recognized Successfully from Speech Dataset.

As shown in "Fig. 5" to "Fig. 12", the proposed system has produced accurate results basis on the four features (frequency, pitch, formant, and amplitude). Because formant has only values and can't be displayed in a plot, only frequency, amplitude, and pitch plots as well as recognition graph is displayed in figures. Sad and Neutral emotions are also recognized successfully which can be shown below given "Fig. 13" to "Fig. 20".

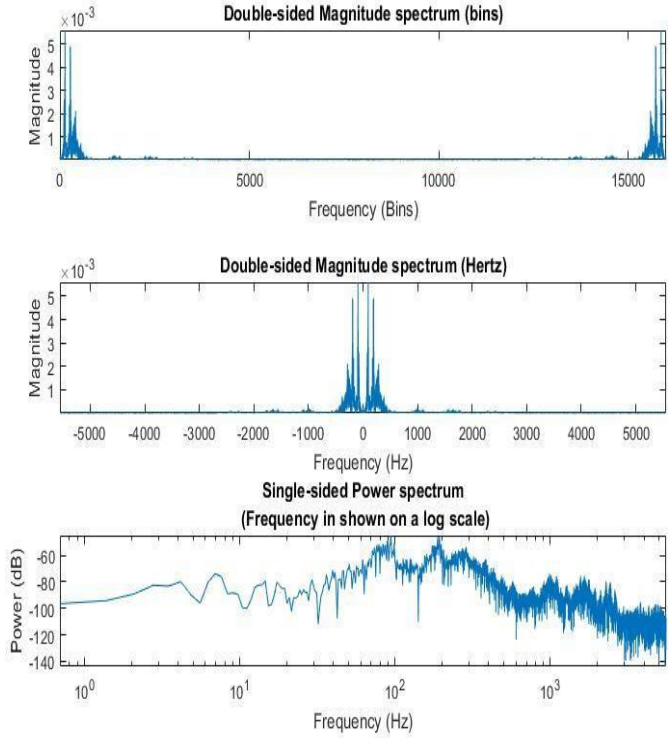

Fig. 13. The Plot of Frequency Feature for Recognizing Neutral Emotion.

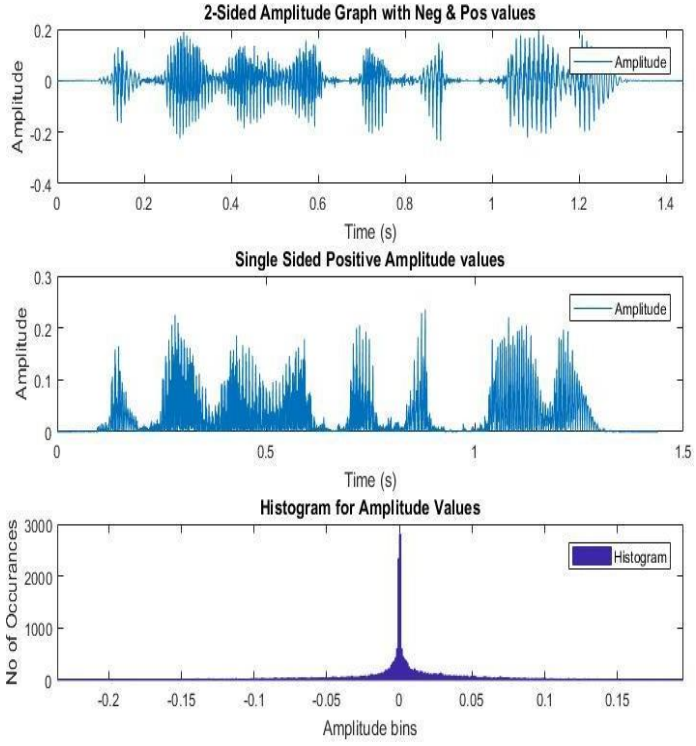

Fig. 14. Amplitude Feature for Speech Data Containing Neutral Emotion.

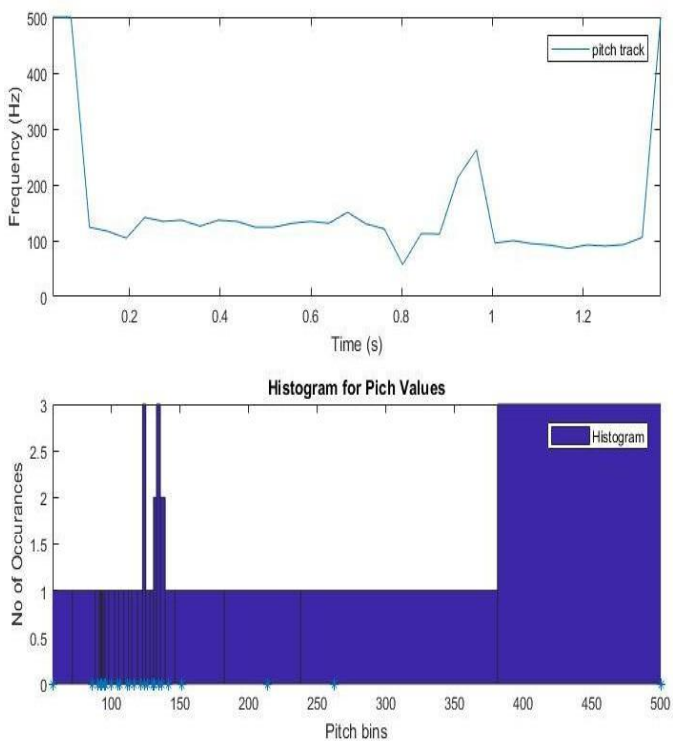

Fig. 15. The Plot of Pitch Feature for Speech Data Containing Happy Emotion.

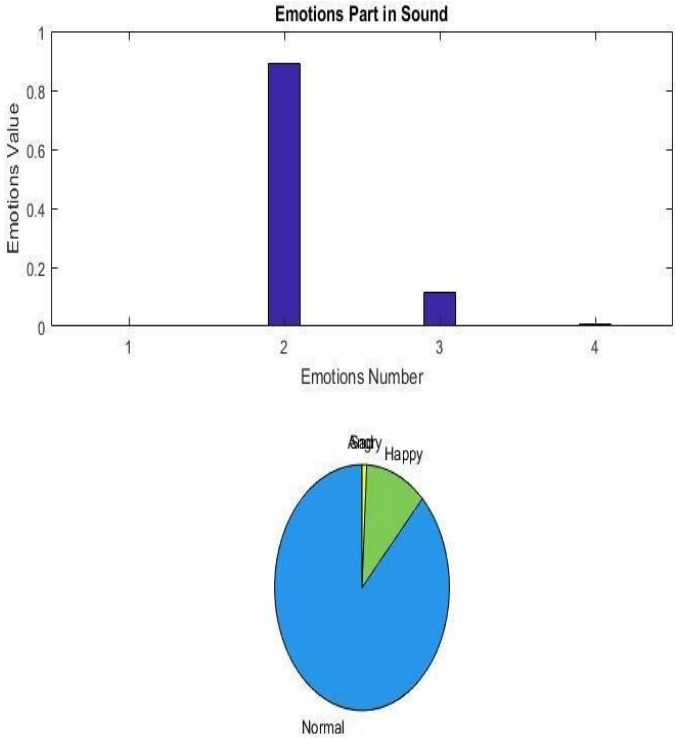

Fig. 16. Neutral Emotion Recognized Successfully from Speech Dataset.

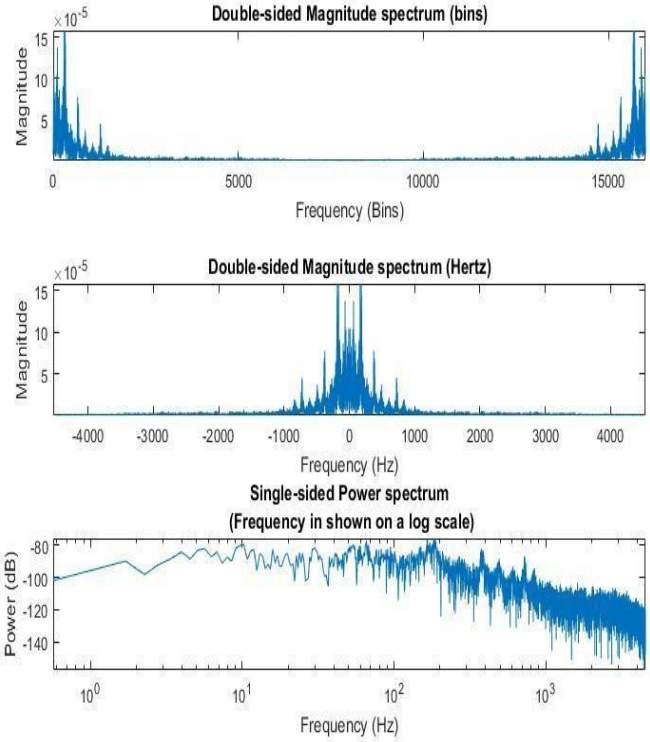

Fig. 17. The Plot of Frequency Feature for Recognizing Sad Emotion.

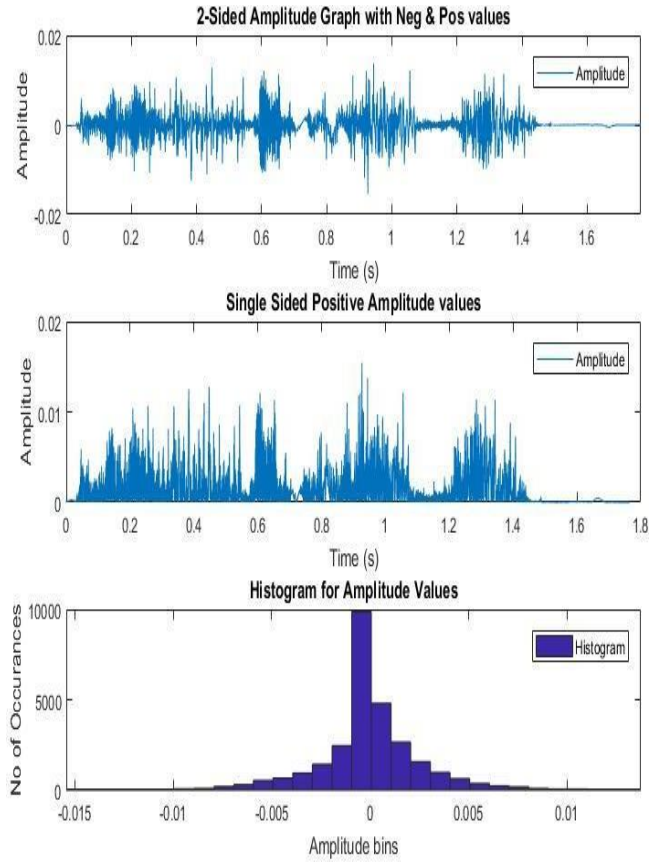

Fig. 18. Amplitude Feature for Speech Data Containing Sad Emotion.

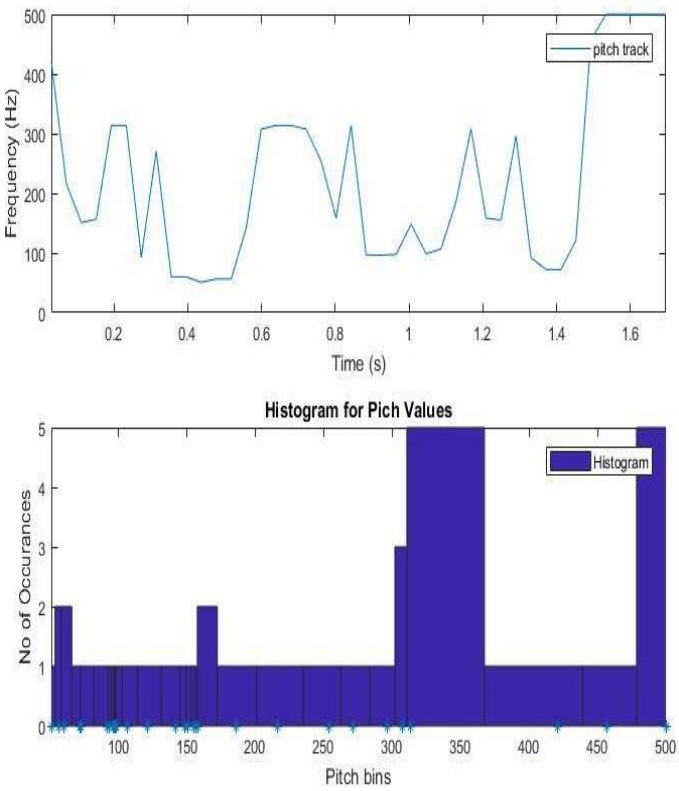

Fig. 19. The Plot of Pitch Feature for Speech Data Containing Sad Emotion.

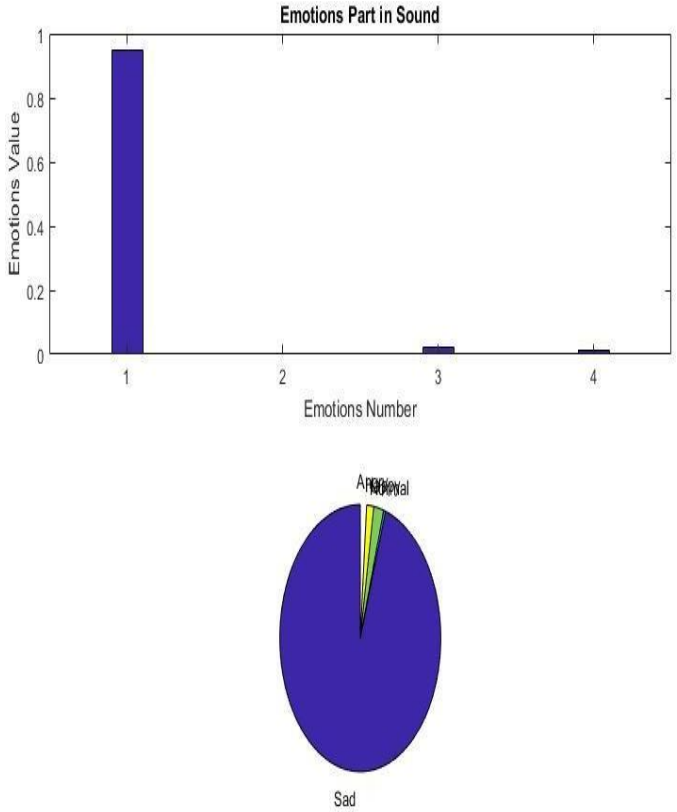

Fig. 20. Sad Emotion Recognized Successfully from Speech Dataset.

TABLE II. COMPARISON OF THE PROPOSED SYSTEM WITH STATE-OF-THE-ART TECHNIQUES

Reference and Year	Methodology	Performance
[11] 2012	Gaussian Mixture Model (GMM)	85% average correct classification for the real speech samples with separate gender voices
[12] 2013	K-nearest neighbor (KNN), artificial neural network (ANN) with modified MFCC approach	54.54% accuracy whereas the modified MFCC approach produced 63.63%.
[13] 2018	vanilla Artificial Neural Network (ANN) with berlin emotion database	72% accuracy with 45 neurons at the hidden layer.
[15] 2016	optimized support vector machine (SVM) with different kernel functions	Radial basis function (RBF) produced 71.89% training accuracy and 88.75% testing accuracy.
[16] 2018	support vector machine (SVM) and artificial neural network (ANN) respectively	artificial neural network (ANN) produced 75 %testing accuracy and support vector machine (SVM) produced 85% accuracy
Proposed System	ANN with Bayesian Regularization	97% training accuracy,85% testing accuracy, and 95% overallaccuracy

V. CONCLUSION

The proposed architecture is based on Artificial Neural Network with a Bayesian Regularization algorithm. An artificial agent is created and trained with Berlin's emotional database having four emotions: angry, sad, neutral, and happiness. The proposed system is tested with 10 neurons at the hidden layer initially, which are increased step by step. The proposed architecture with 70 neurons at the hidden layer, produced 97% training accuracy, 85% testing accuracy, and 95% overall accuracy for the four basic (Angry, Happy, Neutral, and Sad) emotions. This work contributed almost a 5% gain in training accuracy and a 3% gain in testing accuracy using Bayesian Regularization Artificial Neural Network (BRANN). The proposed architecture results are also compared with the state-of-the-art techniques for speech-based emotion recognition. The comparison results show that the proposed system recognizes four basic emotions from speech more accurately than state-of-the-art techniques. Moreover, higher accuracy can be obtained using the combination of more features.

REFERENCES

[1] P. Song, W. Zheng, S. Ou, X. Zhang, Y. Jin, J. Liu and Y. Yu, "Cross-corpus speech emotion recognition based on transfer non-negative matrix factorization," Speech Communication, vol. 83, no. 2016, pp. 34-41, 2016.

[2] A. Milton, S. S. Roy and S. T. Selvi, "SVM Scheme for Speech Emotion Recognition using MFCC Feature," International Journal of Computer Applications, vol. 69, no. 9, pp. 34-39, 2013.

[3] Y. Wang and L. Guan, "Recognizing Human Emotional State From Audiovisual Signals," IEEE TRANSACTIONS ON MULTIMEDIA, vol. 10, no. 5, pp. 936-945, 2008.

[4] A. Joshi and R. Kaur, "A Study of Speech Emotion Recognition Methods," International Journal of Computer Science and Mobile Computing, vol. 2, no. 4, pp. 28-31, 2013.

[5] K. M. Kudiri, G. K. Verma and B. Gohel, "Relative Amplitude based Features for Emotion Detection from Speech," in IEEE International Conference on Signal and Image Processing, Chennai, 2010.

[6] X. MAO, B. ZHANG and Y. LUO, "SPEECH EMOTION RECOGNITION BASED ON A HYBRID OF HMM/ANN," in Proceedings of the 7th WSEAS International Conference on Applied Informatics and Communications, Athens, 2007.

[7] S. Demircan and H. Kahramanlı, "Feature Extraction from Speech Data for Emotion Recognition," Journal of Advances in Computer Networks, vol. 2, no. 1, pp. 28-30, 2014.

[8] S. N. R. A, R. P. Gadhe, R. R. Deshmukh, V. B. Waghmare and P. P. Shrishrimal, "Automatic emotion recognition from speech signals: A Review," International Journal of Scientific & Engineering Research, vol. 6, no. 4, pp. 636-638, 2015.

[9] K. Wang, N. An, B. N. Li and Y. Zhang, "Speech Emotion Recognition Using Fourier Parameters," IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, vol. 6, no. 1, pp. 69-74, 2015.

[10] S. K.Gaikwad, B. W.Gawali and P. Yannawar, "A Review on Speech Recognition Technique," International Journal of Computer Applications, vol. 10, no. 3, pp. 16-23, 2010.

[11] K. S. Rao, T. P. Kumar, K. Anusha, B. Leela, I. Bhavana and S. V. Gowtham, "Emotion Recognition from Speech," International Journal of Computer Science and Information Technologies, vol. 3, no. 2, pp. 3603-3607, 2012.

[12] A. Sapra, N. Panwar and S. Panwar, "Emotion Recognition from Speech," International Journal of Emerging Technology and Advanced Engineering, vol. 3, no. 2, pp. 341-345, 2013.

[13] A. K. Komal Rajvanshi, "An Efficient Approach for Emotion Detection from Speech Using Neural Networks," International Journal for Research in Applied Science & Engineering Technology, vol. 6, no. 5, pp. 1062-1065, 2018.

[14] R. D.Shah and D. Anil.C.Suthar, "Speech Emotion Recognition Based on SVM Using MATLAB," International Journal of Innovative Research in Computer and Communication Engineering, vol. 4, no. 3, pp. 2916-2920, 2016.

[15] B. Yu, H. Li and C. Fang, "Speech Emotion Recognition based on Optimized Support Vector Machine," JOURNAL OF SOFTWARE, vol. 7, no. 12, pp. 2726-2732, 2012.

[16] X. Ke, Y. Zhu, L. Wen and W. Zhang, "Speech Emotion Recognition Based on SVM and ANN," International Journal of Machine Learning and Computing, vol. 8, no. 3, pp. 198-201, 2018.

[17] A. Joshi and R. Kaur, "A Study of Speech Emotion Recognition Methods," International Journal of Computer Science and Mobile Computing, vol. 2, no. 4, p. 28–31, 2013.

[18] N. J. Gogoi and J. Kalita, "Emotion Recognition from Acted Assamese Speech," International Journal of Innovative Research in Science, Engineering and Technology, vol. 4, no. 6, pp. 4116-4121, 2015.

[19] S. Bhadra, U. Sharma and A. Choudhury, "Study on Feature Extraction of Speech Emotion Recognition," ADBU-Journal of Engineering Technology, vol. 4, no. 1, pp. 7-9, 2016.

[20] Y. Pan, P. Shen and L. Shen, "Speech Emotion Recognition Using Support Vector Machine," International Journal of Smart Home, vol. 6, no. 2, pp. 101-106, 2012.

[21] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods," Speech Communication, vol. 48, no. 2006, p. 1162–1181, 2006.

[22] X. m. Cheng, P. y. Cheng and L. Zhao, "A Study on Emotional Feature Analysis and Recognition in Speech Signal," in IEEE 2009 International Conference on Measuring Technology and Mechatronics Automation, Zhangjiajie, 2009.

[23] A. Payal, C. Rai and B. Reddy, "Comparative analysis of Bayesian regularization and Levenberg-Marquardt training algorithm for

localization in wireless sensor network," in IEEE 15th International Conference on Advanced Communications Technology (ICACT), PyeongChang, 2013.

[24] A. Gupta and A. Joshi, "Speech Recognition using Artificial Neural Network," in IEEE 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, 2018.

[25] I. A. Maaly and M. El-Obaid, "Speech Recognition using Artificial Neural Networks," in IEEE 2006 2nd International Conference on Information & Communication Technologies, Damascus, 2006.

[26] K. S and C. E, "A Review on Automatic Speech Recognition Architecture and Approaches," International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 9, no. 4, pp. 393-404, 2016.

[27] A. Milton, S. S. Roy and S. T. Selvi, "SVM Scheme for Speech Emotion Recognition using MFCC Feature," International Journal of Computer Applications, vol. 69, no. 9, pp. 34-38, 2013.

[28] "Berlin Emotional Speech Database," [Online]. Available: http://www.emodb.bilderbar.info/download/.