Performance Analysis of Gender Clustering and Classification Algorithms

Dr.K.Meena
Vice Chancellor, Bharathidhasan University,
Principal and Director,
Shrimathi Indira Gandhi College, Trichy-2.

Dr.K.R.Subramaniam
Professor
Department of Computer Application
Shrimathi Indira Gandhi College, Trichy-2.

M.Gomathy
Research Scholar,
Department of Computer Science,
Shrimathi Indira Gandhi College, Trichy-2.
gomathymphd@gmail.com, m.gomathy.lect@gmail.com

***Abstract*—In speech processing, gender clustering and classification plays a major role. In both gender clustering and classification, selecting the feature is an important process and the often utilized feature for gender clustering and classification in speech processing is pitch. The pitch value of a male speech differs much from that of a female speech. Normally, there is a considerable frequency value difference between the male and female speech. But, in some cases the frequency of male is almost equal to female or frequency of female is equal to male. In such situation, it is difficult to identify the exact gender. By considering this drawback, here three features namely; energy entropy, zero crossing rate and short time energy are used for identifying the gender. Gender clustering and classification of speech signal are estimated using the aforementioned three features. Here, the gender clustering is computed using Euclidean distance, Mahalanobis distance, Manhattan distance & Bhattacharyya distance method and the gender classification method is computed using combined fuzzy logic and neural network, neuro fuzzy and support vector machine and its performance are analyzed.**

Key words: Mahalanobis distance, Manhattan distance, Bhattacharyya distance, Neuro fuzzy, Support vector machine.

I. Introduction

Speech is a common way for communication between the humans in all over the world [1]. Normally during speech communication, first an idea about the speech is formed in the mind of speaker and then it is converted in the form of words, phrases, and sentences by applying proper grammatical rules [2]. Speech signals can be represented using a speech production model that views speech as the outcome of passing a glottal excitation waveform through a time-varying linear filter, which models the resonant characteristics of the vocal tract [7]. The voiced-unvoiced-silence (V/U/S) classification of speech signal gives an elementary acoustic segmentation of speech, which is very essential for speech analysis [5]. A series of individual sounds called phonemes that can almost identical to the sounds of each letter of the alphabet makes the composition of human speech [4].

Speech processing is the study of speech signals, where several techniques are used to process the speech. Speech processing is used in many applications like speech coding, speech synthesis, speech recognition, and speaker recognition technology [8]. Among the above, one of the most important system is speech recognition. The speech recognition process produces a set of words by converting the audio signal received from a microphone or a telephone [9] [10]. Speech recognition is a process of extracting and determining the linguistic information transmitted by a speech signal using computers or electronic circuits [3]. This process is used in several applications such as security device, domestic appliances, cellular phones, automated teller machine (ATM) and computers [6].

Gender classification is a difficult and challenging problem, which is used in several fields such as speech recognition, speaker diarization, speaker indexing, annotation and retrieval of multimedia database, synthesis, smart human-computer interaction, biometrics, social robots, etc [11]. The gender based differences in human speech are partially due to physiological differences such as vocal fold thickness or vocal tract length and partially due to differences in speaking style [12] [14]. The female speakers normally have higher formant frequencies as well as higher fundamental frequency (F_0) and the (F_0) differences are larger than the formant frequency differences between the male and female groups [15]. But for male speakers, the F_0 is lower, because of the qualities like aggressiveness, body size, self-assurance, and assertiveness [13]. Some of the commonly used audio features are energy entropy, pitch, zero crossing rate, spectral roll off, mel-frequency spectral coefficients, spectrogram feature etc [16] [17].

In this paper, gender clustering and classification of speech signals are computed using different algorithms and its performance is analyzed. The rest of the paper is organized as follows: Section 2 briefly reviews the related works; Section 3 describes the proposed technique with sufficient mathematical models and illustrations; Section 4 discusses the implementation results; and Section 5 concludes the paper.

II. RELATED WORKS

Some of the recent research works related to gender classification and clustering in speech processing are as follows.

Rakesh *et al.* [18] have introduced two models by using different speech processing methods and algorithms. Both the models have been used for producing Formant values and pitch values of the voice sample respectively. By utilizing these two models, the gender biased features and pitch value of a speaker have been extracted. The mean of formants and pitch of all the samples of a speaker have been calculated by implementing a model having loop and counters which generates a mean of Formant 1 and pitch value of the speaker. The speaker has been classified between Male and Female by measuring the Euclidean distance from the Mean value of Males and Females of the generated mean values of Formant 1 and Pitch by using a nearest neighbor technique. Using NI Lab VIEW, the algorithm has been implemented in real time.

Rao et al. [19] have utilized the time-varying glottal excitation component of speech for text independent gender recognition studies. Linear prediction (LP) residual has been used as a representation of excitation information in speech. A Hidden Markov Models (HMMs) has been used for capturing the gender-specific information in the excitation of voiced speech. The reduce in the error during training and identifying genders during testing phase near to 100 % accuracy have illustrated that the continuous Ergodic HMM can capably capture the gender-specific information in the excitation component of speech. In their gender identification study, they have calculated the size of testing data on the gender recognition performance by using gender specific features in various HMM states, and mixture components. They have used Texas Instruments and Massachusetts Institute of Technology (TIMIT) database for performing the gender recognition studies.

Nandyala et al. [20] have proposed a technique to create a real time isolated word speech recognition system for human-computer communication. The system has relied on the speaker. Their major task has been to identify the list of words said by the speaker via the microphone. The mel-frequecy cepstarl coefficients (MFCC) that provides good discrimination of the speech signal have been used as features. Using the Dynamic Programming Algorithm, the similarity between the stored template and the test template has been measured for the speech recognition, which provides the optimum distance. The proposed system has achieved a recognition accuracy of 88.0%. They have prepared an elementary list containing ten words of cities names in India and when a particular city name is spoken, the image corresponding to that city name has been displayed. Hence, it can be utilized for many applications, especially for tourism purposes.

Lakshmi et al. [21] have analyzed the acoustical features of speech of deaf persons with the aim of improving the speech recognition rate. Since the speech to text or sound system is available for a normal speaker, the deaf persons can also employ all the computer aided devices as well as can communicate with the normal speaker without any difficulty by designing a speech to text or sound system for deaf. The main characteristics of speech such as the fundamental frequency or the pitch frequency of the vocal fold and resonant frequency of the vocal tract or formants have been analyzed. They have noticed that there is a high variability in the speech of deaf when compared to normal speech. The fundamental frequency contour of deaf person has unusual characteristics and also their formants are extremely closed, so their pitch and formants cannot be used as features for speech recognition. But, the speaker classification can be performed using variation in pitch and formant as they are higher for the deaf than the normal speakers

Devi et al. [22] have discussed that the unwanted background noise from noisy environment such as car, bus, babble, factory, helicopter, street noise and more have disturbed the performance of speech-processing systems like speech coding, speech recognition etc. Hence to improve the performance, it was very essential to perform a noise classification. An important process considered during the design of their signal classification system was the selection of a best set of features that are efficient in separating the signals in the feature space. Noise classification is vital for minimizing the impact of environmental noises on speech processing tasks. A fuzzy adaptive resonance theory (ARTMAP) network and a modified fuzzy ARTMAP network have been developed for the classification of background noise signals. Also, their experimental results have been compared with both back propagation networks and Radial Basis Function Network (RBFN).

Revathi et al. [23] have proposed a robust perceptual features and iterative clustering approach for performing both speech and speaker recognition. Procedure used for the creation of training speech was different for developing training models for speaker independent speech and text independent speaker recognition. They have mainly stressed that better accuracy of 91%, 91% and 99.5% for mel frequency perceptual linear predictive cepstrum in connection with the three categories namely, speaker identification, isolated digit recognition and continuous speech recognition can be obtained by the use of clustering models designed for the training data.

Sedaaghi [24] have performed a comparative study of gender and age classification algorithms applied in speech signal. Their experiments have been conducted on the database such as Danish Emotional Speech database (DES) and English Language Speech Database for Speaker Recognition (ELSDSR). They have experimentally compared different classifiers namely, Bayes classifier using sequential floating forward selection (SFFS) for feature selection, probabilistic Neural Networks (PNNs), support vector machines (SVMs), the K nearest neighbor (K-NN) and Gaussian mixture model (GMM), in order to find an excellent classifier for gender and age classification when speech signal is processed. They have proved that around 95% of accuracy can be achieved for gender classification done using the speech signal either from male and female individually or from both genders.

III. ANALYZING SPEECH CLUSTERING AND CLASSIFICATION ALGORITHMS

Clustering and Classification are the two important processes in speech processing. The gender clustering is done by using Euclidean distance method in [25] and the gender classification is done by using combined neural network and fuzzy logic method in [27] and their results are analyzed. In this paper, the performance of both clustering and classification method are analyzed by applying different algorithms. Both clustering and classification results are obtained based on feature selection. In most of the research works, gender classification and clustering process is done by using pitch as feature. Normally, the pitch value of male and female speech has a great difference. Also, the frequency value of male is higher than female speech. But, in some cases the frequency of male is nearly similar to female or frequency of female is similar to male. In such situation, it is difficult to recognize the exact gender. By considering all these drawbacks into account, in our proposed method three features such as energy entropy, zero crossing rate and short time energy are used for identifying the gender. By using the aforesaid features, the gender clustering and classification is performed. Initially, we discuss about the features that are used in speech clustering and classification.

A. Feature Analysis for Gender Classification

The most important process in gender clustering and classification is feature selection. The three features that are employed in our method are:

i. Energy Entropy
ii. Short Time Energy
iii. Zero Crossing Rate

Among these three features, the salient feature is zero crossing rate. These features are explained briefly in [25]. We can see the basic operation of these three features one by one.

i. Energy Entropy

Energy entropy in speech signal indicates the abrupt changes in the energy level of a speech signal. For computing the energy entropy, first the speech signal is divided into k frames and then the normalized energy for each frame is calculated. The energy entropy is calculated using the equation given below.

$$\text{Energy entropy, } E = -\sum_{i=0}^{k-1} \sigma^2 . \log_2 (\sigma^2) \qquad (1)$$

$$\text{where, } \sigma^2 = \sum_{a=1}^{N} \frac{\left(N * \frac{L}{M} \right)^2}{F}$$

σ^2 is the normalized energy, N is the total number of blocks, L is the window length, M is the number of short blocks, F is the frequency.

The above equation is applied to the sample signals. From the result obtained it is clear that the energy entropy for male is low and distributed, whereas for females it is high and remains only for a short period of time.

ii. **Short Time Energy**

The sudden rise in energy signal is said to be the short time energy of speech signal. For computing the short time energy, first the signal is divided into s windows and then the windowing function is calculated for each window. The short time energy is calculated using the equation given below.

$$\text{Short time energy, } S = \sum_{r=-\infty}^{\infty} y(r)^2 . h(s-r) \tag{2}$$

where, $h(s)$ is the impulse response, $y(r)$ is the input signal.

The above equation is applied to the sample signals. From the result obtained it is obvious that the short time energy for male is low, but for female it is high and continuous.

iii. **Zero crossing rate**

Among the three features, the zero crossing rate is the most important feature considered in our method. The ratio of number of time-domain zero crossings occurred to the frame length is called as zero crossing rate. The zero crossing rate is calculated using the equation given below.

$$Z = \frac{1}{N} \sum_{i=1}^{N-1} \frac{\operatorname{sgn}\{x(i)\} - \operatorname{sgn}\{x(i-1)\}}{2} \tag{3}$$

where, N is the length of the sample, $\operatorname{sgn}\{x(i)\}$ represent the sign function, i.e.

$$\operatorname{sgn}\{x(i)\} = \begin{cases} 1; x(i) > 0 \\ 0; x(i) = 0 \\ -1; x(i) < 0 \end{cases} \tag{4}$$

The above equation is applied to the sample signals. It is clear from the result that the ZCR for female speech is higher when compared to the male speech. The clustering and classification process is done by using the abovementioned three features. First we discuss about the performance for speech clustering method.

B. Performance Analysis of Speech Clustering

In [25] the speech clustering is done by using Euclidean distance method. In this method, initially the threshold value is calculated for all the three features given above and using this threshold value, the male and female speech signal are splitted. For a particular speech signal, if all the features are belongs to one gender then that speech signal is belongs to that gender itself. If one speech signal is present in both male and female gender based on grouping using the three features, then that speech belongs to which gender is identified using Euclidean distance method. In this paper, the gender grouping performance is analyzed by applying different algorithms like Mahalanobis distance, Manhattan distance and Bhattacharyya distance. Initially, we see about the performance of gender clustering using Mahalanobis distance method.

1) Gender Clustering using Mahalanobis Distance Method

Here, the Mahalanobis distance method is applied instead of Euclidean distance method. Equation to calculate distance using Mahalanobis distance method is as follows.

$$D = \sqrt{(x_s - y_s) * C^{-1} * (x_s - y_s)^T} \tag{5}$$

where, C is covariance matrix, $x = \{x_1, x_2, ...x_s\}^T$ and $y = \{y_1, y_2, ...y_s\}^T$.

Using the above equation, the distance for the given set of speech signal is computed and based on the result the given set of speech signal is grouped. Next we see about the Manhattan distance method.

2) Gender Clustering using Manhattan Distance Method

The Manhattan distance method is applied instead of Euclidean distance method. Equation to calculate distance using Manhattan distance method is as follows.

$$D = \sum_{i=1}^{T} |x_i - y_i| \tag{6}$$

where, $x = \{x_1, x_2, ...x_s\}^T$ and $y = \{y_1, y_2, ...y_s\}^T$.

Using the above equation, the distance for the given set of speech signal is computed and based on the result the given set of speech signal is grouped. Next we see about the Bhattacharyya distance method.

3) Gender Clustering using Bhattacharyya Distance Method

The Bhattacharyya distance method is applied instead of Euclidean distance method. Equation to calculate distance using Bhattacharyya distance method is as follows.

$$D = 0.125 * (m_1 - m_2)^T P^{-1} * (m_1 - m_2) + 0.5 * \ln\left(\frac{\det P}{\sqrt{\det P_1 * \det P_2}}\right) \quad (7)$$

where, m_i and P_i are the mean and covariance of the distributions.

$$P = \frac{P_1 + P_2}{2} \quad (8)$$

Using the above equation, the distance for the given set of speech signal is computed and based on the result the given set of speech signal is grouped.

C. Performance Analysis of Speech Classification

In [27] speech classification is done by using combined neural network and fuzzy logic. In this method, the gender classification is performed by using the above explained features. Initially, the fuzzy logic and neural network is trained based on the testing signals. In the testing stage if one signal is given as input, the three features are calculated for that signal and the value of three features are given as input to the fuzzy logic and neural network. Neural network and fuzzy logic gives the percentage of male and female feature as output and then mean is calculated for fuzzy logic and neural network. This mean value gives the exact gender to which the speech signal belongs. In this paper, instead of using fuzzy and neural network, neuro fuzzy and support vector machine are used and their performances are analyzed. First, we see about the performance of gender classification using neuro fuzzy.

1) Gender Classification using Neuro Fuzzy

Here, neuro fuzzy is applied instead of neural network and fuzzy logic. Generally, neuro fuzzy is a combination of neural network, where the fuzzy rules are generated based on input variables and using these rules, the neural network is trained. Input variables to neuro fuzzy are three features namely energy entropy, zero crossing rate and short time energy and output is male/female gender. By considering this input and output variables, fuzzy rules are generated and then the neural network is trained using this generated fuzzy rules. In the testing stage, if one speech signal is given as input, it gives the speech signal that belongs to male or female gender as output. The structure of neural network used in our method is as follows.

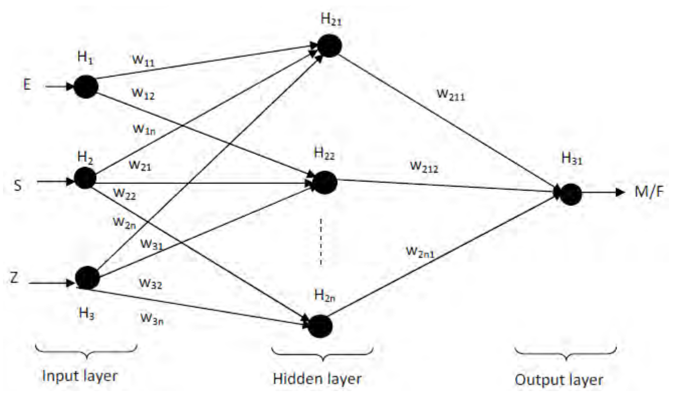

Figure 1. Structure of neural network used in our proposed technique.

2) Gender Classification using Support Vector Machine

Here, support vector machine (SVM) is applied instead of neural network and fuzzy logic. To identify the speech signal belongs to which gender, the SVM must be trained. The SVM is trained using an optimization function and that optimization function is shown in equation [9].

$$f = \sum_{i=1}^{n} \alpha_i y_i K(s_i, x) + b \quad (9)$$

where, α is the coefficient, n is the number of samples, y is the label value of sample, K is the kernel function, b is the bias value, s is the supporting vector and x is the feature vector to be classified.

Using the above equation, the SVM is trained and after the completion of training, if we given a speech signal as input, it gives the output as the given speech signal belongs to male or female.

IV. RESULT AND DISCUSSIONS

The proposed method is implemented in working platform of MATLAB 7.11 and it is tested using Harvard-Haskins database [26]. Here, 80 speech signals are taken from the Harvard-Haskins database, and they are splitted into four datasets and each dataset contains equal number of male and female speech signal. Then, the gender clustering and classification is performed for the abovementioned dataset. The performance of our method is analyzed using different equations as mentioned in section 4.1.

A. Performance Analysis

Initially, true positive, true negative, false positive and false negative values are computed from the result obtained from clustering and classification methods and using these values, performance parameters given below are computed.

$$False\ positive\ rate,\alpha = \frac{FP}{(FP+TN)} \tag{10}$$

$$False\ negative\ rate,\beta = \frac{FN}{(TP+FN)} \tag{11}$$

$$Sensitivity = Power = 1-\beta \tag{12}$$

$$Likelihood\ \ ratio\ \ positive = \frac{Sensitivit\,y}{(1-Specificit\,y)} \tag{13}$$

$$Likelihood\ ratio\ negative = \frac{(1-Sensitivity)}{Specificity} \tag{14}$$

$$Specificity = \frac{TN}{(FP+TN)} \tag{15}$$

$$Sensitivity = \frac{TP}{(TP+FN)} \tag{16}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{17}$$

$$\Pr ecision = \frac{TP}{(TP+FP)} \tag{18}$$

First, we see about the performance of different distance algorithms in gender clustering.

B. Gender Clustering

In [25], gender clustering is done by using Euclidean distance method. In this paper, instead of Eulidean distance, three different distance algorithms namely Mahalanobis distance, Manhattan distance and Bhattacharyya distance are applied and its results are analyzed.

TABLE 1 . PERFORMANCE ANALYSIS USING DISTANCE ALGORITHMS

	Dataset	**Euclidean distance method**	**Mahalanobis distance**	**Manhattan distance**	**Bhattacharyya distance**
Specificity	1	0.75	0.75	1	1
	2	0.5	0.5	0.5	0.5
	3	1	1	1	1
	4	0.5	0.5	0.5	0.5
Sensitivity	1	0.75	1	0.75	0.75
	2	1	1	1	1
	3	0.25	0.25	0.5	0
	4	0.75	0.75	0.75	0.75
True Positive	1	3	4	3	3
	2	4	4	4	4
	3	1	1	2	0
	4	3	3	3	3
True Negative	1	3	3	4	4
	2	2	2	2	2
	3	4	4	4	4
	4	2	2	2	2
False Positive	1	1	1	0	0
	2	2	2	2	2
	3	0	0	0	0
	4	2	2	2	2

False Negative	1	1	0	1	1
	2	0	0	0	0
	3	3	3	2	4
	4	1	1	1	1
Alpha	1	0.25	0.25	0	0
	2	0.5	0.5	0.5	0.5
	3	0	0	0	0
	4	0.5	0.5	0.5	0.5
Beta	1	0.25	0	0.25	0.25
	2	0	0	0	0
	3	0.75	0.75	0.5	1
	4	0.25	0.25	0.25	0.25
LRP	1	3	4	0	0
	2	2	2	2	2
	3	0	0	0	0
	4	1.5	1.5	1.5	1.5
LRN	1	0.33	0	0.25	0.25
	2	0	0	0	0
	3	0.75	0.75	0.5	1
	4	0.5	0.5	0.5	0.5
Accuracy	1	0.75	0.87	0.875	0.875
	2	0.75	0.75	0.75	0.75
	3	0.625	0.625	0.75	0.5
	4	0.625	0.625	0.625	0.625
Precision	1	0.75	0.8	1	1
	2	0.667	0.667	0.667	0.667
	3	1	1	1	0
	4	0.6	0.6	0.6	0.6

The most important parameter in performance analysis is accuracy. The average accuracy of gender clustering using Eulidean distance is 68.75%, Mahalanobis distance method-71.75%, Manhattan distance method-75% & Bhattacharyya distance method -68.75%. From the analysis result it is clear that Manhattan distance method is better than the Euclidean distance method, Mahalanobis distance method and Bhattacharyya distance method. In the future works if Manhattan distance method is used instead of Euclidean distance method the accuracy will be high.

The accuracy, sensitivity and precision values obtained from each dataset for Euclidean distance method, Mahalanobis distance method, Manhattan distance method and Bhattacharyya distance method are shown in figure 2, 3 & 4 respectively.

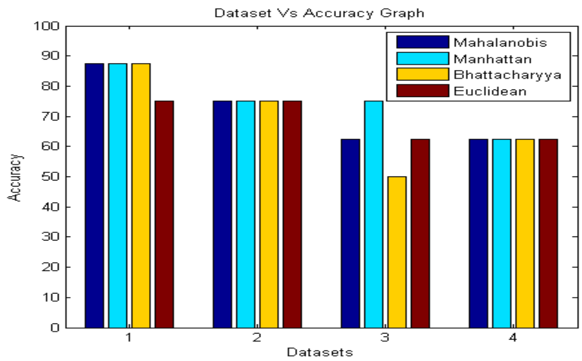

Figure 2. Dataset vs Accuracy comparison graph

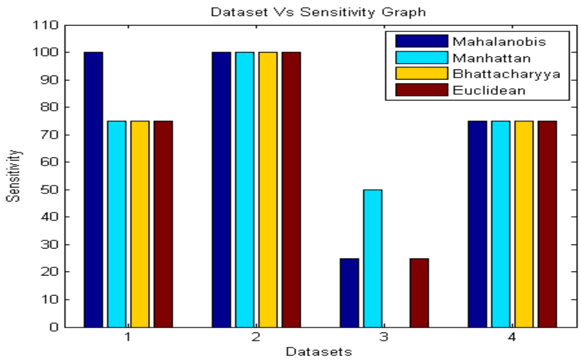

Figure 3. Dataset vs Sensitivity comparison graph

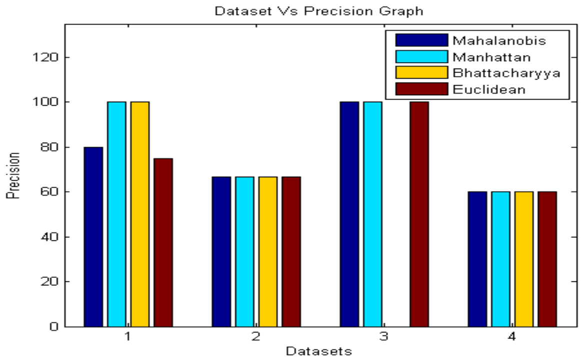

Figure 4. Dataset vs Precision comparison graph

C. Gender Classification

In [27], gender classification is done by using combined fuzzy logic and neural network. In this paper, instead of using combined fuzzy logic and neural network, two classification methods such as neuro fuzzy and SVM are used for gender classification in speech processing. In [27], the threshold value taken for gender classification in speech processing is 0.5, but here by varying the threshold values as 0.1 and 0.8, the gender classification results are analyzed. Initially, we see about the comparison of various classification methods with threshold value as 0.5.

1) At Threshold value 0.5

TABLE 2 . PERFORMANCE ANALYSIS AT THRESHOLD VALUE=0.5

	Dataset	**Combined fuzzy logic and neural network method**	**Neuro fuzzy**	**SVM**
Specificity	1	0.4	0.4	0.9
	2	0.4	0.8	0.7
	3	0.9	0.7	0.6
	4	0.8	0.3	0.9
Sensitivity	1	0.4	0.2	0.5
	2	0.4	0.4	0.3
	3	0.4	0.3	0.3
	4	0.5	0.3	0.6
True Positive	1	4	2	5
	2	4	4	3
	3	4	3	3
	4	5	3	6
True Negative	1	4	4	9
	2	4	8	7
	3	9	7	6
	4	8	3	9
False Positive	1	6	6	1
	2	6	2	3
	3	1	3	4
	4	2	7	1

False Negative	1	6	8	5
	2	6	6	7
	3	6	7	7
	4	5	7	4
Alpha	1	0.6	0.6	0.1
	2	0.6	0.2	0.3
	3	0.1	0.3	0.4
	4	0.2	0.7	0.1
Beta	1	0.6	0.8	0.5
	2	0.6	0.6	0.7
	3	0.6	0.7	0.7
	4	0.5	0.7	0.4
LRP	1	0.667	0.333	5
	2	0.667	2	1
	3	4	1	0.75
	4	2.5	0.428	6
LRN	1	1.5	2	0.556
	2	1.5	0.75	1
	3	0.667	1	1.1667
	4	0.625	2.333	0.444
Accuracy	1	0.4	0.3	0.7
	2	0.4	0.6	0.5
	3	0.65	0.5	0.45
	4	0.65	0.3	0.75
Precision	1	0.4	0.25	0.833
	2	0.4	0.667	0.5
	3	0.8	0.5	0.428
	4	0.714	0.3	0.857

The average accuracy of gender classification using combined neural network and fuzzy logic is 52.5%, Neuro fuzzy-42.5%, & SVM-60%. From the analysis result it is clear that SVM is better than combined neural network and fuzzy logic and neuro fuzzy. In the future works if SVM is used instead of combined neural network and fuzzy logic the accuracy will be high.

The accuracy, specificity, sensitivity and precision values obtained from each dataset for combined fuzzy logic and neural network method, Neuro fuzzy and SVM are shown in figure 5, 6, 7 &8 respectively.

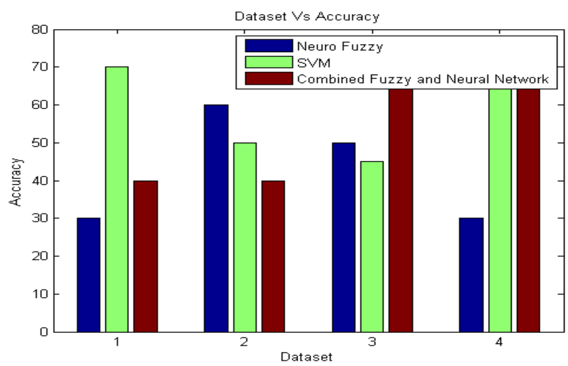

Figure 5. Dataset vs Accuracy comparison graph

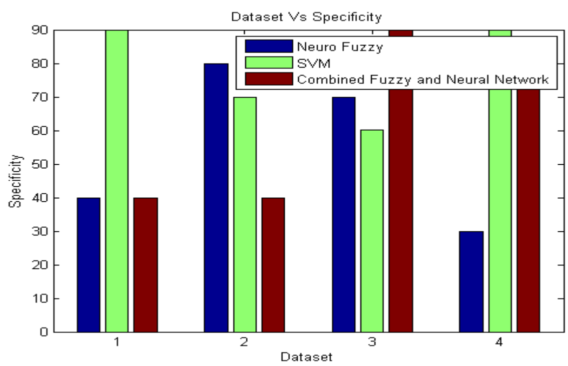

Figure 6. Dataset vs Specificity comparison graph

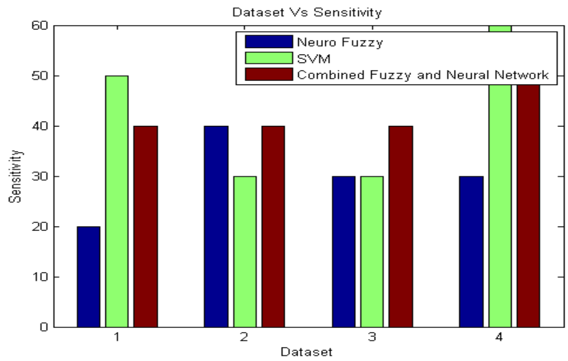

Figure 7. Dataset vs Sensitivity comparison graph

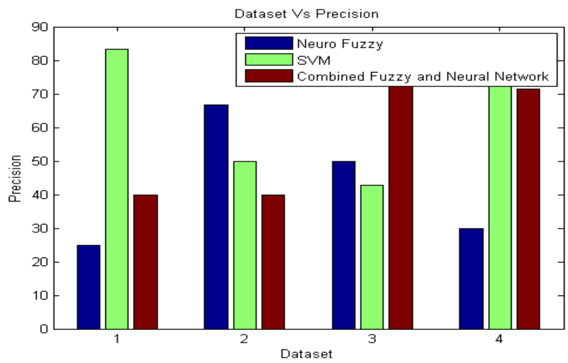

Figure 8. Dataset vs Precision comparison graph

2) At Threshold value 0.1

TABLE 3. PERFORMANCE ANALYSIS AT THRESHOLD VALUE=0.1

	Dataset	Combined fuzzy logic and neural network method	Neuro fuzzy	SVM
Specificity	1	0.4	0.6	0.9
	2	0.4	0.3	0.7
	3	0.7	0.4	0.6
	4	0.4	0.2	0.9
Sensitivity	1	0.7	0.5	0.5
	2	0.6	0.3	0.3
	3	0.5	0.5	0.3
	4	0.3	0.3	0.6
True Positive	1	7	5	5
	2	6	3	3
	3	5	5	3
	4	3	3	6
True Negative	1	4	6	9
	2	4	3	7
	3	7	4	6
	4	4	2	9
False Positive	1	6	4	1
	2	6	7	3
	3	3	6	4
	4	6	8	1
False Negative	1	3	5	5
	2	4	7	7
	3	5	5	7
	4	7	7	4
Alpha	1	0.6	0.4	0.1
	2	0.6	0.7	0.3
	3	0.3	0.6	0.4
	4	0.6	0.8	0.1
Beta	1	0.3	0.5	0.5

	2	0.4	0.7	0.7
	3	0.5	0.5	0.7
	4	0.7	0.7	0.4
LRP	1	1.166	1.25	5
	2	1	0.428	1
	3	1.667	0.833	0.75
	4	0.5	0.37	6
LRN	1	0.75	0.833	0.55
	2	1	2.33	1
	3	0.714	1.25	1.166
	4	1.75	3.5	0.44
Accuracy	1	0.55	0.55	0.7
	2	0.5	0.3	0.5
	3	0.6	0.45	0.45
	4	0.35	0.25	0.7
Precision	1	0.538	0.55	0.833
	2	0.5	0.3	0.5
	3	0.625	0.454	0.428
	4	0.33	0.272	0.857

The average accuracy of gender classification using combined neural network and fuzzy logic is 50%, Neuro fuzzy-38.75%, & SVM-58.75%. From the analysis result it is clear that SVM is better than combined neural network and fuzzy logic and neuro fuzzy. In the future works if SVM is used instead of combined neural network and fuzzy logic the accuracy will be high.

The accuracy, specificity, sensitivity and precision values obtained from each dataset for combined fuzzy logic and neural network method, Neuro fuzzy and SVM are shown in figure 9, 10, 11 &12 respectively.

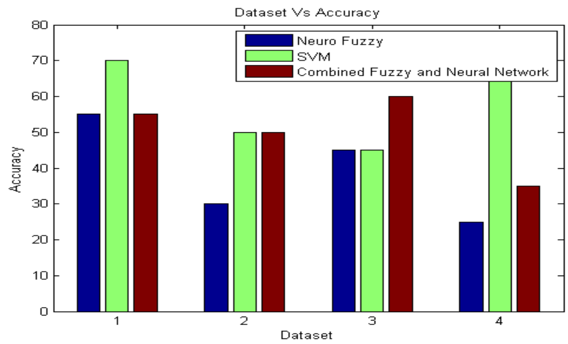

Figure 9. Dataset vs Accuracy comparison graph

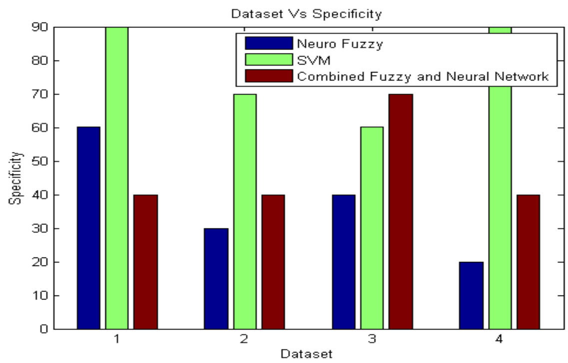

Figure 10. Dataset vs Specificity comparison graph

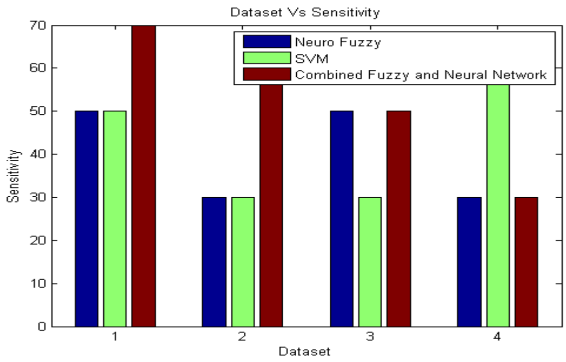

Figure 11. Dataset vs Sensitivity comparison graph

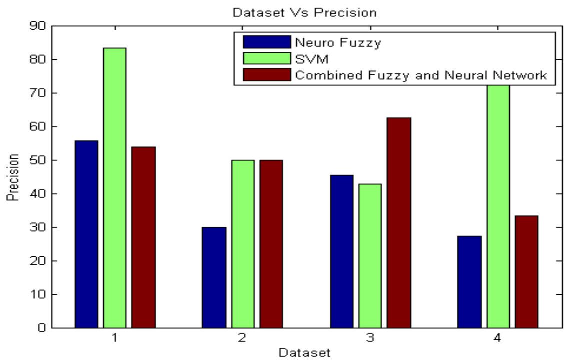

Figure 12. Dataset vs Precision comparison graph

3) At Threshold value 0.8:

Table 4. Performance analysis at Threshold value=0.8

	Dataset	**Combined fuzzy logic and neural network method**	**Neuro fuzzy**	**SVM**
Specificity	1	0.6	0.7	0.9
	2	0.7	0.4	0.7
	3	0.7	0.6	0.6
	4	0.6	0.6	0.9
Sensitivity	1	0.3	0.4	0.5
	2	0.4	0.1	0.3
	3	0.1	0.2	0.3
	4	0.3	0.1	0.6
True Positive	1	3	4	5
	2	4	1	3
	3	1	2	3
	4	3	1	6
True Negative	1	6	7	9
	2	7	4	7
	3	7	6	6
	4	6	6	9
False Positive	1	4	3	1
	2	3	6	3
	3	3	4	4
	4	4	4	1
False Negative	1	7	6	5
	2	6	9	7
	3	9	8	7
	4	7	9	4
Alpha	1	0.4	0.3	0.1
	2	0.3	0.6	0.3
	3	0.3	0.4	0.4
	4	0.4	0.4	0.1
Beta	1	0.7	0.6	0.5
	2	0.6	0.9	0.7
	3	0.9	0.8	0.7
	4	0.7	0.9	0.4
LRP	1	0.75	1.33	5
	2	1.33	0.1667	1
	3	0.33	0.5	0.75
	4	0.75	0.25	6
LRN	1	1.166	0.857	0.55
	2	0.857	2.25	1
	3	1.285	1.33	1.167
	4	0.75	1.5	0.44
Accuracy	1	0.45	0.55	0.65
	2	0.55	0.25	0.45
	3	0.4	0.4	0.45
	4	0.45	0.35	0.70
Precision	1	0.428	0.57	0.833
	2	0.57	0.1428	0.5
	3	0.25	0.33	0.428
	4	0.428	0.2	0.85

The average accuracy of gender classification using combined neural network and fuzzy logic is 46.25%, Neuro fuzzy-38.75%, & SVM-56.25%. From the analysis result it is clear that SVM is better than combined neural network and fuzzy logic and neuro fuzzy. In the future works if SVM is used instead of combined neural network and fuzzy logic the accuracy will be high.

The accuracy, specificity, sensitivity and precision values obtained from each dataset for combined fuzzy logic and neural network method, Neuro fuzzy and SVM are shown in figure 13, 14, 15 &16 respectively.

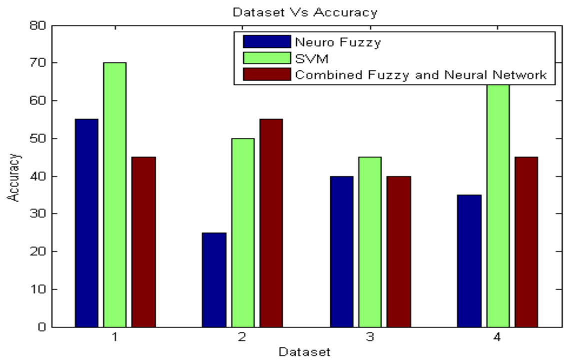

Figure 13. Dataset vs Accuracy comparison graph

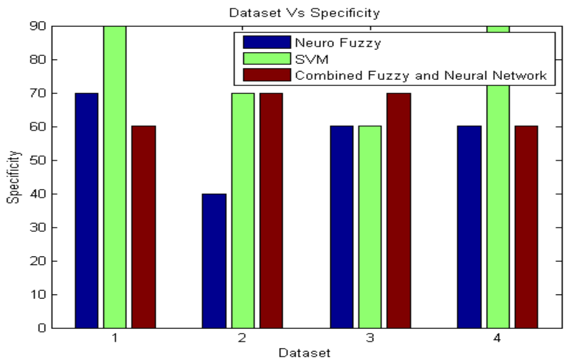

Figure 14. Dataset vs Specificity comparison graph

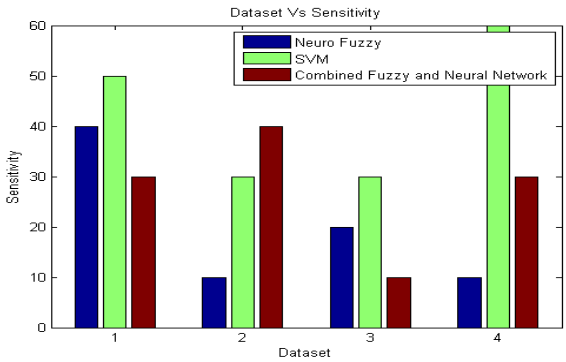

Figure 15. Dataset vs Sensitivity comparison graph

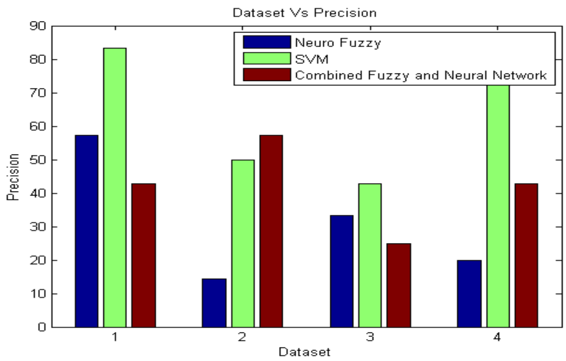

Figure 16. Dataset vs Precision comparison graph

Gender classification is analyzed for combined Fuzzy logic and neural network, SVM and neuro fuzzy for different threshold values like 0.5, 0.1 and 0.8. From the analysis results it is clear that SVM method is better than other methods and also the best threshold value is 0.5.

V. CONCLUSION

In this paper, gender clustering and classification of speech signal were computed using different algorithms and its results were analyzed. The gender clustering method was computed using Mahalanobis distance, Manhattan distance & Bhattacharyya distance and the gender classification method was computed using neuro fuzzy and support vector machine. The performance, accuracy, precision and sensitivity of all this gender clustering method were compared with Euclidean distance method and the gender classification method was compared with the combination of fuzzy logic and neural network. From the performance analysis result it is clear that for gender clustering Manhattan distance method is the best method and for gender classification SVM is the best method and best threshold value for gender classification is 0.5.

REFERENCE

[1] Anandthirtha. B. Gudi, H.K. Shreedhar and H. C. Nagaraj, "Signal Processing Techniques to Estimate the Speech Disability in Children", IACSIT International Journal of Engineering and Technology, Vol. 2, No. 2, pp. 169-176, April 2010.

[2] Anandthirtha. B. Gudi, and H. C. Nagaraj, "Optimal Curve Fitting of Speech Signal for Disabled Children", International Journal of Computer science & Information Technology (IJCSIT), Vol. 1, No 2, pp. 99-107, Nov 2009.

[3] Ramzi A. Haraty and Omar El Ariss, "CASRA+: A Colloquial Arabic Speech Recognition Application", American Journal of Applied Sciences, Vol. 4, No.1, pp. 23-32, 2007.

[4] James A. Rodger, Parag C. Pendharkar, "A field Study of the Impact of Gender and User’s Technical Experience on the Performance of Voice-Activated Medical Tracking Application", Int. J. Human-Computer Studies, Vol. 60, pp. 529–544, 2004.

[5] Yingyong Qi and Bobby R. Hunt, "Voiced-Unvoiced-Silence Classifications of Speech using Hybrid Features and a Network Classifier", IEEE Transactions on Speech and Audio Processing, Vol. 1, No.2, pp. 250-255, April 1993.

[6] Ibrahim Patel and Y. Srinivas Rao, "Speech Recognition using HMM with MFCC- an Analysis using Frequency Specral Decomposion Technique", Signal & Image Processing : An International Journal (SIPIJ), Vol. 1, No. 2, pp.101-110, Dec 2010.

[7] R.J. McAulay and T.F. Quatieri, "Speech Processing Based on a Sinusoidal Model", The Lincoln Laboratory Journal, Vol. 1, No. 2, pp. 153-168, 1988.

[8] M. Faúndez-Zanuy, S. McLaughlin, A. Esposito, A. Hussain, J. Schoentgen, G. Kubin, W. B. Kleijn and P. Maragos, "Non-linear Speech Processing: Overview and Applications, Control & Intelligent Systems", ACTA Press, Vol.30, No.1, pp. 1-10, 2002.

[9] Gurpreet Singh, Akhil Junghare and Priyam Chokhani, "Multi Utility E-Controlled cum Voice Operated Farm Vehicle", International Journal of Computer Applications, Vol. 1, No. 13, pp. 109-113, 2010.

[10] Akram M. Othman, and May H. Riadh, "Speech Recognition using Scaly Neural Networks", World Academy of Science, Engineering and Technology Vol. 38, pp. 253-258, 2008.

[11] Kotti. M and Kotropoulos. C, "Gender Classification In Two Emotional Speech Databases", In Proceedings of 19th International Conference on Pattern Recognition, pp. 1-4, Tampa, Dec 2008.

[12] Yu-Min Zengi, Zhen-Yang Wu, Tiago Falk and Wai-Yip Chan, "Robust GMM based Gender Classification using Pitch and Rasta-PLP Parameters of Speech", In Proceedings of Fifth International Conference on Machine Learning and Cybernetics, pp.13-16, Dalian, Aug 2006.

[13] Yoko Hasegawa and Kazue Hata, "Non-Physiological Differences Between Male and Female Speech: Evidence from the Delayed F0 Fall Phenomenon in Japanese", In Proceedings of 1994 International Conference on Spoken Language Processing, pp. 1179-82, 1994.

[14] Yen-Liang Shue and Markus Iseli, "The Role of Voice Source Measures on Automatic Gender Classification", In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4493-4496, Las Vegas, 2008.

[15] Yoko Hasegawa and Kazue Hata, "The Function of F0-Peak Delay in Japanese", In Proceedings of 21st Annual Meeting of the Berkeley Linguistics Society, pp. 141-151, 1995.

[16] Theodoros Giannakopoulos, Aggelos Pikrakis and Sergios Theodoridis, "A Multi-Class Audio Classification Method with Respect to Violent Content in Movies using Bayesian Networks", In Proceedings of IEEE 9th Workshop on Multimedia Signal, pp. 90-93, Crete, 2007.

[17] Theodoros Giannakopoulos, Aggelos Pikrakis and Sergios Theodoridis, "A Speech/Music Discriminator for Radio Recordings using Bayesian Networks", In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 809-812, Toulouse, 2006.

[18] Kumar Rakesh, Subhangi Dutta and Kumara Shama, "Gender Recognition using Speech Processing Techniques in LABVIEW", International Journal of Advances in Engineering & Technology, Vol. 1,No. 2, pp. 51-63, May 2011.

[19] R. Rajeshwara Rao, A. Prasad, "Glottal Excitation Feature based Gender Identification System using Ergodic HMM", International Journal of Computer Applications, Vol. 17, No. 3, pp. 31-36, March 2011.

[20] Siva Prasad Nandyala and T. Kishore Kumar, "Real Time Isolated Word Speech Recognition System for Human Computer Interaction", International Journal of Computer Applications, Vol. 12, No. 2, pp. 1-7, Nov 2010.

[21] C.Jeyalakshmi and V.Krishnamurthi and A.Revathy, "Deaf Speech Assesment using Digital Processing Techniques", Signal & Image Processing: An International Journal (SIPIJ), Vol. 1, No. 1, pp. 14-25, Sep 2010.

[22] T.Meera Devi, N.Kasthuri and A.M.Natarajan, "Performance Comparison of Noise Classification Using Intelligent Networks", International Journal of Electronics Engineering, Vol. 2, No.1, pp. 49-54, 2010.

[23] A.Revathi, R.Ganapathy and Y.Venkataramani, "Text Independent Speaker Recognition and Speaker Independent Speech Recognition using Iterative Clustering Approach", International Journal of Computer science & Information Technology (IJCSIT), Vol. 1, No 2, pp. 30-42, Nov 2009.

[24] M. H. Sedaaghi, "A Comparative Study of Gender and Age Classification in Speech Signals", Iranian Journal of Electrical & Electronic Engineering, Vol. 5, No. 1, pp. 1-12, March 2009.

[25] M.Gomathy, K.Meena and K.R.Subramaniam, “Gender grouping in speech recognition using statistical metrics of pitch strength”, European Journal of Scientific Research, Vol.61, No.4, pp.524-537, 2011.

[26] Harvard-Haskins database: http://vesicle.nsi.edu/users/patel/download.html

[27] M.Gomathy, K.Meena and K.R.Subramaniam, “Classification of speech signal based on Gender: A hybrid approach using Neuro-Fuzzy systems”, International Journal of Speech Technology, Vol.14, No. 4, pp.377-391, 2011

Dr.K.Meena, M.Sc, M.Phil, M.E (Computer Science and Engineering), M.I.E., Ph.D. She is the Vice-Chancellor of Bharathidhasan University. She is the principal and Director (M.B.A and M.C.A) of Shrimathi Indira Gandhi College, Trichirapalli. She has rich experience in the development of software tools for the assessment of specially abled children. She also provides consultancy for organizing specific programmes for creating awareness/literacy about the computer and information technology among specific cross-sections of the society (Co-ordinator of the novel project IT ON WHEELS – from Lab to Land). Provides counseling for higher education, career placement and training.

Dr.K.R.Subramaniam received B.Sc, M.Sc (Maths), M.A (English), M.Ed, M.Sc (I.T) and Ph.D (Maths and Computer Applications) from Madras, Annamalai University, Madurai and Bharathidhasan Universities, Tamil nadu, India in the years 1966, 1969, 1982, 1977, 1983, 2009 and 2003 respectively. From 1969 to 2007 he has been an Educationist for Mathematics, English, Educational Technology and Computer applications as Lecturer and Professor. He has headed the Department of Master of Computer Applications, Shrimathi Indira Gandhi College, Trichy-2, from 2007 to 2010.

M.Gomathy received B.Sc (Chemistry) degree in Holy Cross College in the year 1998. She has completed her M.S.I.T degree in the year 2001 from Shrimathi Indira Gandhi College, Trichy, Bharathidhasan University, Then, She has completed her M.Phil degree from St.Josephs College, Trichy in the year 2002. From 2003 to 2010, She has been an Educationist for computer applications, Information Technology as Lecturer and Professor.