International Journal of Advance Research in Science and Engineering
Volume No. 10, Issue No. 08, August 2021
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Classification and Testing of Speaker Verification System

N K Kaphungkui[1], Dr Aditya Bihar Kandali[2]

Department of Electronics and Communication, Dibrugarh University, Assam, India.

Department Of Electrical Engineering, Jorhat Engineering College, Assam, India.

ABSTRACT:

Automatic Speaker Recognition is a system which validates the speaker's identity to gain access into the system by using a unique speaker's speech characteristics. In simple term it is a man-machine interface system. The objective of this paper is to classify twenty (20) speakers with acceptable accuracy rate. Mel frequency Cepstral Coefficient (MFCC) is used for the extraction of speech features and Neural Network as speech classifier. After the speakers are correctly classified, the classifier is tested with the registered speaker's new speech data for their recognition. The back propagation neural network used a Scaled conjugate gradient as its training function. A speech database from a group of ten male and ten female is created uttering the same sentence repeatedly. The total number of data set of 20 speakers for classification is 28,414. The classification accuracy obtained is 87.8% with small misclassification of 12.2%. The overall precision and sensitivity scores are 87.67%, 88.07% respectively which are good enough. Matlab tool is used for the entire simulation.

Keywords: Speaker recognition, MFCC, Back Propagation Neural Network, Accuracy, training, testing.

I. INTRODUCTION

Speaker recognition system identifies the person based on the characteristics of their voice signal. It can be broadly classified as open set & close set, text dependent & text independent, and identification & verification[2], [13]. The system is an open set type if it can have many trained speakers. Whereas in closed set, the system has a fix number of registered speakers in the system. The system is Text-Dependent type if the same speech is used for both training as well as the testing session by the system. Therefore the system required cooperation from the test speaker for successful recognition [1]. Whereas in text independent type the test speaker can speak anything while testing as no prior information about the utterances is available to the speaker [4]. Lastly but not the least, in Speaker identification type, it finds the unknown speaker's identity by comparing the registered speech data into the system with the given utterance. Thus it is one to many comparison procedure. Whereas in Speaker verification system, as it is to either accept or dismiss the claimed speaker it follows the procedure of one to one mapping process [17]. Two modules govern the whole of Automatic Speaker Recognition, which are speech features extraction module and speech feature matching module. Several methods used for speech features extraction techniques are Linear Prediction coefficients, Mel-Frequency Cepstral Coefficients, Gammatone Frequency Cepstral Coefficients, Linear Predictive Cepstral Coefficients and Perceptual Linear Predictive. Similarly for Speech features matching methods, several classifiers such as Dynamic Time Warping, Gaussian Mixture Model, Hidden Markov Model, K- Mean Clustering and Vector

Quantization are used. The important phases that any Speaker Recognition should undergo are the training phase and testing phase [19]. A researcher have reported that an accuracy of 81.8% is obtained for class ten classification with the combination of MFCC, pitch and rms in feed forward neural network (FFNN) [12]. For multiple classes's classification wrt accuracy, Artificial Neural Network outperforms fuzzy logic based systems if speech is recorded in a clean environment. With ANN an Accuracy of 74% is achieved against 72% accuracy rate with fuzzy logic [6]. Others work have also achieved the highest accuracy of 92% for 10 speakers classification with the combination of Mel Frequency Cepstrum Coefficients and BPNN in text dependent speaker recognition system. [15]. With the combination of LPC and MFCC for Assamese Speaker Recognition using Artificial Neural Network for 10 users, a moderate accuracy is obtained [3].This work will be implemented for 20 speakers with the combination of MFCC and BPNN to achieve an accuracy of 87.8%.

II. MEL FREQUENCY CEPSTRAL COEFFICIENT

Many experts working in speech processing have already reported that speech features extraction with MFCC approach has more success rate. This is due to the fact that it is modelled as human auditory system that don't perceived over 1KHz frequency and showing more robust against noisy environment [5], [8]. Mel Frequency Cepstral Coefficients is an algorithm which generates the speaker's voice coefficients that are unique to every speaker [9]. A new way of using weighted MFCC is also reported for speaker recognition system claiming that the recognition rate is superior to non-weighted MFCC [20]. Speaker recognition with 16-order MFCC and LPCC as speech classifier is also presented [18]. The overall MFCC's steps for the extraction of speech feature are shown in Figure. 1 [7].

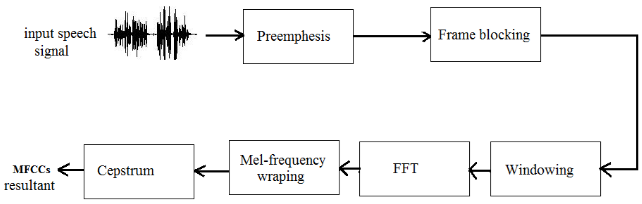
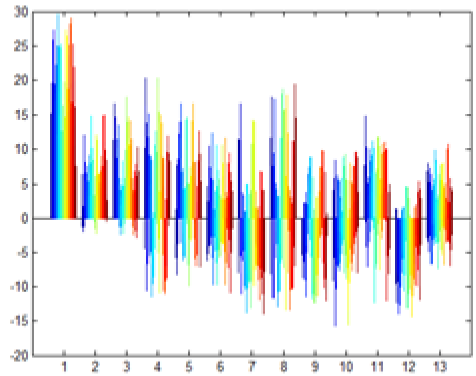


Figure. 1 MFCC computation steps

In this work also MFCC technique will be used for the speech feature extraction technique. All the collected speech samples are represented with a unique 13 coefficients for each speaker following a 13-order MFCC. The parameters for the computation of MFCC are set as frame duration of 0.026 sec, frame shift of 0.01 sec, Pre-emphasis coefficient of 0.97, filter banks channel number of 20, filter's Lower frequency limit of 300 Hz, filter's Upper frequency limit of 3700 Hz and Cepstral sine lifter parameter of 22. A speech data base is created from 20 different speakers by uttering the same sentence repeatedly. The data base consist of ten female f1, f2, f3, f4, f5, f6, f7, f8, f9,f10 and ten male speakers as m1, m2, m3, m4, m5, m6, m7, m8, m9 and m10. The voice samples are recorded and collected in a relatively noise less environment.The speech samples length varies for different person as it depends on a time taken for completing the utterance. The speech length will be more if more time is taken to complete the utterance and vice versa.The sampling of the voice is at 48000 Hz with a 16 bit, bit depth. The MFCC resultants are in matrix form consisting of fix 13 rows and variable column for different speaker as shown in Figure. 2 (a). This can be represented as M x N where M is fix 13 rows and N is variable columns.

Sl	Speakers		MFCC Size (M x N)
1	f1	s1	13 x 2064
2	f2	s2	13 x 1615
3	m1	s3	13 x 1050
4	m2	s4	13 x 1096
5	f3	s5	13 x 1426
6	m3	s6	13 x 1085
7	f4	s7	13 x 1544
8	f5	s8	13 x 1473
9	m4	s9	13 x 1419
10	m5	s10	13 x 1918
11	m6	s11	13 x 1242
12	f6	s12	13 x 1686
13	f7	s13	13 x 1302
14	m7	s14	13 x 1397
15	f8	s15	13 x 1408
16	f9	s16	13 x 1556
17	m8	s17	13 x 1226
18	m9	s18	13 x 1094
19	m10	s19	13 x 1466
20	f10	s20	13 x 1347

(a)



(b)

Figure.2 MFCC result for different speakers

The N values will be more when a speaker takes more time to utter the given sentence. The pictorial representation of a particular speaker's MFCCs is shown in Figure 2 (b).

III. BACK PROPAGATION NEURAL NETWORK

The multi-layer Perceptron neural network which consists of three different layers that is input layer, hidden layer and output layer is shown in Figure. 3.

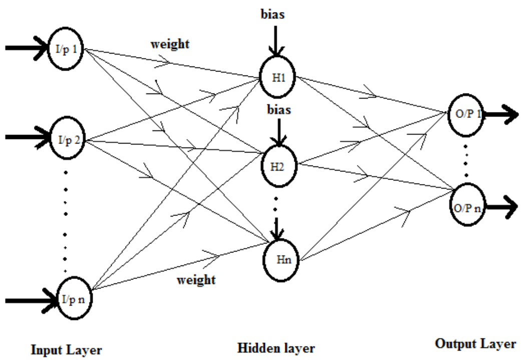


Fig. 3. Basic network of Multilayer Perceptron Neural Network.

The Back propagation algorithm consists of two basic steps for obtaining a desire output that is the feed forward propagation and the back propagation. In feed forward direction the data will propagates from the input node towards the output node through the hidden layer along with the assigned initial network's weights and bias. If the output produced by the network is not equal to the set target then back propagation process will takes place by updating the network weights and bias backward from output node towards input node. It is an algorithm for supervised learning of Artificial Neural Network. Initially when the network is designed, any random values are

International Journal of Advance Research in Science and Engineering
Volume No. 10, Issue No. 08, August 2021
www.ijarse.com
IJARSE
ISSN 2319 - 8354

assigned as weights to predict the desired target. By doing this if there exist huge variation between the network's output and the desire set target, weights and bias are updated to minimize this huge error. The specific network parameters which minimize the error function between the output and the set target will be the solution of the network leaning. The entire steps of BPNN are summarized as follows

- Error Calculation of network – in feed forward direction the difference between network output and desire output is obtained.
- Minimizing the error – Check whether the error is minimum or not
- The network parameters are updated – if huge difference occurs between the actual output and set target, the parameters are updated and the error is checked again. These processes will be repeated until the least minimum error is obtained.
- The Network training session end and ready to predict the set target – The training of the network stop when it finally obtained the least mean square error (MSE).

Figure. 4 show the flowchart of a Back Propagation Algorithm. The aim of a back propagation algorithm is to minimize the value of the mean square error (MSE) functions by updating the weights of the network [10], [11], [16]. The final weights minimizing the MSE function is the solution of neural network learning algorithm. Accuracy is high and performance is better if more number of training data set is used for training the network model [14]

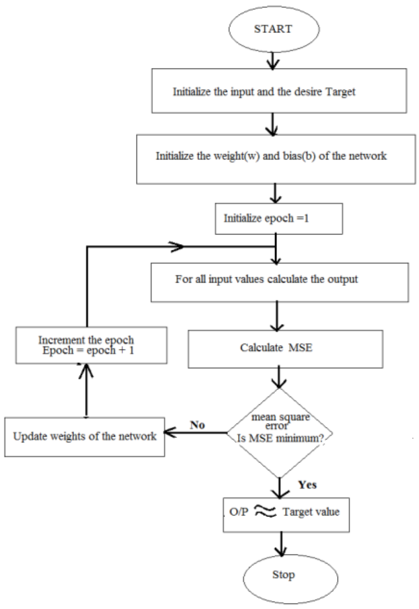


Figure.4 Flowchart of BPNN

In ideal condition, training session of the network will be stop when the network's actual output is equal to the desired set target value. To achieved this desire condition, the loop will run continuously with the increment of epoch number along with updating the network's weights.

IV. IMPLEMENTATION AND SIMULATION RESULT

The speech classifier model is built with multi-layer perceptron neural network consisting of 13 input nodes, 362 hidden neurons and 20 output nodes for twenty class's classification as shown in Figure 5. The input data's

International Journal of Advance Research in Science and Engineering
Volume No. 10, Issue No. 08, August 2021
www.ijarse.com
IJARSE
ISSN 2319 - 8354

for the neural network are obtained from MFCCs result of twenty different persons. The 13 order MFCC of each speaker are concatenated and fed as input training data to the neural network. The size of the input is in matrix form of the order 13 x 28,414. (Rows x Columns) and the neural network desire target is set at 20 x 28,414. The simulation result of twenty class's pattern classification is given in Figure. 6.

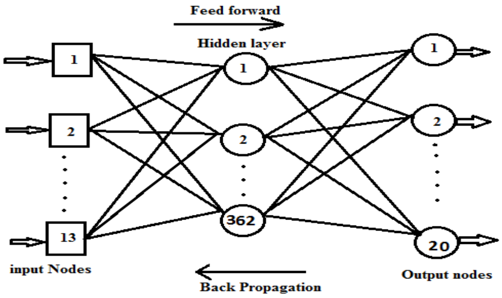


Figure. 5. Speech Classifier Model

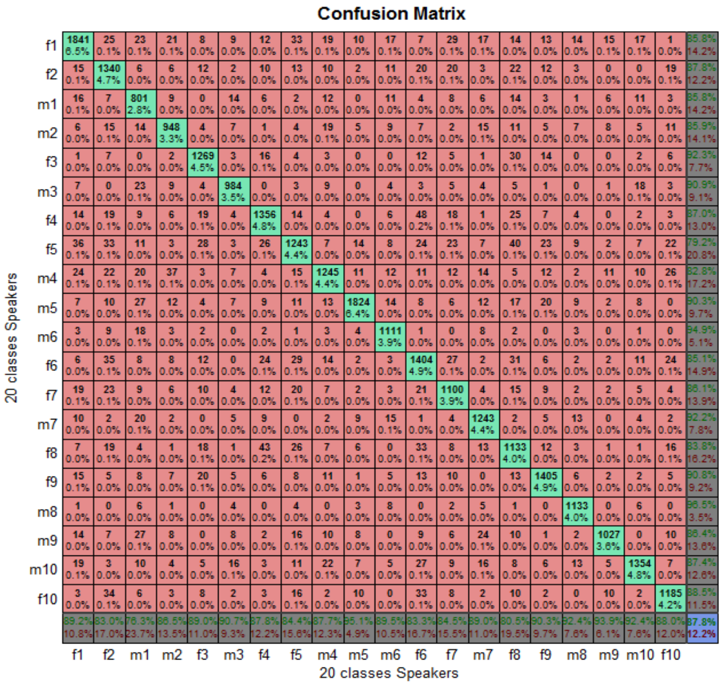


Figure. 6 Confusion matrix of 20 classes classification

The characteristics of confusion matrix can be define by the following terminoly.

- *Accuracy = (Sum of Diagonal elements)/(Sum of all elements)*
- *Precision = True Positive/(True positive+ False Positive).*
- *Sensitivity = True Positive/(True Positive + False Negative)*

The overall accuracy obtained from the classification is found to be 87.8%. The Precision and Sensitivity for all the speakers are also extracted from the confusion matrix and shown in Figure.7.

International Journal of Advance Research in Science and Engineering
Volume No. 10, Issue No. 08, August 2021
www.ijarse.com
IJARSE
ISSN 2319 - 8354

1	Precision of s1	89.2 %	1	Sensitivity of s1	85.8%
2	Precision of s2	83 %	2	Sensitivity of s2	87.8%
3	Precision of s3	76.3%	3	Sensitivity of s3	85.8%
4	Precision of s4	86.5%	4	Sensitivity of s4	85.9%
5	Precision of s5	89%	5	Sensitivity of s5	92.3%
6	Precision of s6	90.7%	6	Sensitivity of s6	90.9%
7	Precision of s7	87.8%	7	Sensitivity of s7	87%
8	Precision of s8	84.4%	8	Sensitivity of s8	79.2 %
9	Precision of s9	87.7%	9	Sensitivity of s9	82.8%
10	Precision of s10	95.1%	10	Sensitivity of s10	90.3 %
11	Precision of s11	89.5%	11	Sensitivity of s11	94.9%
12	Precision of s12	83.3 %	12	Sensitivity of s12	85.1%
13	Precision of s13	84.5%	13	Sensitivity of s13	88.1%
14	Precision of s14	89%	14	Sensitivity of s14	92.2%
15	Precision of s15	80.5%	15	Sensitivity of s15	83.8%
16	Precision of s16	90.3 %	16	Sensitivity of s16	90.8%
17	Precision of s17	92.4%	17	Sensitivity of s17	96.5%
18	Precision of s18	93.9%	18	Sensitivity of s18	86.4%
19	Precision of s19	92.4%	19	Sensitivity of s19	87.4%
20	Precision of s20	88%	20	Sensitivity of s20	88.5%

Figure. 7 Precision and Sensitivity table

The graphical plot of the classifier model's performance wrt precision and sensitivity of all the classes is shown in Figure.8.

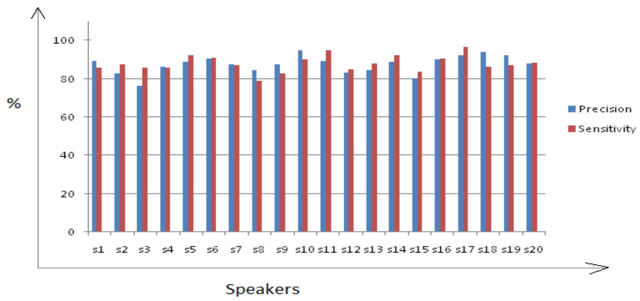


Figure. 8. Performance parameters of different speakers

The Receiver output characteristic (ROC) summarizes the overall performance of the neural network classifier. It is a graphical plot of the True Positive Rate along the y-axis against the False Positive Rate along x-axis. When the resultants lean sharply towards the true positive rate, the accuracy obtained will be high and better will be the classifier's classification. The best validation performance is 0.0081538 at 1000 epochs which is also shown in Figure. 9.

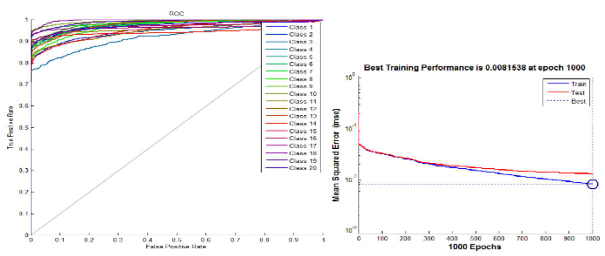


Figure. 9 Classifier's ROC and Validation Performance

After successful classification, the classifier is now ready for testing with new input data's. For this to execute, the voice samples of all registered speakers are collected again for the same utterance. The MFCCs for each speaker are extracted again and given as testing input to the network. The diagonal elements score from the

IJARSE
ISSN 2319 - 8354

confusion matrix while testing with different speakers is shown from Figure. 10 to Figure. 14. The score of diagonal element corresponding to the right speaker will always be maximum while all other elements will be the least. Therefore, by examining the row and column to which speaker this maximum score belongs, the speaker can be successfully recognized.

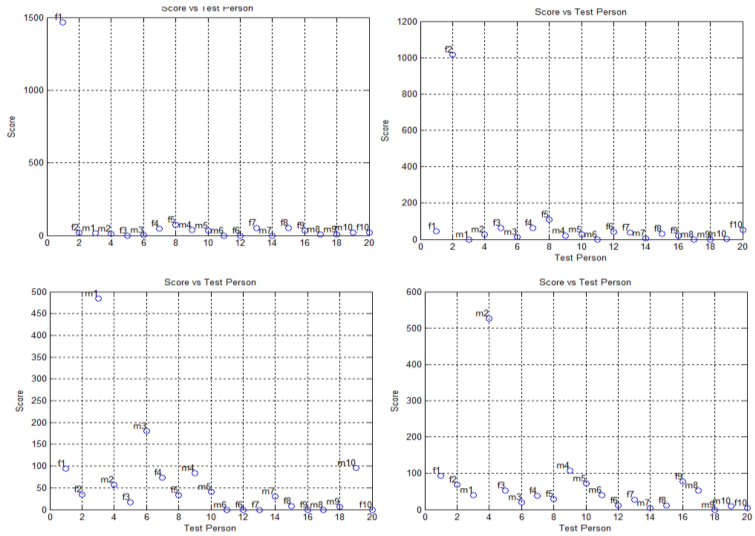


Figure. 10 The plot of diagonal score of confusion matrix while testing the network with s1, s2, s3 and s4 speakers.

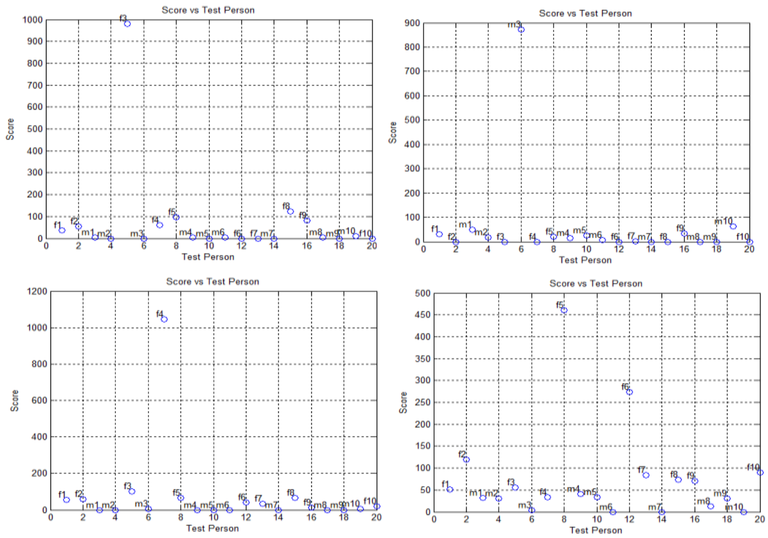


Figure. 11 The plot of diagonal score of confusion matrix while testing the network with s5, s6, s7 and s8 speakers.

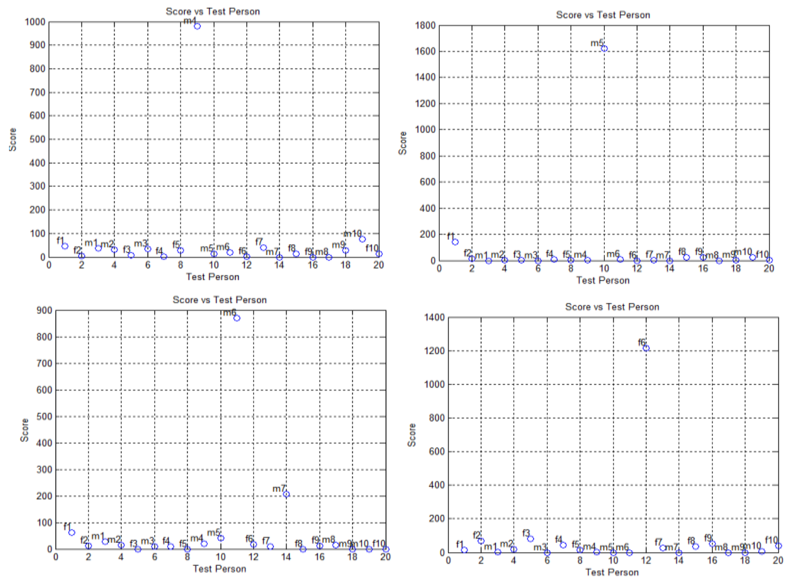


Figure. 12 The plot of diagonal score of confusion matrix while testing the network with s9, s10, s11 and s12 speakers

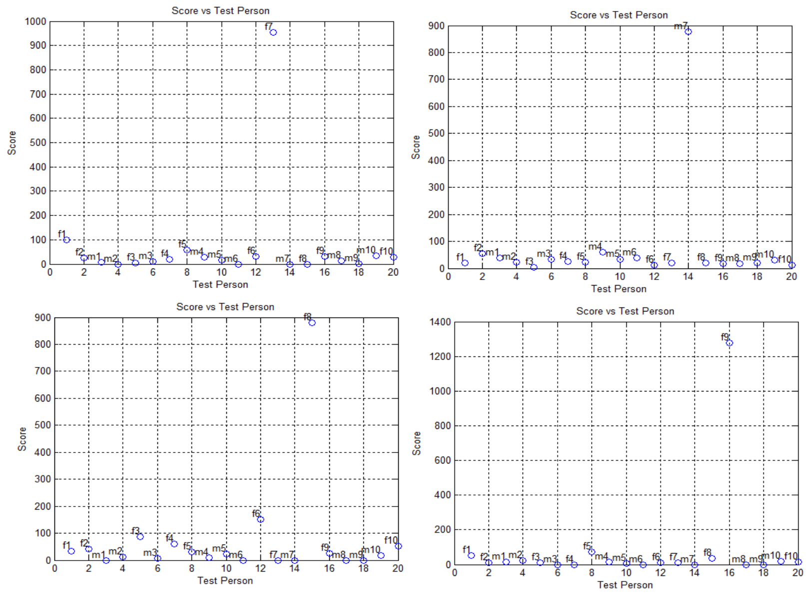


Figure. 13 The plot of diagonal score of confusion matrix while testing the network with s13, s14, s15 and s16 speakers.

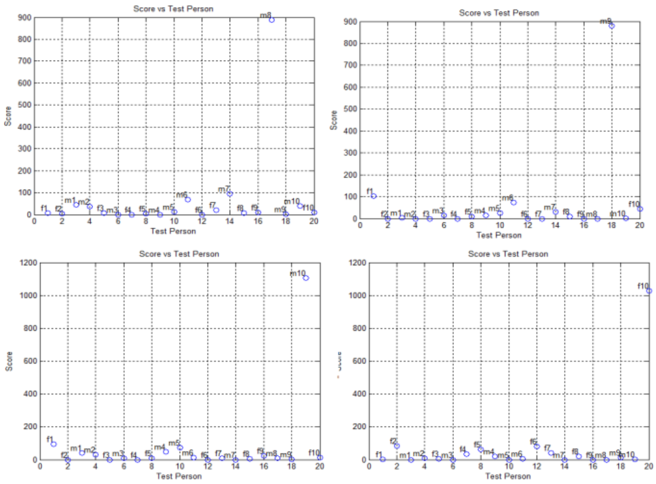


Figure. 14 The plot of diagonal score of confusion matrix while testing the network with s17, s18, s19 and s20 speakers

V. CONCLUSION

When the neural network is trained with sufficient number of training data, its accuracy is improved. Therefore more training data set is reserved for training the network. In this paper, the total data set is divided into two set as training data and testing data. Out of the total data set 85 % is reserved and used for training session and remaining 15% for testing the network. The classification accuracy obtained is 87.8%**.** The classifier's performance is acceptable good as the overall precision and sensitivity of all the classes score 87.67 % and 88.07% respectively. The classifier correctly recognizing the unknown speaker when testing with his/her input speech is also shown in the simulation result. The maximum score in the diagonal element of confusion matrix will always corresponds to the right speaker as shown from Figure 10 to Figure 14.

REFERENCES

1. Abdelmajid Hassan Mansour[1], Gafar Zen Alabdeen Salh[2], HozayfaHayderZeen Alabdeen[3] "Voice recognition Using back propagation algorithm in neural networks" International Journal of Computer Trends and Technology (IJCTT) ISSN: 2231-2803 volume 23 Number 3, pp 132-139 May 2015.
2. AbhinavAnand, RuggeroDonidaLabati, MadasuHanmandluy, Vincenzo Piuri_, Fabio Scotti, "Text-Independent Speaker Recognition for Ambient Intelligence Applications by Using Information Set Features" IEEE International Conference on Computational Intelligence and virtual Environments for Measurement system and Applications (CIVEMSA),Year: 26-28 June, 2017, Pages: 30 – 35.
3. Bhargab Medhi1, Prof. P.H.Talukdar2 "Assamese Speaker Recognition Using ArtificialNeural Network" International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 3, pp 321-324, March 2015.
4. Dany Ishac1, 3, Antoine Abche2 and Elie Karam2, Georges Nassar3, Dorothée Callens3, "A Text-dependent Speaker-Recognition System" 2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Year: 22-25 May 2017 page 1-6, IEEE Conference Publications.
5. E.C. Gordon,Signal and Linear System Analysis.John Wiley & Sons Ltd., New York, USA,1998.
6. J Sirisha Devi "Language and Text Independent Speaker Recognition System using Artificial Neural Networks and Fuzzy Logic" International Journal of Recent Technology and Engineering (IJRTE), Volume-7, Issue-6, pp 327-330, March 2019.

7. Jorge MARTINEZ*, Hector PEREZ, Enrique ESCAMILLA, Masahisa Mabo SUZUKI, “Speaker recognition using Mel Frequency Cepstral Coefficients (MFCC) and Vector Quantization (VQ) Techniques” CONIELECOMP 2012, 22nd International Conference on Electrical Communications and Computers, 27-29 Feb. 2012 pages: 248 - 251,IEEE Conference Publications.

8. Khan Suhail Ahmad1, Anil S. Thosar2, Jagannath H. Nirmal3 and Vinay S. Pande4 “A Unique Approach in Text Independent Speaker Recognition using MFCC Feature Sets and Probabilistic Neural Network” 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR) Year: 5-7 June2015 Pages: 1 - 6, IEEE Conference Publications.

9. Koustav Chakraborty, AsmitaTaleleProf.SavithaUpadhya, “Voice Recognition Using MFCC Algorithm” International Journal of Innovative Research in Advanced Engineering (IJIRAE) Volume 1 Issue 10 (November 2014), page 158-161.

10. Md. Ali Hossain1, Md. Mijanur Rahman2, Uzzal Kumar Prodhan3, Md. Farukuzzaman Khan4 “Implementation Of Back-Propagation Neural Network For Isolated Bangla Speech Recognition” International Journal of Information Sciences and Techniques (IJIST) Vol.3, No.4, pp 1-9 July 2013.

11. N.AYSHWARYA[1], G.LOGESHWARI[2], G.S.ANANDHA MALA[3] “FEED FORWARD BACK PROPAGATION NEURAL NETWORK FOR SPEAKER INDEPENDENT SPEECH RECOGNITION”, International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-2, Issue-8, pp 36-39, Aug.-2014.

12. Neha Chauhan “Speaker recognition using pattern recognitionneural network and feedforward neural network, International Journal of Scientific & Engineering Research, Volume 8, Issue 3, pp 1444-1446, March-2017.

13. Parvati J.Chaudhary1, Kinjal M. Vagadia2 “A Review Article on Speaker Recognition with Feature Extraction International Journal of Emerging Technology and Advanced Engineering, Volume 5, Issue 2, February 2015 page 94-97.

14. Roger Achkar*, Mustafa El-Halabi*, ElieBassil*, Rayan Fakhro*, Marny Khalil* “Voice Identity Finder Using the Back Propagation Algorithm of an Artificial Neural Network” Complex Adaptive Systems, Publication 6, Conference Organized by Missouri University of Science and Technology 2016 - Los Angeles, CA.

15. S. S.Wali, S.M. Hatture, S. Nandyal “MFCC Based Text-Dependent Speaker Identification Using BPNN” International Journal of Signal Processing System VOL.3, No.1, pp 30-34, June 2015.

16. Siddhant C. Joshi1, Dr. A.N.Cheeran2, “MATLAB Based Back-Propagation Neural Network for Automatic Speech Recognition” International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3, Issue 7, pp 10498-10504, July 2014.

17. Varun Sharma, Dr. P K Bansal, “A Review On Speaker Recognition Approaches And Challenges” International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 5, May – 2013 page 1581-1588.

18. XinxingJing1, Jinlong Ma2, Jing Zhao3 , Haiyan Yang4 “Speaker Recognition Based on Principal Component Analysis of LPCC and MFCC” 2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC) 5-8 Aug. 2014, Pages: 403 - 408 IEEE Conference Publications.

19. Yi Wang, Dr. Bob Lawlor, “Speaker Recognition Based on MFCC and BP Neural Networks” 28th Irish Signals And systems Conference Year: 20-21 June 2017, Pages: 1 - 4, IEEE Conference publication.

20. Zhang Wanli, Li Guoxin, “The Research of Feature Extraction Based on MFCC for Speaker Recognition” Proceedings of 2013 3rd International Conference on Computer Science and Network Technology, 12-13 Oct. 2013, Pages: 1074 - 107IEEE Conference Publications.