REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 02 Aug 2024, Revised: 5 Sep 2024, Accepted: 10 Sep 2024

Text Dependent Speaker Identification And Intruder Detection System

N K Kaphungkui[1*], Gurumayum Robert Michael[2], N Hemarjit Singh[3], Moirangthem Tiken Singh[4], Rabinder Kumar Prasad[5]

[1*,2,3]Department of Electronics and Communication Engineering, Dibrugarh University, Assam, India
[4,5]Department of Computer Science Engineering, Dibrugarh University, Assam, India

***Corresponding author:** N K Kaphungkui,
Email:- ningshen@dibru.ac.in

Abstract

This research paper will present an automatic Speaker Identification System using MFCC (Mel-Frequency Cepstral Coefficients) and BPNN (Back Propagation Neural Network). The objective of this work is to classify 20 speakers' pattern and to identify each registered speaker correctly while testing with new input speech without any false identification. MFCC is used for the extraction of speech features from each speaker and BPNN is used for identification of the test speaker. The developed classifier model is tested with both registered and unregistered speakers and found that it successfully identifies all the registered speakers correctly and reject the intruder speakers. Scaled conjugate gradient training function is used for training the BPNN. A speech database consisting of 20 speakers is created from a group of 10 male and 10 female speakers with the same sentence spoken twice. The classification accuracy rate obtained from the classification is 92.1% and the correct identification rate obtained is 100%. Matlab simulation tool is used in this work

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

1. Introduction

Speech is the most efficient way of communication between human. In the same way many researchers have already implemented efficient method for the interaction of man and machine successful. Speaker recognition system is one of its applications to identify and verify a speaker by the system. Speaker recognition revealed the identity of speaker by analysing the personal features and characteristics associated with that speaker [15]. For authentication purposes speaker recognition system is widely used [16]. Speaker recognition system process is based on the principle that each and every speaker possesses a unique characteristic which are different from each other [17]. In any automatic speaker recognition system, there are two main modules which govern the whole system, they are the speech features extraction module and speech feature matching module. The two important phases that a Speaker Recognition system undergo are the training phase and testing phase [12]. A work has already reported that for 10 classes' classification. An accuracy of 81.8% is obtained with the combination of MFCC, pitch and rms in feed forward neural network (FFNN) [9]. Artificial Neural Network shows better result in terms of accuracy than fuzzy logic-based systems when a speech is recorded in a noiseless environment. Accuracy obtained with ANN is 74% against 72% with fuzzy logic [3]. In text dependent speaker recognition system of 10 speakers' an accuracy of 92% is achieved with the combination of MFCC and BPNN [11]. A moderate accuracy for 10 speakers is also achieved with the combination of LPC and MFCC using Artificial Neural Network for Assamese Speaker Recognition [1]. Speaker recognition system is made and presented by using Discrete Wavelet Transform as speech features extractor and Gaussian Mixture Models (GMM) as classifier. The accuracy result obtained under this work is 96.18% for 32 set speaker which is good enough [13]. A combination of Praat and Matlab for computing MFCCs is also reported. The voice samples are initially denoised using praat. When extracting the MFCC coefficients Delta energy function is taken into account. This draws a conclusion that MFCC coefficient can be increased according to one's requirement. A 39 MFCC coefficients is extracted by adding velocity and acceleration [14]. A real-time text dependent speaker recognition and authentication system is presented with 16 MFCC coefficients along with the derived Delta and Delta Delta Coefficients for speech feature extraction and Dynamic Time Warping as speech classifier [18]. This work will be implemented for 20 speakers with the combination of MFCC and BPNN to achieve a classification accuracy rate of 92.1% and correct identification rate of 100%.

2. Mel Frequency Cepstral Coefficient

MFCC is the most widely used for the extraction of speech features. MFCC becomes the most popular speech feature extraction technique due to its fewer complexes in implementation and more effective [19], [20]. High success rate is obtained with MFCC due to the fact that it is modelled as human auditory system. It fails to perceived signal over 1KHz and showing more robust against noisy environment [2], [5]. Mel Frequency Cepstral Coefficients generates the voice signal coefficients which are unique to every individual speaker [6]. The overall MFCC's steps for the extraction of speech feature are shown in Figure. 1 [4].

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 02 Aug 2024, Revised: 5 Sep 2024, Accepted: 10 Sep 2024

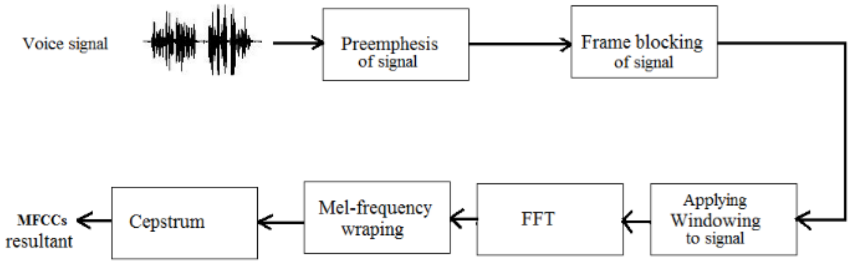


Figure. 1 Block diagram of MFCC computation.

A speech data base of 20 different speakers is created which consist of 10 female speakers label as f1, f2, f3, f4, f5, f6, f7, f8, f9,f10 and 10 male speakers as m1, m2, m3, m4, m5, m6, m7, m8, m9 and m10. The voice samples are collected in a relatively noise free environment. All the collected speech samples are process with MFCC methods and represented with a unique 13 coefficients for each speaker following a 13-order MFCC. The MFCC resultants which is in matrix form consisting of fix 13 rows and variable column for all the speakers as shown in Figure. 2. This can be represented as M x N where M is fix 13 rows and N is variable columns. The N values will be more when a speaker takes more time to utter the given sentence and vice versa.

Speakers	MFCCs Dimension
m1	13 x 731
m2	13 x 545
m3	13 x 612
m4	13 x 703
m5	13 x 618
m6	13 x 988
m7	13 x 712
m8	13 x 548
m9	13 x 540
m10	13 x 550

Speakers	MFCCs Dimension
f1	13 x 671
f2	13 x 787
f3	13 x 737
f4	13 x 659
f5	13 x 861
f6	13 x 750
f7	13 x 833
f8	13 x 734
f9	13 x 767
f10	13 x 1062

Figure.2 MFCC result for different speakers

3. Back Propagation Neural Network

The structure of multi-layer Perceptron neural network is shown in Figure. 4. It consists of 3 layers as input layer, hidden layer and output layer. The Back propagation algorithm is used to train a neural network through a chain rule method. It consists of 2 basic steps which are forward pass of the inputs through the network and back propagation which execute a backward pass by adjusting the parameters of the network. In feed forward direction the input data propagates towards the output node through the hidden layer along with the assigned initial model's parameters i.e. weights and bias. If the output produced by the network is not equal to the set target, then back propagation process will take place by updating the network parameters backward from output node towards input node. Initially when the model is designed, any random values of weights and bias are assigned to predict the set target. After the forward pass execution if there exist a variation between the network's output and the set target, the parameters are updated backward to minimize this huge error. The particular weights and bias which result in minimizing the error function between the output and the set target is the solution of the network leaning. The flowchart of a Back Propagation Algorithm is shown in Figure. 3. The objective of a back propagation algorithm is to minimize the mean square error (MSE) functions between the actual output and the desire set target by updating the parameters of the network [7], [8]. The final network parameters which minimize the MSE function is the solution of neural network learning algorithm. Accuracy and performance of the network is better when a greater number of training data set is used for training the model [10].

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 02 Aug 2024, Revised: 5 Sep 2024, Accepted: 10 Sep 2024

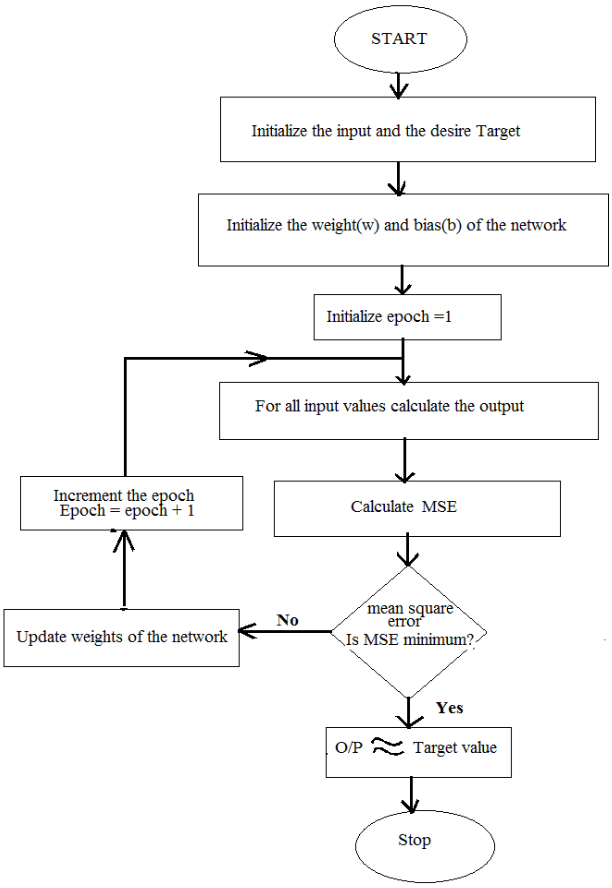


Figure.3 Flowchart of BPNN

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

4. Implementation and Simulation result

The classifier model is implemented with multi-layer perceptron neural network consisting of 13 input nodes, 330 hidden neurons and 20 output nodes for 20 class's classification as shown in Figure 4. The input data for the neural network is fed from the MFCCs result of twenty different persons. The 13 order MFCC of each speaker are concatenated and gave as input data to the neural network. The result input data in matrix form is 13 x 14408 and the desire target is also set at 20 x 14408. The simulation result confusion matrix of 20 class's pattern classification is shown in Figure. 5. The accuracy of the classification is 92.1% and its expression is given in equation 1.

Accuracy = (Sum of diagonal elements)/(Sum of entire elements) (1)

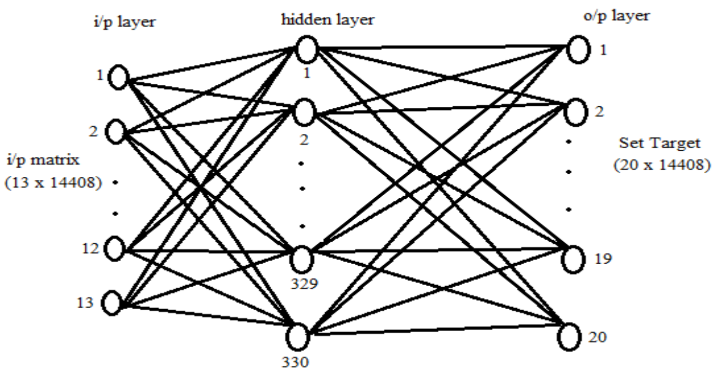


Figure. 4 Speech Classifier Model

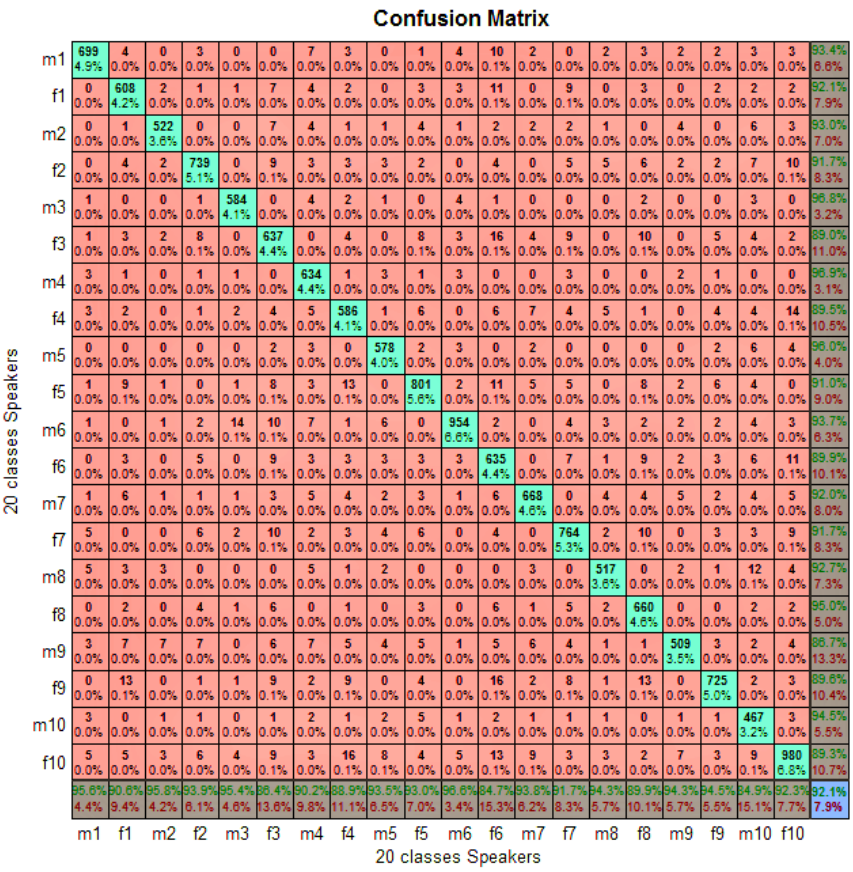


Figure 5. Conclusion Matrix

The Receiver output characteristic (ROC) is used for analysing the classification accuracy. It summarizes the overall performance of the neural network model. It is a graphical representation 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 classification accuracy rate obtained will be high and better will be the classifier's performance. The plot of ROC and the best validation performance at 1000 epochs is also shown in Figure. 6.

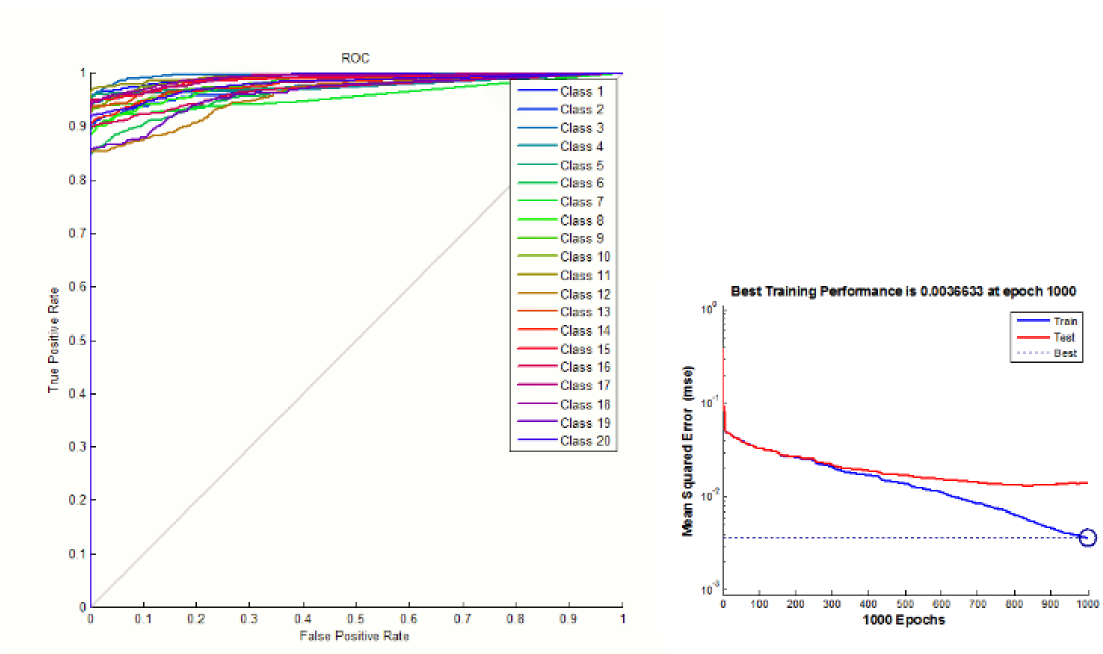


Figure. 6 Classifier's ROC and Validation Performance

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 02 Aug 2024, Revised: 5 Sep 2024, Accepted: 10 Sep 2024

After classification of the registered speakers, the classifier model is ready for testing with new input data's and identify the speaker. For this process, the voice samples of all registered speakers are collected again with the same sentence which is used while training. The MFCCs for each speaker are computed again and given as new testing input data to the model which is shown in Figure 7.

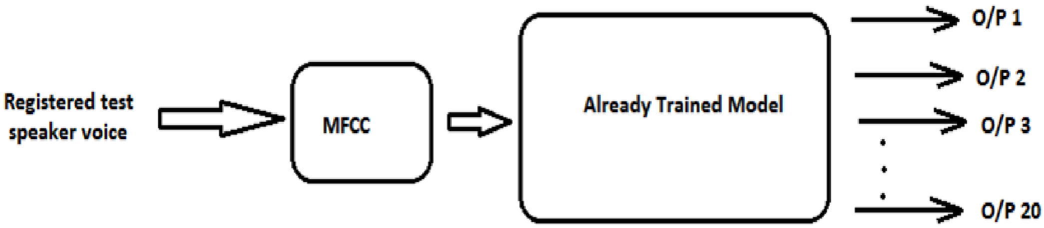


Figure .7 Testing of the model to identify the speaker

The method of how the developed model will identify the speaker correctly will base on which o/p node the maximum score belongs. The 20 o/p nodes of 20 speakers will correspond to 20 diagonal cells of the resultant confusion matrix. The diagonal elements score of the confusion matrix while testing the model with different speakers one at a time is shown from Figure. 8 to Figure. 12. The diagonal elements corresponding to the right speaker always score maximum while another score is minimum. Therefore, by examining the row and column to which speaker this maximum score belongs, the speaker can be successfully recognized without any ambiguity.

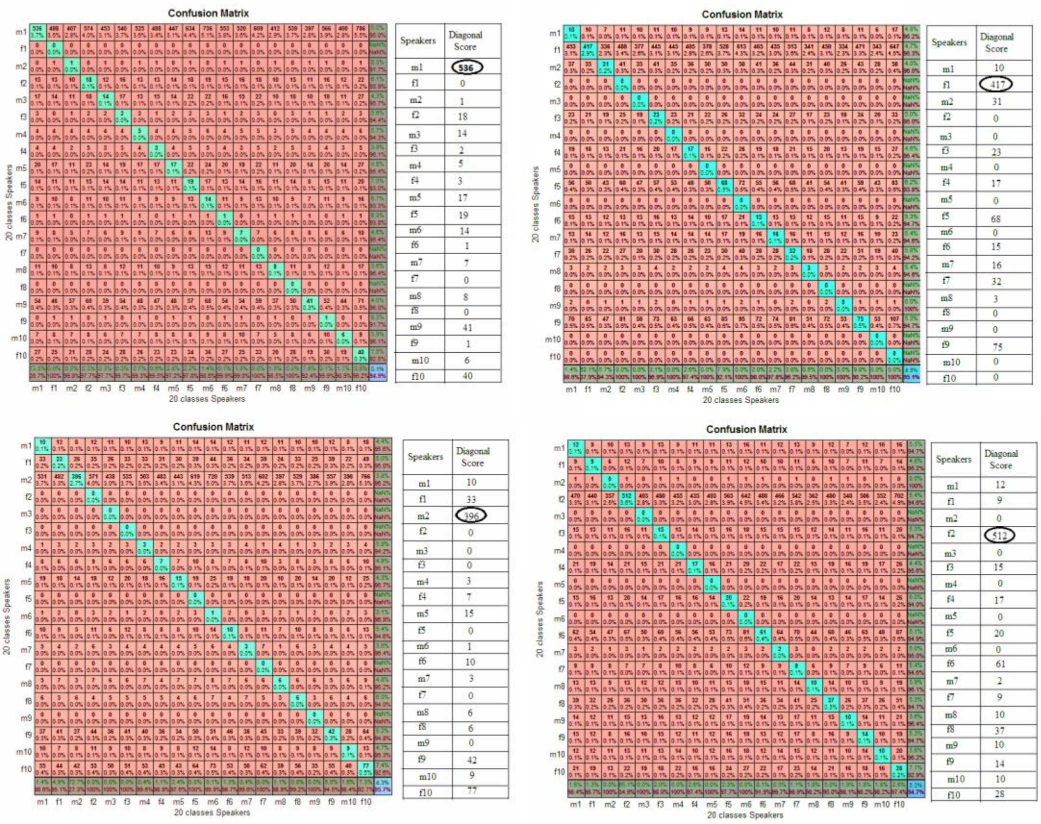
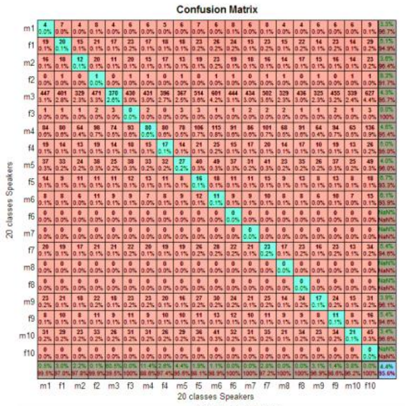
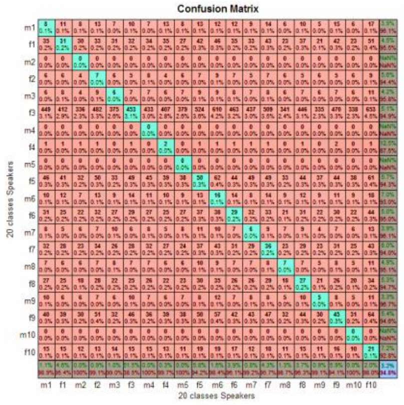


Figure. 8 Confusion matrix result while testing the model with m1, f1, m2 and f2 speakers.

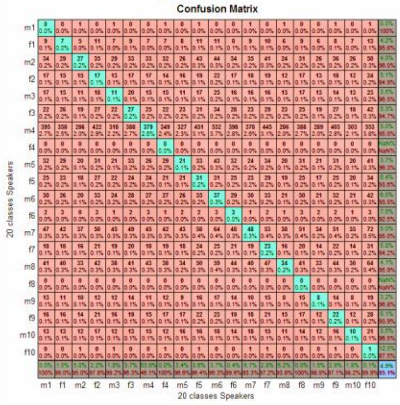
REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 02 Aug 2024, Revised: 5 Sep 2024, Accepted: 10 Sep 2024



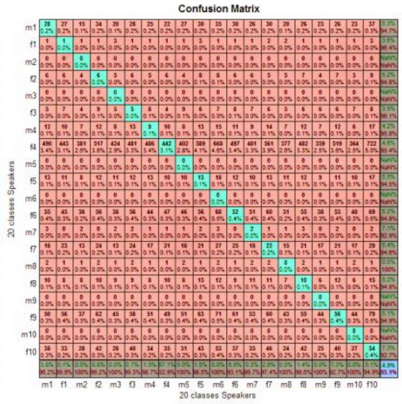
Speakers	Diagonal Score
m1	4
f1	20
m2	12
f2	1
m3	370
f3	0
m4	80
f4	17
m5	27
f5	16
m6	11
f6	0
m7	0
f7	23
m8	0
f8	0
m9	17
f9	11
m10	21
f10	0



Speakers	Diagonal Score
m1	8
f1	31
m2	0
f2	7
m3	6
f3	453
m4	0
f4	2
m5	0
f5	50
m6	16
f6	29
m7	6
f7	36
m8	7
f8	27
m9	5
f9	45
m10	0
f10	21

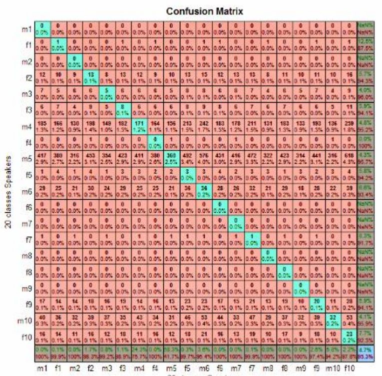


Speakers	Diagonal Score
m1	0
f1	7
m2	27
f2	17
m3	11
f3	27
m4	379
f4	0
m5	21
f5	31
m6	37
f6	0
m7	48
f7	23
m8	34
f8	0
m9	8
f9	22
m10	10
f10	1

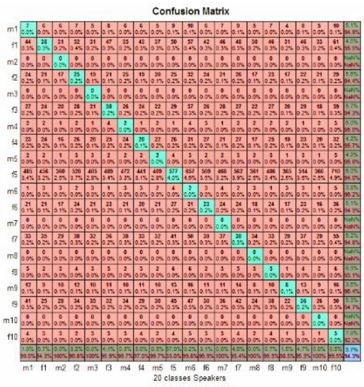


Speakers	Diagonal Score
m1	28
f1	1
m2	0
f2	6
m3	0
f3	5
m4	9
f4	442
m5	0
f5	13
m6	0
f6	52
m7	2
f7	22
m8	0
f8	10
m9	0
f9	56
m10	0
f10	54

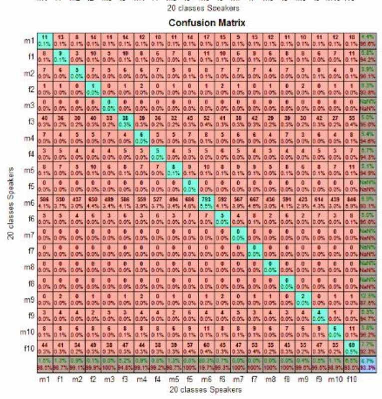
Figure. 9 Confusion matrix result while testing the network with m3, f3, m4 and f4 speakers.



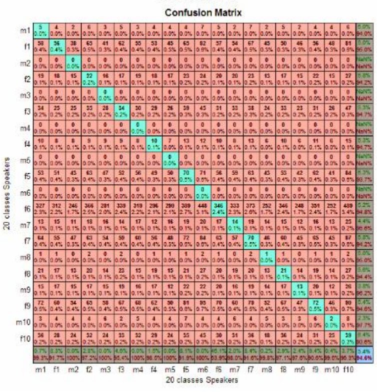
Speakers	Diagonal Score
m1	0
f1	1
m2	0
f2	13
m3	5
f3	8
m4	171
f4	0
m5	360
f5	3
m6	36
f6	0
m7	0
f7	1
m8	0
f8	0
m9	0
f9	20
m10	32
f10	23



Speakers	Diagonal Score
m1	7
f1	38
m2	0
f2	25
m3	0
f3	30
m4	2
f4	20
m5	2
f5	577
m6	2
f6	23
m7	0
f7	38
m8	0
f8	5
m9	8
f9	36
m10	0
f10	5



Speakers	Diagonal Score
m1	11
f1	9
m2	5
f2	1
m3	0
f3	38
m4	6
f4	5
m5	8
f5	0
m6	791
f6	5
m7	0
f7	0
m8	0
f8	0
m9	2
f9	4
m10	6
f10	69



Speakers	Diagonal Score
m1	5
f1	56
m2	0
f2	22
m3	0
f3	34
m4	0
f4	10
m5	0
f5	70
m6	0
f6	546
m7	14
f7	70
m8	1
f8	21
m9	13
f9	72
m10	2
f10	39

Figure. 10 Confusion matrix result while testing the network with m5, f5, m6, f6 speakers.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 02 Aug 2024, Revised: 5 Sep 2024, Accepted: 10 Sep 2024

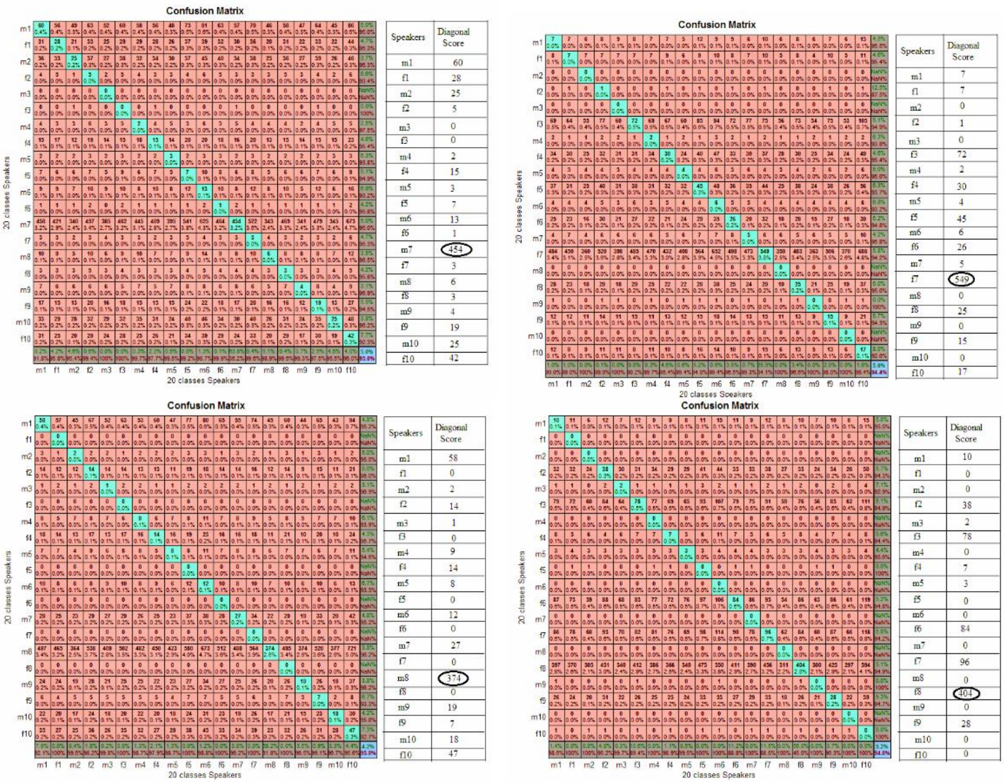


Figure. 11 Confusion matrix result while testing the model with m7, f7, m8 and f8 speakers.

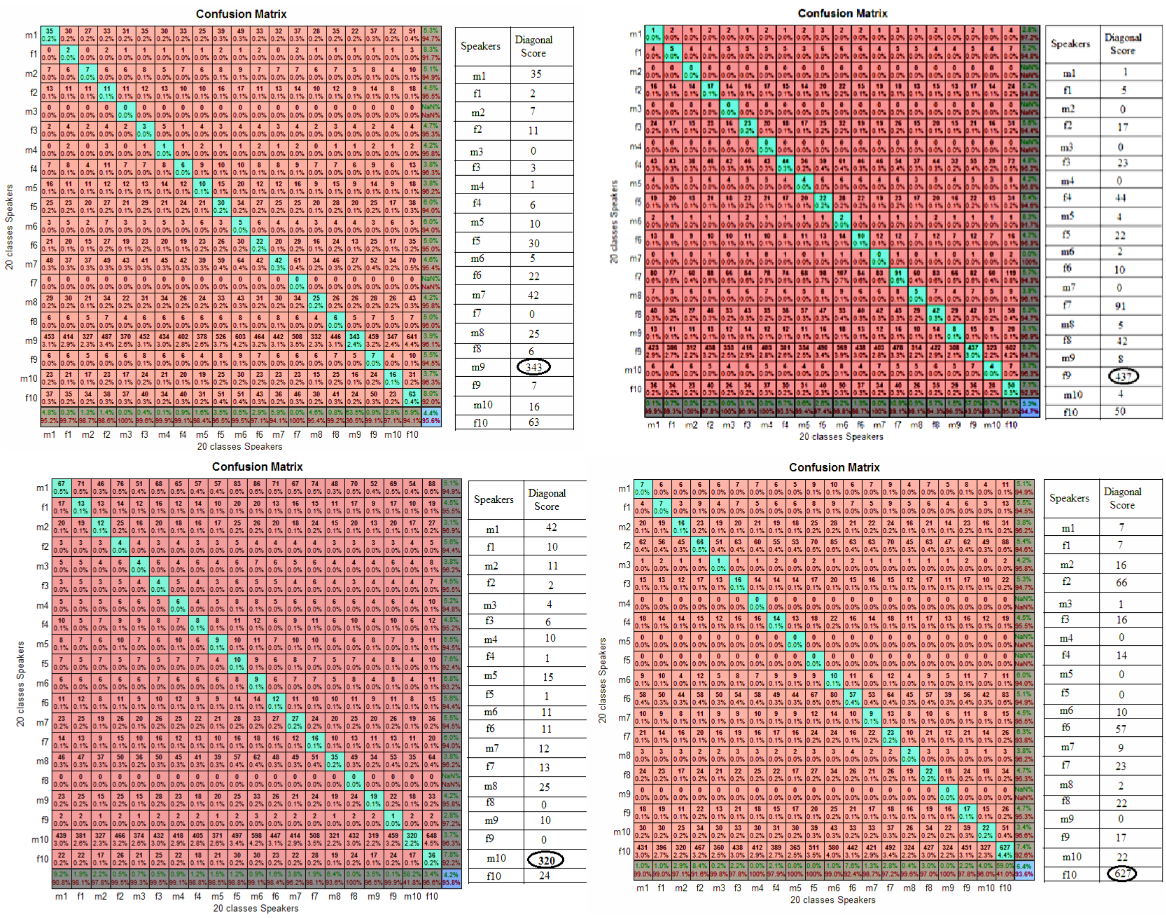


Figure. 12 Confusion matrix result while testing the model with m9, f9, m10 and f10 speakers.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 02 Aug 2024, Revised: 5 Sep 2024, Accepted: 10 Sep 2024

The identification accuracy of the developed model is 100% as all 20 registered speakers are correctly identified while testing the network. The comparison of the classification score of each speaker and the testing score of the classifier are also shown in Figure 13 and 14. The testing score percentage is evaluated wrt the classification score. From this result a threshold level is set i.e testing score rate should be at least 50% of the classification score to correctly identified a registered test speaker. If the threshold level score is set too high there is chances of misdetection and if it is set too low there is chances of false detection. Hence testing score rate of 50% of the classification score is chosen in this work.

Speakers	Classification Score	Testing Score	% of Testing Score
m1	699	536	76.7
m2	522	396	75.8
m3	584	370	63.3
m4	634	379	59.7
m5	578	360	62.2
m6	954	793	83.1
m7	668	454	67.9
m8	517	374	72.3
m9	509	343	67.3
m10	467	320	68.5

Figure. 13 Testing score rate of male speakers

Speakers	Classification Score	Testing Score	% of Testing Score
f1	608	417	68.5
f2	739	512	69.2
f3	637	453	71.1
f4	586	442	75.4
f5	801	577	72
f6	635	346	54.4
f7	764	549	71.8
f8	660	404	61.2
f9	725	437	60.2
f10	980	627	63.9

Figure. 14 Testing score rate of female speakers

The performance of speech classifier's is best evaluated not only by its Classification Accuracy rate but along with three more parameters. They are Precision, Sensitivity and Specificity which are express as

$$\text{Precision} = TP/(FP + TP). \quad (2)$$
$$\text{Sensitivity} = TP/(FN + TP). \quad (3)$$
$$\text{Specificity} = TN/(TN + FN) \quad (4)$$

Where TP is True Positive, FP is False Positive, TN is true negative and FN is False Negative.
The score of each male speaker with respect to performance parameters is also shown in Figure. 15 and its graphical plot is shown in Figure. 16. From the result it is found that Specificity of all male speakers score above 99% and varies from 99.4% to 99.83%. In terms of sensitivity, it varies from 84.9% to 96.6% whereas precision varies from 86.7% to 96.9%.

Speakers	Precision
m1	93.4%
m2	93%
m3	96.8%
m4	96.9%
m5	96%
m6	93.7%
m7	92%
m8	92.7%
m9	86.7%
m10	94.5%

Speakers	Sensitivity
m1	95.6%
m2	95.8%
m3	95.4%
m4	90.2%
m5	93.5%
m6	96.6%
m7	93.8%
m8	94.3%
m9	94.3%
m10	84.9%

Speakers	Specificity
m1	99.76%
m2	99.83%
m3	99.79%
m4	99.49%
m5	99.71%
m6	99.74%
m7	99.67%
m8	99.77%
m9	99.77%
m10	99.40%

Figure .15 Performance parameters score of male speakers

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 02 Aug 2024, Revised: 5 Sep 2024, Accepted: 10 Sep 2024

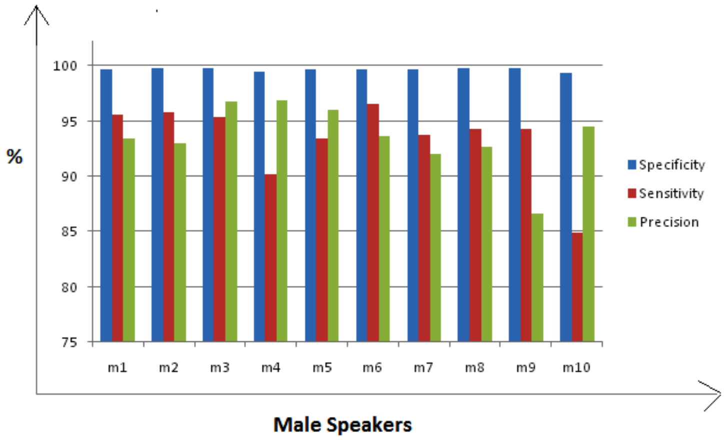


Figure. 16 Graphical plot of performance parameters of male speakers.

In similar manner, for female speakers the classifier performance parameters score is shown in Figure. 17 and its graphical representation is shown in Figure 18.

Speakers	Precision
f1	92.1%
f2	91.7%
f3	89%
f4	89.5%
f5	91%
f6	89.9%
f7	91.7%
f8	95%
f9	89.6%
f10	89.3%

Speakers	Sensitivity
f1	90.6%
f2	93.9%
f3	86.4%
f4	88.9%
f5	93%
f6	84.7%
f7	91.7%
f8	89.9%
f9	94.5%
f10	92.3%

Speakers	Specificity
f1	99.54%
f2	99.64%
f3	99%
f4	99.46%
f5	99.55%
f6	99.16%
f7	99.49%
f8	99.46%
f9	99.69%
f10	99.38%

Figure. 17 Performance parameters score of female speakers.

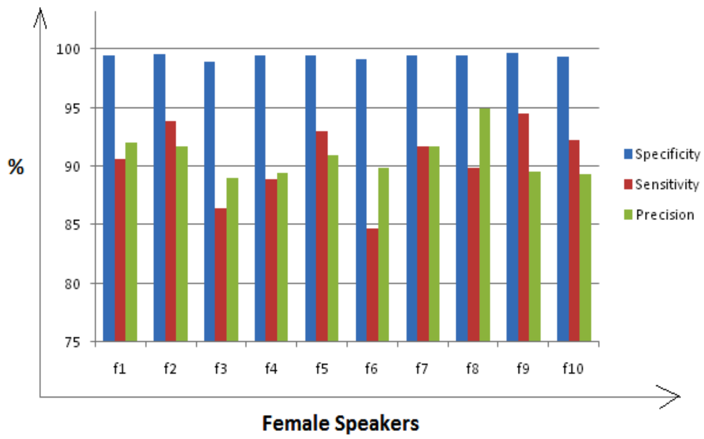


Figure. 18 Graphical plot of performance parameters of female speakers.

5. Testing the classifier model with intruder speaker

To test the trained model to detect intruder, 4 voice sample consisting of 2 male and 2 female unregistered speakers are collected again with the same sentence.The MFCC features of all the sample are extracted and gave as testing input data to the BPNN. The resultant maximum score in one cell of the confusion matrix while testing with 4 unregistered speaker corresponds to f1, m6, m4, m6 is shown in Figure 18.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 02 Aug 2024, Revised: 5 Sep 2024, Accepted: 10 Sep 2024

Test Speakers	Classification Score	Testing Score	% of Testing Score
Test speaker 1	f(1)608	415	23.8
Test speaker 2	m(6)954	189	19.8
Test speaker 3	m(4)634	189	29.8
Test speaker 4	m(6)954	176	18.4

Figure. 18 Testing score rate obtained for unregistered speakers

Here test speaker 1 and test speaker 2 are female unregistered speaker and test speaker 3 and test speaker 4 are male unregistered speaker. While testing the model with test speaker 1 and 2, the maximum score corresponds to f1 and m6 cell. In order to correctly identify the registered speaker, the obtained maximum score of the diagonalelement should be at least 50% of the classification score. As both the test speaker score 23.8% and 19.8% in the diagonal element, they will be rejected and label as intruder with false identity as the score is below 50% of the classification score. Similarly,the classifier is tested again with 2 male unregistered speaker and the maximum score belongs to m4 and m6 cell. These 2 male speakers will also be rejected as the diagonal scoreis only 29.8% and 18.4% which is below 50% of the classification score. The proposed model to either accept and identify the test speaker or reject and labelas intruder speaker is shown in Figure 19. If the output score corresponding to the diagonal cell of confusion matrix is at least 50% of the classification score it will correctly identify the test speaker. But if the output score is below 50% of the classification score, it will label as intruder speaker.

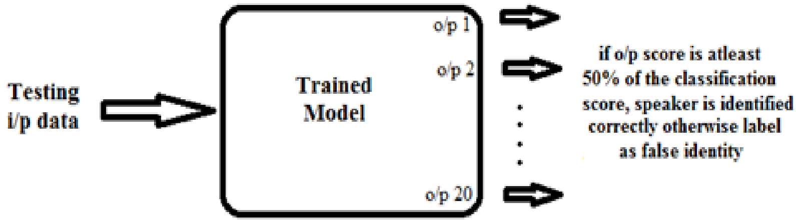


Figure. 19 Classifier Model to identify or reject the test speaker.

6.Conclusion

When ample number of training dataset is given to the neural network, it will learn the nature of its input data's correctly for its classification by improving its performance in terms of classification accuracy. Hence out of total number of dataset available, a greater number of data set is reserved for training the neural network. Here out of the total input data set, 85 % is used for training and remaining 15% is used for testing the network. The speech classifier classification accuracy obtained is 92.1% with 7.9% misclassification. The classifier's performance is acceptable high and good as the overall precision, sensitivity and specificity of all the twenty classes score 92.22%, 92.02% and 99.56% respectively. After classification is done, this trained classifier will correctly identify all the registered speaker correctly when testing with its new input speech. Finally, from the experiment a threshold level is set i.e testing score rate should be at least 50% of the classification score to correctly identified a registered test speaker. If the testing score rate is below 50% then the test speaker will be label as intruder. Hence from this work, a correct speaker identification rate of 100% is obtained with the detection of intruder.

References

1. Bhargab Medhi1, Prof. P.H.Talukdar2 "Assamese Speaker Recognition Using Artificial Neural Network" International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 3, pp 321-324, March 2015.
2. E.C. Gordon,Signal and Linear System Analysis.John Wiley & Sons Ltd., New York, USA,1998.
3. 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.
4. 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
5. 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.

6. Koustav Chakraborty, AsmitaTalele Prof.SavithaUpadhya, "Voice Recognition Using MFCC Algorithm" International Journal of Innovative Research in Advanced Engineering (IJIRAE) Volume 1 Issue 10 (November 2014), page 158-161.
7. 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.
8. N.AYSHWARYA1,G.LOGESHWARI2,G.S.ANANDHA MALA3 "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.
9. Neha Chauhan "Speaker recognition using pattern recognition neural network and feedforward neural network, International Journal of Scientific & Engineering Research, Volume 8, Issue 3, pp 1444-1446, March-2017.
10. 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.
11. 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.
12. 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.
13. Sajeer, K., & Rodrigues, P.. "Speaker Recognition System Based on Wavelet Features and Gaussian Mixture Models". International Journal of Engineering and Advanced Technology, 9(1), 5363-5367, 2019.
14. Singh, P. P., & Rani, P. " An approach to extract feature using mfcc" IOSR Journal of Engineering, 4(8), 21-25. 2014
15. Matsui, T., & Tanabe, K. "Comparative study of speaker identification methods: dPLRM, SVM and GMM. IEICE" transactions on information and systems, 89(3), 1066-1073. (2006).
16. Doddington, G. R. "Speaker recognition -Identifying people by their voices" Proceedings of the IEEE, 73(11), 1651-1664. Fang, (1985).
17. Fang, C.. From dynamic time warping (DTW) to hidden markov model (HMM). University of Cincinnati, 3, 19. (2009).
18. Rakesh, D. R., & Jayasimhan, N. K. "Speaker recognition and authentication" International Journal of Computer Science and Mobile Computing, 2(5), 402-407. 2013.
19. Maesa, A., Garzia, F., Scarpiniti, M., & Cusani, R. " Text independent automatic speaker recognition system using mel-frequency cepstrum coefficient and gaussian mixture models" Journal of Information Security, 3(04), 335-340. 2012
20. Reynolds, D. A. (2002, May). An overview of automatic speaker recognition technology. In 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 4, pp. IV-4072). IEEE