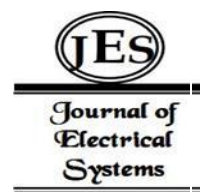


¹ N K Kaphungkui^{1*}Gurumayum Robert
Michael²Hemarjit
Ningombam³Moirangthem Tiken
Singh⁴Rabinder Kumar
Prasad⁵

A Novel Approach for Text-Dependent Speaker Identification System



Abstract: - This research paper will present an Automatic Speaker Identification System using MFCC (Mel-Frequency Cepstral Coefficients) to extract speech features and BPNN (Back Propagation Neural Network) to classify speakers. This paper aims to classify 16 speakers' patterns with acceptable classification accuracy rates and identify each registered speaker correctly while testing with new speech data without any false identification of the registered test speaker. MFCC method is used to extract the speech features from each speaker and BPNN is used to identify the registered test speaker. The speech classifier model is built with 13 input nodes with 1 hidden layer consisting of 298 hidden neurons and 16 output nodes for 16 speaker classifications. Initially, the developed classifier model is trained for speaker's pattern classification and later it is tested with all the registered speakers and found that it successfully identifies all the registered speakers correctly achieving a 100% correct identification rate. A scaled conjugate gradient training function will be used for training the BPNN. A speech database consisting of 16 speakers from different age groups is created in a relatively noise-free environment with the same sentence spoken once. The classification accuracy rate obtained from the classification is 92.6%. MATLAB simulation tool is used in the entire work.

Keywords: Speaker recognition, MFCC, BPNN, Training, Testing

1.0 INTRODUCTION

Speech is the most efficient way of communication between humans. In the same way, many researchers have already implemented efficient methods for the interaction of man and machine successfully. A speaker recognition system is one of its applications to identify and verify a speaker in the system. Speaker recognition reveals the identity of a speaker by analyzing the personal features and characteristics associated with that speaker [1]. For authentication purposes speaker recognition system is widely used [2]. The speaker recognition system process is based on the principle that every speaker possesses a unique characteristic that is different from each other [3]. In any automatic speaker recognition system, 2 main modules govern the whole system. They are the extraction of unique speech features and the matching of speech features. The 2 important steps that a Speaker Recognition system undergo are the training and testing phase [4]. A work has already reported that for 10 classes' classification an accuracy of 81.8% is obtained with the combination of MFCC, pitch and root mean square in feed forward neural network [5]. Artificial Neural Networks give better results of accuracy rate than fuzzy logic-based systems when a speech is recorded in a quite environment. Accuracy obtained with ANN is 74% against 72% with fuzzy logic [6]. In text-dependent speaker recognition system of 10 speakers' an accuracy of 92% is achieved with the combination of MFCC and BPNN [7]. A moderate accuracy for 10 speakers is also achieved with the combination of LPC and MFCC using Artificial Neural Network for Assamese Speaker Recognition [8]. The speaker recognition system is made and presented by using Discrete Wavelet Transform as speech features extractor and Gaussian Mixture Models (GMM) as a classifier. The accuracy result obtained under this work is 96.18% for 32 set speaker which is good enough [9]. 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 concludes that MFCC

^{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

Copyright © JES 2024 on-line : journal.esrgroups.org

coefficient can be increased according to one’s requirements. A 39 MFCC coefficient is extracted by adding velocity and acceleration [10]. 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 [11]. This work will be implemented for 16 speakers with the combination of MFCC and BPNN to achieve a classification accuracy rate of 92.6% and a correct identification rate of 100%.

2.0 MEL-FREQUENCY CEPSTRAL COEFFICIENTS

MFCC is used widely for speech feature extraction methods. MFCC becomes the most popular speech feature extraction technique due to its fewer complexes in implementation and more effective [12], [13]. Every individual in terms of their speeches is not identical because of their different vocal tract shapes, variation in larynx sizes and other parts of the voice production organs [14]. Feature extraction is a method used for finding the features of speech that are spoken or uttered by different speakers. The feature extraction captures the specific properties of a speaker by data reduction process [15]. High success rate is obtained with MFCC as it is modeled similarly to the human auditory system which fails to perceived signal over 1000Hz [16], [17]. Mel Frequency Cepstral Coefficients generate the voice signal coefficients which are unique to every individual speaker [18].

The sequences of MFCC computation for speech feature extraction technique are shown in Figure. 1 [19].

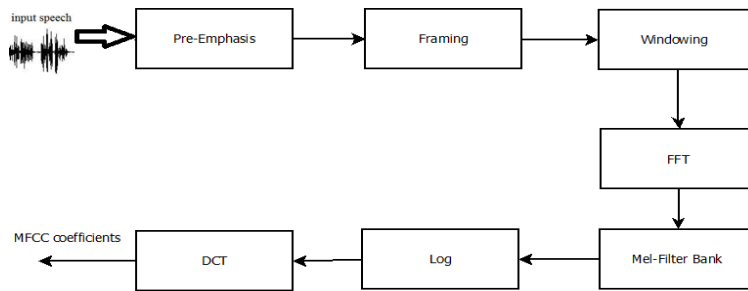


Figure. 1 MFCC block diagram

The MFCC algorithm will convert any speech wav file into 13 MFCC coefficients for each speaker. The obtained MFCC features are in matrix order consisting of 13 rows and variable columns for all the registered speakers. It can be represented as (A x B) where A is fixed 13 rows and B is variable columns. The values of B will vary because all speakers will take different time to utter the given sentence. A speech database of 16 different speakers is created with the same sentence spoken once. The voice samples are collected with the same recording device in a relatively less noisy environment. All the collected speech samples are processed with MFCC methods and represented with a unique 13 coefficients for each speaker following a 13-order MFCC. The simulation result of MFCC for a speaker sp1 is shown in Figure 2.

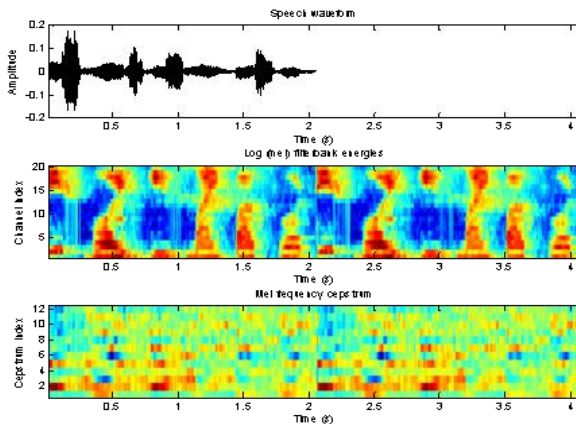


Figure.2 Mel frequency Cepstral coefficient

Similarly, for all the speakers, the MFCC is computed and represented with a fixed 13 rows and variable column as shown in Figure. 3

Speakers	MFCCs Dimension	Speakers	MFCCs Dimension
sp1	13 x 324	sp9	13x 358
sp2	13 x 361	sp10	13 x 361
sp3	13 x 251	sp11	13 x 289
sp4	13 x 404	sp12	13 x 361
sp5	13 x 342	sp13	13 x 258
sp6	13 x 342	sp14	13 x 256
sp7	13 x 407	sp15	13 x 404
sp8	13 x 340	sp16	13 x 409

Figure.3 MFCC order of 16 speakers

3.0 BACK PROPAGATION NEURAL NETWORK

In computer systems, neural networks are general-purpose programs that can learn difficult and non-linear input-output relationships. When these networks are trained extensively enough, they adapt to the data's feeding to them [20]. ANN is a computer architecture that is modeled after the human brain. This network consists of hundreds of simple processing units/neurons that are wired together in a complex inter-connected network. Each of these neurons or processing units can fire or send off a new signal when it receives a strong signal from the other connected neuron [21]. The multi-layer Perceptron neural network is shown in Figure. 5. 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 from the input nodes pass towards the output node through the hidden layer along with the assigned weights and bias of the model. If the network output is not equal to the set target, the backpropagation of data takes place by updating the network parameters backward from the output node toward the 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 exists a variation between the network's output and the set target, the parameters are updated backward to minimize the error. The particular bias and weight that result in minimizing the error function between the network output and the target is the solution of the network leaning. The flowchart of a Back Propagation Algorithm is shown in Figure. 4. The objective of a backpropagation algorithm is to minimize the mean square error (MSE) functions between the actual output and the desired set target by updating the parameters of the network [22], [23], [24]. The final network parameters that minimize the MSE function is the solution of the neural network learning algorithm. The accuracy and performance of the network are better when a greater number of training data sets is used for training the model [25].

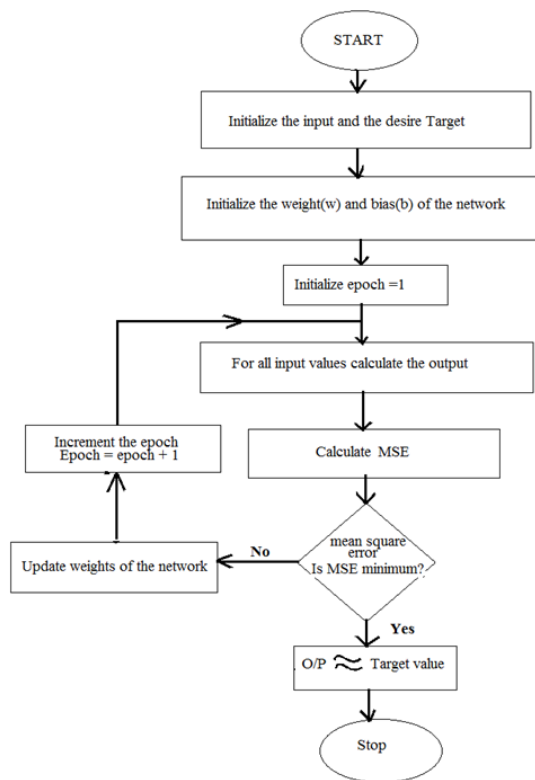


Figure.4 BPNN flowchart

4.0 IMPLEMENTATION AND SIMULATION RESULTS

The speech classifier model as shown in Figure. 5 consists of 13 input nodes, 298 hidden neurons and 16 output nodes for 16 class classifications. The input data for the neural network is fed from the MFCC results of 16 different persons. The 13-order MFCC of each speaker is concatenated and given as input data to the neural network. The result input data in matrix form is 13 x 5467 and the desired target is also set at 16 x 5467.

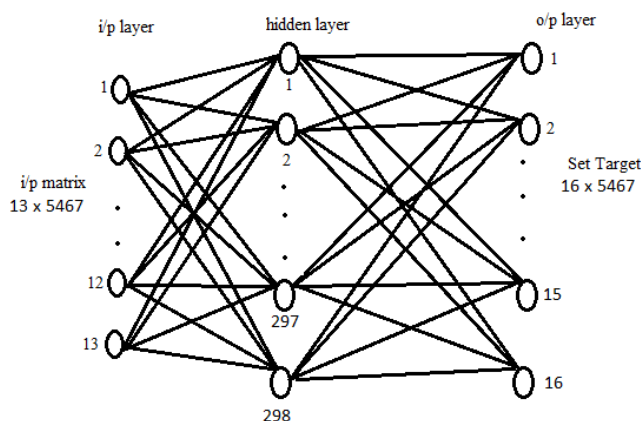


Figure. 5 Speech Classifier Model.

In this experiment the simulation of the classifier model is started by taking number of hidden neurons from 100 onward with a multiple of 4. The obtained classification accuracy rate from the simulation is shown in Table 1 and Table 2. From the result it is found that with 298 hidden neurons the highest classification accuracy rate is obtained. Therefore, the classifier model is implemented with 13 i/p nodes, 1 hidden layer of 298 neurons and 16 o/p nodes.

No. of hidden neurons	Classification Accuracy	No. of hidden neurons	Classification Accuracy	No. of hidden neurons	Classification Accuracy	No. of hidden neurons	Classification Accuracy
100	7.8%	102	81.9%	172	70.5%	174	53.4%
104	85.3%	106	82.7%	176	6.6%	178	15.9%
108	84.8%	110	31.8%	180	89.6%	182	87.1%
112	50.2%	114	81.7%	184	23.5%	186	46.6%
116	51.6%	118	86.2%	188	78.6%	190	5.9%
120	11.6%	122	90.8%	192	21.1%	194	87.4%
124	79.1%	126	46.6%	196	20.9%	198	35.2%
128	34.1%	130	68.9%	200	86.6%	202	15.5%
132	11%	134	39.5%	204	72%	206	5.9%
136	10.1%	138	55.8%	208	67.8%	210	30.8%
140	42%	142	49.6%	212	89%	214	90.2%
144	54%	146	84.5%	216	90.5%	218	83.2%
148	86.7%	150	77.8%	220	92%	222	67.4%
152	69.3%	154	83.5%	224	51%	226	37.3%
156	39.8%	158	82.3%	228	37.8%	230	15.8%
160	74.4%	162	43.4%	232	35%	234	80.7%
164	6.2%	166	57.2%	236	90.1%	238	88.4%
168	88.2%	170	49%	240	70.1%	242	36.4%
				244	22.9%	246	90.0%

Table 1. Variation of Classification accuracy rate wrt number of hidden neurons

No. of hidden neurons	Classification Accuracy	No. of hidden neurons	Classification Accuracy
248	85%	250	31.3%
252	17.4%	254	7.6%
256	85%	258	88.9%
260	47.8%	262	41%
264	37.9%	266	66.1%
268	75.7%	270	57.4%
272	83.7%	274	26.9%
276	85.4%	278	33%
280	87.3%	282	22.6%
284	84.4%	286	37.4%
288	84.4%	290	64.2%
292	66.1%	294	12.1%
296	32.8%	298	92.6%
300	62.7%	302	78.1%
304	72.1%	306	53%
308	42.1%	310	90.8%
312	4.7%	314	83.9%
316	90.9%	318	46.9%

Table 2. Variation of Classification accuracy rate wrt number of hidden neurons

The sequence of steps for training the BPNN using Scaled Conjugate gradient are listed below.

Step 1: A speech database for 16 people is created in a noise-free environment.

Step 2: MFCC feature vector consisting first 13 coefficients is computed from all samples collected in Step 1. These are randomly divided into 85% of the training set, and 15% for the validation set.

Step 3: Training of BPNN is done by taking input MFCC feature vector randomly from 85% of the sample computed in Step 2.

Step 4: Validation of BPNN is done by taking input MFCC feature vector randomly from 15% of samples computed in Step 2.

Step 5: Repetition of Steps 3 and 4 are fixed at 1000 epochs by randomly dividing 100% of the MFCC feature vector into 85% of the training set and 15% into validation.

Step 6: Steps 2, 3, 4, 5 are repeated and at 453 epoch the best validation performance is obtained and finally average confusion matrix is computed.

Step 7: To identify the speaker with new inputs, the column index of Diagonal Square having the highest score is taken as the identified person for the input taken from the person given by the row index.

The simulation result of the confusion matrix for 16 speaker pattern classification is shown in Figure. 6.

Confusion Matrix

20 classes Speakers	sp1	296 5.4%	4 0.1%	0 0.0%	0 0.0%	3 0.1%	2 0.0%	5 0.1%	0 0.0%	2 0.0%	0 0.0%	2 0.0%	1 0.0%	0 0.0%	12 0.2%	0 0.0%	90.5%	
	sp2	2 0.0%	319 5.8%	1 0.0%	3 0.1%	8 0.1%	0 0.0%	3 0.1%	1 0.0%	0 0.0%	1 0.0%	3 0.1%	2 0.0%	0 0.0%	0 0.0%	0 0.0%	92.7%	
	sp3	0 0.0%	0 0.0%	246 4.5%	0 0.0%	1 0.0%	6 0.1%	1 0.0%	0 0.0%	3 0.1%	0 0.0%	2 0.0%	0 0.0%	6 0.1%	0 0.0%	0 0.0%	92.6%	
	sp4	1 0.0%	3 0.1%	0 0.0%	381 7.0%	0 0.0%	0 0.0%	0 0.0%	2 0.0%	1 0.0%	3 0.1%	0 0.0%	0 0.0%	4 0.1%	4 0.1%	1 0.0%	95.3%	
	sp5	0 0.0%	0 0.0%	2 0.0%	2 0.0%	299 5.5%	4 0.1%	2 0.0%	0 0.0%	3 0.1%	3 0.1%	2 0.0%	4 0.1%	0 0.0%	0 0.0%	4 0.1%	91.7%	
	sp6	1 0.0%	0 0.0%	1 0.0%	4 0.1%	5 0.1%	303 5.5%	2 0.0%	0 0.0%	8 0.1%	0 0.0%	0 0.0%	2 0.0%	0 0.0%	3 0.1%	2 0.0%	91.5%	
	sp7	3 0.1%	4 0.1%	0 0.0%	0 0.0%	8 0.1%	4 0.1%	379 6.9%	2 0.0%	6 0.1%	3 0.1%	0 0.0%	1 0.0%	3 0.1%	1 0.0%	2 0.0%	91.1%	
	sp8	6 0.1%	7 0.1%	0 0.0%	2 0.0%	1 0.0%	3 0.0%	328 6.0%	3 0.1%	0 0.0%	0 0.0%	0 0.0%	4 0.1%	2 0.0%	2 0.0%	0 0.0%	91.5%	
	sp9	4 0.1%	4 0.1%	0 0.0%	2 0.0%	8 0.1%	4 0.1%	3 0.1%	0 0.0%	321 5.9%	0 0.0%	0 0.0%	5 0.1%	1 0.0%	0 0.0%	3 0.0%	90.4%	
	sp10	0 0.0%	0 0.0%	0 0.0%	0 0.0%	3 0.1%	0 0.0%	10 0.2%	0 0.0%	0 0.0%	352 6.4%	0 0.0%	6 0.1%	0 0.0%	2 0.0%	1 0.0%	94.1%	
	sp11	0 0.0%	6 0.1%	1 0.0%	1 0.0%	1 0.0%	2 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	271 5.0%	1 0.0%	3 0.1%	7 0.1%	0 0.0%	91.9%	
	sp12	3 0.1%	2 0.0%	0 0.0%	1 0.0%	3 0.1%	0 0.0%	0 0.0%	1 0.0%	0 0.0%	3 0.1%	2 0.0%	338 6.2%	0 0.0%	2 0.0%	0 0.0%	94.7%	
	sp13	1 0.0%	3 0.1%	0 0.0%	6 0.1%	0 0.0%	1 0.0%	0 0.0%	1 0.0%	2 0.0%	0 0.0%	2 0.0%	1 0.0%	239 4.4%	1 0.0%	0 0.0%	92.6%	
	sp14	0 0.0%	7 0.1%	0 0.0%	2 0.0%	3 0.1%	2 0.0%	0 0.0%	0 0.0%	2 0.0%	0 0.0%	2 0.0%	0 0.0%	229 4.2%	0 0.0%	1 0.0%	92.3%	
	sp15	7 0.1%	2 0.0%	0 0.0%	2 0.0%	1 0.0%	2 0.0%	5 0.1%	2 0.0%	1 0.0%	0 0.0%	2 0.0%	1 0.0%	0 0.0%	0 0.0%	386 6.7%	93.4%	
	sp16	0 0.0%	0 0.0%	0 0.0%	0 0.0%	3 0.1%	1 0.0%	0 0.0%	1 0.0%	5 0.1%	0 0.0%	2 0.0%	2 0.0%	2 0.0%	2 0.0%	6 0.1%	397 7.3%	
	sp1	91.4% 8.8%	88.4% 11.6%	98.0% 2.0%	94.3% 5.7%	87.4% 12.6%	88.0% 11.4%	93.1% 6.9%	95.5% 4.5%	89.7% 10.3%	97.5% 2.5%	93.8% 6.2%	93.6% 6.4%	92.6% 7.4%	89.6% 10.6%	94.4% 5.6%	90.6% 9.4%	97.1% 2.9%
	sp1	sp2	sp3	sp4	sp5	sp6	sp7	sp8	sp9	sp10	sp11	sp12	sp13	sp14	sp15	sp16		
	16 classes Speakers																	

Figure 6. Conclusion Matrix

The Receiver output characteristic (ROC) of the speech classifier model is used for analyzing the rate of classification accuracy. It is a graphical representation of the True Positive Rate along the positive y-axis against the False Positive Rate along the positive 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 representation of ROC and the least mean square error of the classifier model at 453 epochs is shown in Figure 7.

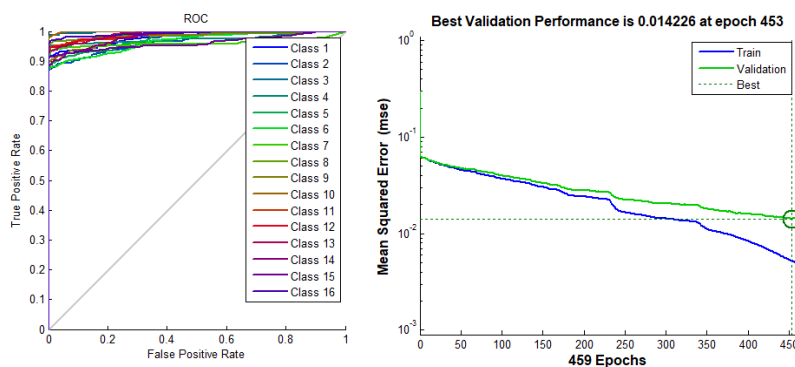


Figure 7 Classifier’s ROC and Validation Performance

After the classification of the registered speakers, the classifier model will be tested with new input data for identifying the registered test speaker. For this step, the voice samples are collected from all the registered speakers, speaking the same sentence which is used while for training sessions. The MFCCs for individual speakers is computed again and given as testing input data to the proposed model as shown in Figure 8.

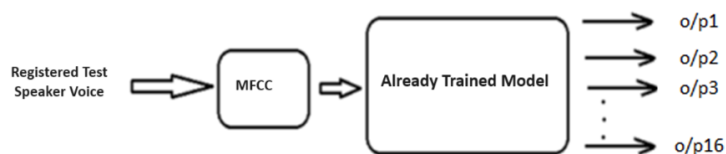


Figure .8 Testing of the model to identify the speaker

The novel method of how the proposed model will identify the speaker correctly will base on which o/p node the maximum score belongs. The 16 o/p nodes will correspond to 16 diagonal cells of the resultant confusion matrix. Here o/p 1 is assigned as sp1 speaker node, o/p 2 as sp2 speaker node, o/p 3 as sp3 speaker node, o/p 4 as sp4 speaker node, o/p 5 as sp5 speaker node, o/p 6 as sp6 speaker node, o/p7 as sp7 speaker node, o/p 8 as sp8 speaker node, o/p 9 as sp9 speaker node, o/p 10 as sp10 speaker node, o/p 11 as sp11 speaker node, o/p 12 as sp12 speaker node, o/p 13 as sp13 speaker node, o/p 14 as sp14 speaker node, o/p 15 as sp15 speaker node, o/p 16 as sp16 speaker node, The diagonal cells score of the confusion matrix when testing the trained model with all the registered speakers one at a time is shown from Figure. 9 to Figure. 12. The maximum score always corresponds to the right person claimed to be. Hence, by examining to which speaker this maximum output score belongs, the speaker which is claimed to be can be identified correctly without any ambiguity.

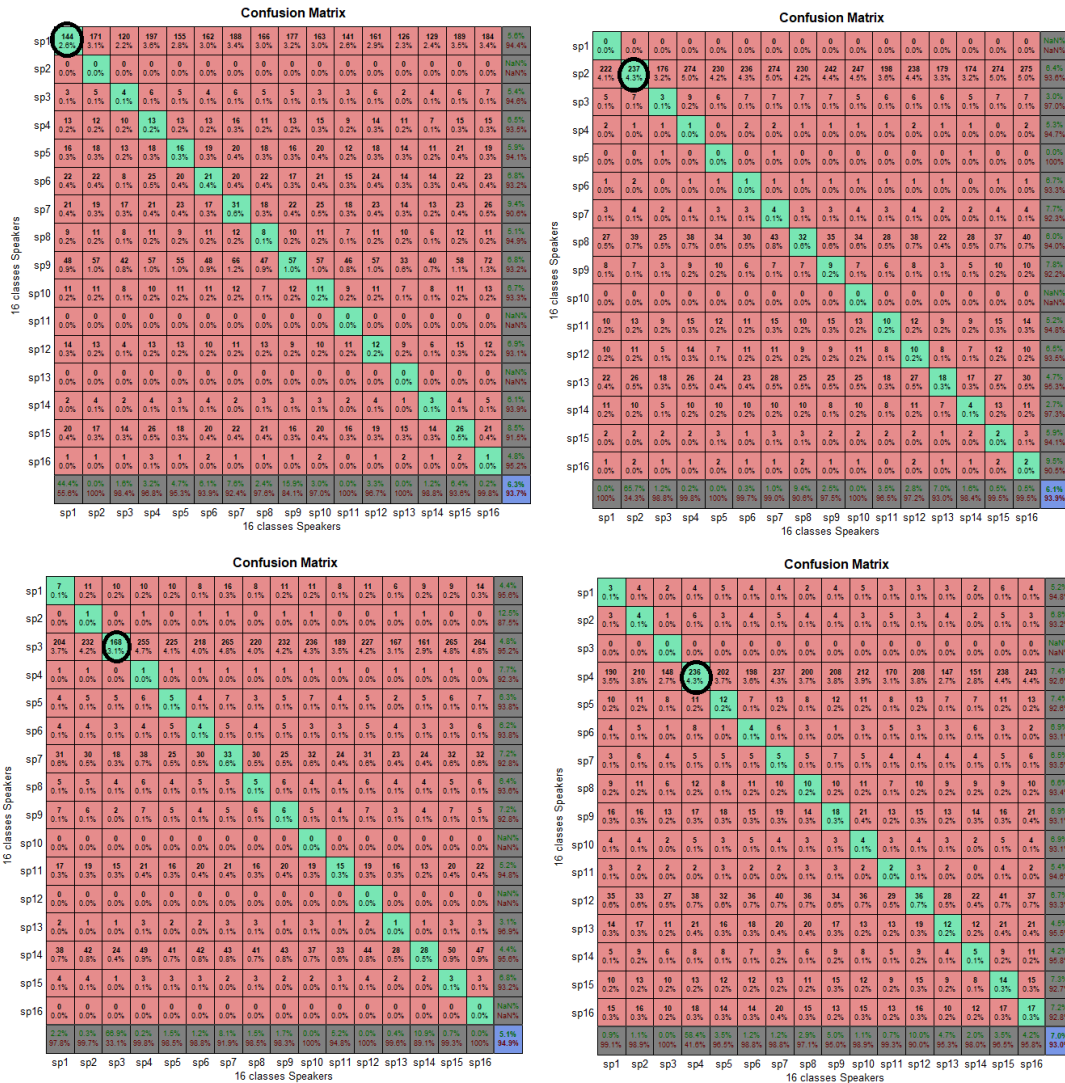


Figure. 9 The diagonal cells score of the confusion matrix while testing the network with sp1, sp2, sp3, and sp4 speakers.

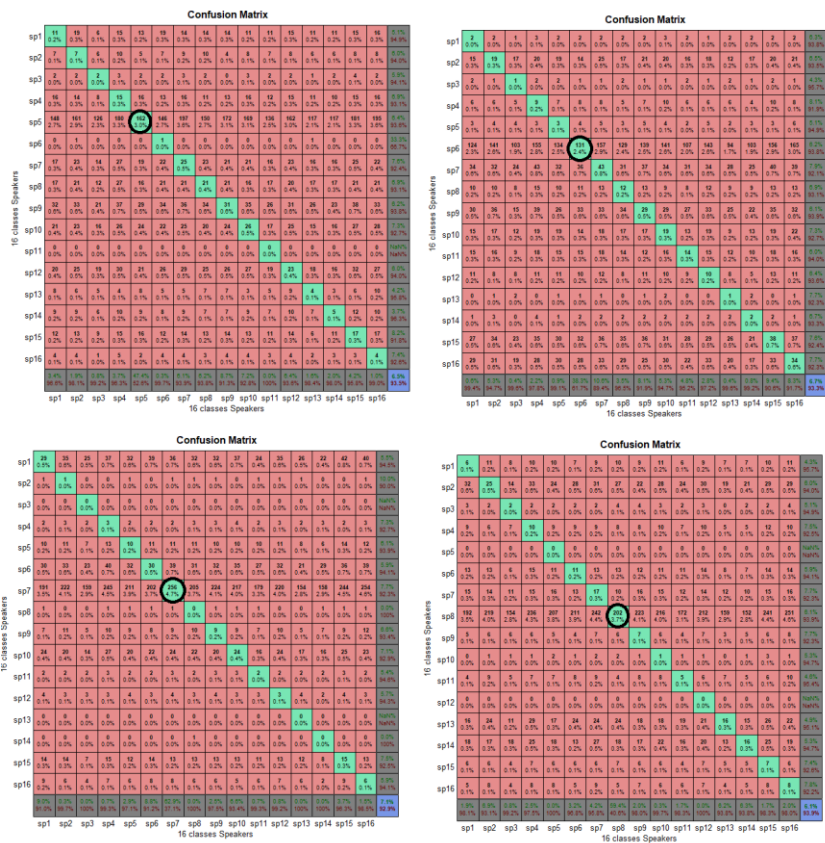


Figure. 10 The diagonal cells score of the confusion matrix when testing the model with sp5, sp6, sp7, and sp8 speakers.

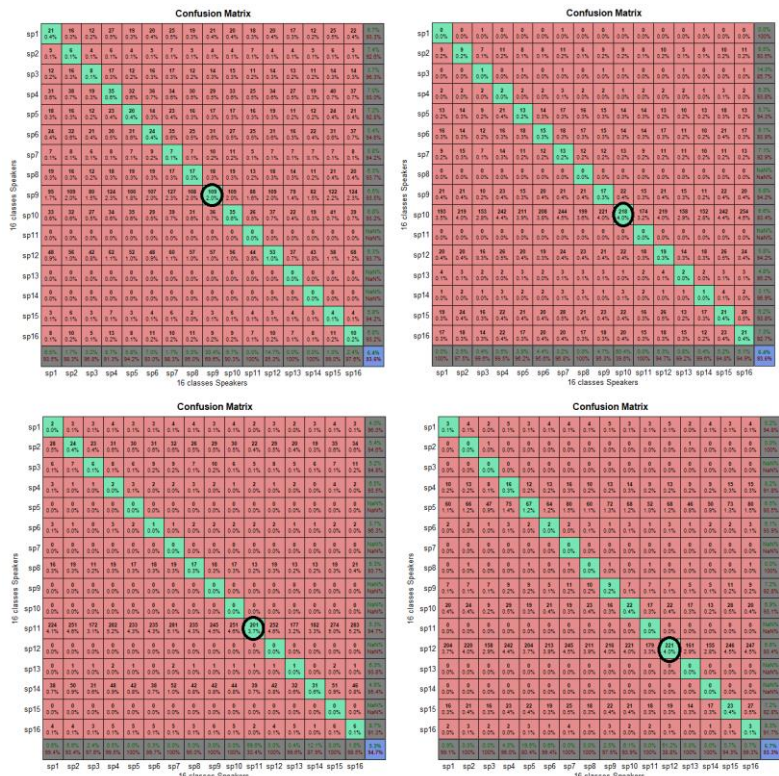


Figure. 11 The diagonal cells score of the confusion matrix when testing the model with sp9, sp10, sp11, and sp12 speakers.

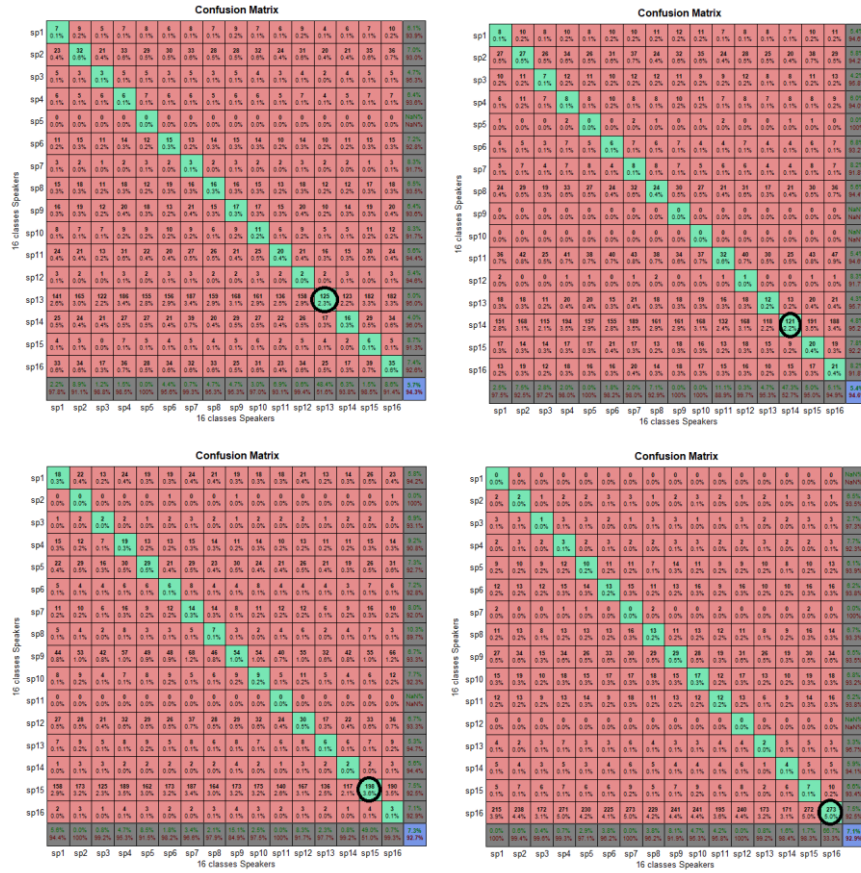


Figure. 12 The diagonal cells score of the confusion matrix when testing the model with sp13, sp14, sp15 and sp16 speakers

5.0 Discussion

The identification accuracy of this model is 100% as all 16 registered speakers are correctly identified while testing the network with registered speaker voice. The comparison of the classification score of each speaker and the testing score of the classifier are also shown in Table 3. The testing score percentage is evaluated wrt the classification score.

Speakers	Classification score	Testing Score	Percentage of testing score
Sp1	296	144	48.64%
Sp2	319	237	74.29%
Sp3	246	168	68.29%
Sp4	381	236	61.94%
Sp5	299	162	54.18%
Sp6	303	131	43.23%
Sp7	379	256	67.54%
Sp8	328	202	61.58%
Sp9	321	118	36.76%
Sp10	352	218	61.93%
Sp11	271	201	74.16%
Sp12	338	221	65.38%
Sp13	239	125	52.30%
Sp14	229	121	52.83%
Sp15	366	198	54.09%
Sp16	397	273	68.76%

Table 3. Testing score rate of the speakers

In a text-dependent speaker identification system, the system requires cooperation from the test speaker for its correct identification. If a test speaker can almost replicate the training and testing voice data, the percentage of the testing score rate is high like for sp2, sp3 and sp16. But if training and testing voice data are different, the testing score rate is low like for sp9 and sp6. The speaker should maintain the same volume, pitch, tone and pace of the voice with clarity for both training and testing to achieve a high testing score rate. From this work, it is also found that an optimum number of hidden neurons of the neural network is required to obtain a high classification accuracy rate.

6.0 CONCLUSION

When an ample amount of training data is given to the neural network, it will learn the nature of its input data correctly for its classification thereby improving the classification accuracy rate. Therefore, out of a total number of input data, an adequate number of data sets is used for training the neural network. For this work, out of the total data set 85 % is used for training and the remaining 15% for validating the network. The speech classifier classification accuracy obtained is 92.1% with 7.9% misclassification. After classification is done, the trained classifier identifies all the registered test speakers correctly while testing with its new speech inputs. Hence from this work, a correct speaker identification rate of 100% is obtained with the proposed new method.

References

- [1] 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).
- [2] Doddington, G. R. "Speaker recognition -Identifying people by their voices" *Proceedings of the IEEE*, 73(11), 1651-1664. Fang, (1985).
- [3] Fang, C.. From dynamic time warping (DTW) to hidden markov model (HMM). *University of Cincinnati*, 3, 19. (2009).
- [4] 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
- [5] 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.
- [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] 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.
- [8] 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.
- [9] 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.
- [10] Singh, P. P., & Rani, P. " An approach to extract feature using mfcc" *IOSR Journal of Engineering*, 4(8), 21-25. 2014.
- [11] Rakesh, D. R., & Jayasimhan, N. K. "Speaker recognition and authentication" *International Journal of Computer Science and Mobile Computing*, 2(5), 402-407. 2013.
- [12] 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.
- [13] 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.
- [14] Sharma, V., & Bansal, P. K. (2013). A Review On Speaker Recognition Approaches And Challenges. *International Journal of Engineering Research and Technology (IJERT)*, 2(5), 1581-1588.
- [15] Sivan, S., & Gopakumar, C. (2014). An MFCC based speaker recognition using ANN with improved recognition rate. *International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)*, 8(4), 365-369.
- [16] E.C. Gordon, Signal and Linear System Analysis. John Wiley & Sons Ltd., New York, USA, 1998. 2
- [17] 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 June 2015 Pages: 1 - 6, IEEE Conference Publications.

- [18] Koustav Chakraborty, Asmita Talele Prof. Savitha Upadhyaya, "Voice Recognition Using MFCC Algorithm" *International Journal of Innovative Research in Advanced Engineering (IJIRAE)* Volume 1 Issue 10 (November 2014), page 158-161.
- [19] 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.
- [20] Shah, F. (2009, December). Automatic emotion recognition from speech using artificial neural networks with gender-dependent databases. In *2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies* (pp. 162-164). IEEE
- [21] Singh, S., & Rajan, E. G. (2011). Vector quantization approach for speaker recognition using MFCC and inverted MFCC. *International journal of computer applications*, 17(1), 1-7.
- [22] 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.
- [23] 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.
- [24] Ayshwarya, N., Logeshwari, G., & MALA, G. A. (2014). Feed forward back propagation neural network for speaker independent speech recognition. *International Journal of Industrial Electronics and Electrical Engineering* ISSN, 2347-6982. 2(8), 36-39.
- [25] Roger Achkar*, Mustafa El-Halabi*, Elie Bassil*, 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.