Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 10, Number 3 (2017) pp. 445-460

http://www.ripublication.com

Designing and Modeling of Speech and Speaker Recognition System to Control the Wheelchair

Gurpreet Kaur[1, 2,*], Mohit Srivastava[3] and Amod Kumar[4]

[1] *Research Scholar, I.K Gujral Punjab Technical University, Kapurthala, India.*
[2] *Assistant Professor, University Institute of Engineering & Technology, Panjab University, Chandigarh-160025, India.*
[3] *Professor, Chandigarh Engineering College, Landran, Mohali- 140307, India.*
[4] *Scientist, Central Scientific Instruments Organisation, Chandigarh-160030, India.*

Abstract

The proposed work presents the speech operated wheelchair for handicapped person which is controlled with the voice command from handicapped person. Speech user interface is imperative for movement of wheelchair based on the speaker voice. The learning curve of the technique is usually very steep in starting. To guarantee good accuracy and less learning time, feature extraction process is very important for the any recognition system. In order to deal with these metrics, we have proposed the automated speaker dependent speech recognition working application for disabled persons in MATLAB using MFCC feature extraction with Artificial Bee Colony (ABC) Algorithm to optimize the extracted features and generating an appropriate feature set. In the first phase, feature extraction using MFCC is executed to find out the feature set from the speech signal and after that feature optimization is used to optimize the MFCC feature using ABC algorithm and in last phase, training is conducted using the Feed Forward Back Propagation Neural Network (FBPNN). In the proposed work, comparative research is proposed with different types of classifiers. For the comparative study, Back Propagation Neural Network (BPNN) is used with MFCC and ABC algorithm. In the end, evaluation and validation of the proposed work model is done by setting real environment in MATLAB 2015a and to check the efficiency of proposed work, parameters like accuracy, precision rate, recall rate, sensitivity and specificity have been calculated.

* Corresponding Author

Keywords: Speech Recognition, MFCC, Feed Forward Back Propagation Neural Network (FBPNN), Back Propagation Neural Network (BPNN), Artificial Bee Colony Algorithm (ABC), Speaker Dependent System

1. INTRODUCTION

The robotic technology wheelchair [1] expands the abilities of conventional controlled gadgets by presenting control and navigational knowledge. These gadgets can facilitate the lives of numerous handicapped individuals, especially those with extreme weaknesses by expanding their scope of portability. For debilitated individuals, human have found a wheel [2] seat which can be moved by utilizing hands for the individuals who don't have legs. In any case, the people who don't have legs and even cannot move their wheel seat by themselves [3] and require some other individual to move their wheel seat. But, once in a while such individual confronts such a large number of issues in the event that they didn't get any individual to move their wheel seat.

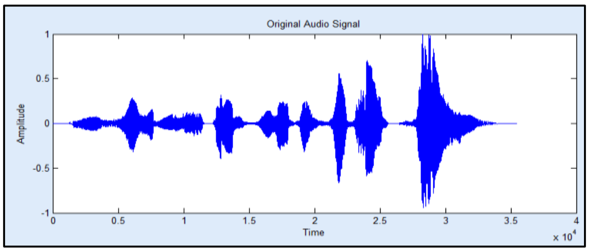

Figure 1: Original Audio Signal of Collected db without noise

Above figure represents the speech signal [4] without any type of distortion in the form of graph and the x coordinate of graph is time and the y coordinate is amplitude of the signal.

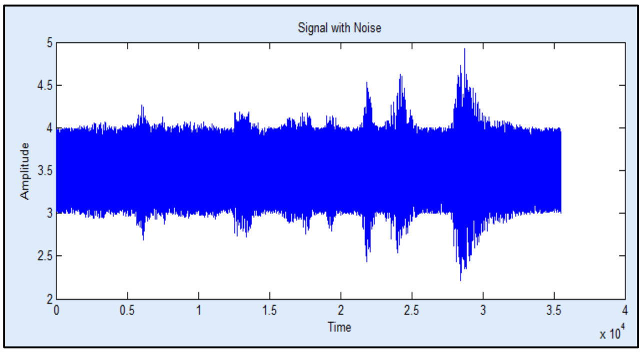

Figure 2: Original Audio Signal of Collected db with noise

Above figure represents the speech signal with distortion in the form of graph and the x-coordinate of graph is time and the y-coordinate is the amplitude of the signal and amplitude of signal shows the noisy signal.

2. PROPOSED TECHNIQUE

There have been different methodologies proposed for parametrically speaking to the discourse signal for the speaker acknowledgment errand, for example, Linear Prediction Coding (LPC) [5], Mel Frequency Cepstral Coefficients (MFCC) [6], and others. MFCC is maybe the best known and most prominent feature extraction system for the speech signals [7]. The primary phase in any speech recognition system is to extract features i.e. classify the components of the speech signal that are excellent for identifying the linguistic content and discarding all the other stuff which carries information like background distortion, noise, emotion etc. MFCCs [8] basically depends on the known variety of the component basic data that transfer capacities with recurrence channels dispersed straightly at low frequencies and logarithmically at high frequencies have been utilized to catch the phonetically essential attributes of discourse vital in the component extraction module.

After the feature extraction using MFCC technique [9], ABC algorithm [10] is used to optimize the extracted feature set using their objective function. In the proposed work, colony of artificial bees consists of three types of bees: (i) employed bees, (ii) onlookers and (iii) scouts. The first half part of the colony consists of the only employed artificial bees and the second half part includes the onlooker's artificial bees. The quantity of employed bees is equal to the number of extracted feature set using the MFCC. The employed bee whose feature set has been exhausted by the bees becomes a scout bee.

In the proposed work, the control parameters of ABC Algorithm [11] are:

Defining the total bees	
Objective function	
Number of employed bees	1^{st} 50% of the total bees
Number of onlookers bees	2^{nd} 50% of the total bees
Number of scouts bee	1

In the feature matching module, a comparative module is presented using the feed forward back propagation neural network [14] and back propagation neural network [18] are used to confirm the command spoken by matching the previously extracted features of the command.

The main challenge for speech recognition is the irregularities in the linguistic rules. They are also known as homophones. Other problem is caused by dependency of phonemes. A feed forward back propagation neural network [15] is a type of Artificial Neural Network architecture where the connections are "fed forward", that do not generate cycles. A feed forward back propagation network [16] is a network that just happened to be trained with a back propagation training algorithm. The feed forward back propagation training algorithm subtracts the training output from the target to obtain the error signal.

The expression "Feed forward" is also used when the user input something at the input layer and it travels from input to hidden and from hidden to output layer. In proposed work, we have used feed forward back propagation neural network [17] to train the optimized feature set using the ABC algorithm [12]. Feed forward back propagation neural network creates a structure with input feature dataset and according to the categories of feature, they separates the feature data into different groups. After the training, we have created 5 different categories of output on the basis of speech signal. In proposed work, we have also created groups on the basis of speaker [8] and store all categories of speech according to the speakers. In other side, we have used only back propagation neural network training algorithm [19] to compare the performance of proposed work. The structure and training scenario of feed forward back propagation neural network is given below.

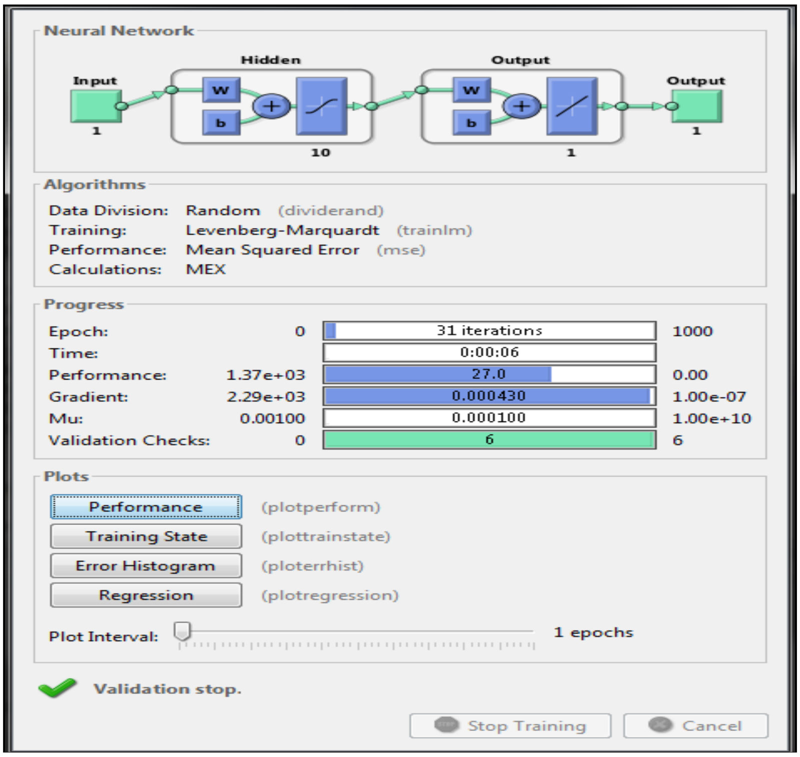

Figure 3: Structure of FBPNN

After the feature optimization using ABC Algorithm [13], we have passed the optimized feature sets to feed forward back propagation neural network [20] for training purpose of proposed module. Artificial Neural Network Toolbox supports a variety of supervised and unsupervised architectures for the training in pattern recognition. Through the toolbox's modular approach of building networks, the user can develop custom network architectures for the specific problem. User can view the network architecture including all inputs, layers, outputs, and interconnections. For the feed forward back propagation neural network training, 'train' function for the training of data according to the initialized function is used.

A Glance of Existing Techniques

The main goal of speech recognition area is to develop techniques and systems for speech input to machine. Below table 1 shows the work done in this area:

Table 1: Related Work

Author	**Year**	**Proposed Work**	**Technique Used**	**Outcome**
Anoop.V, P.V. Rao	2016	Adaptive filtering function generation for various noise levels	Artificial Bee Colony ; Cuckoo Search Algorithm	Cuckoo search has given 45 % more improvement as compare to Artificial bee colony
Sunanda M endiratta, Dr. Neelam Turk, Dr. Dipali Bansal	2016	Presented an effective recognize with speech recognition for producing an interface among	Artificial Bee Colony (ABC); Particle Search Optimization (PSO)	System performance is high and the proposed hybrid algorithm is best for speech recognition
Roberto A. Vazquez, Beatriz A. Garro	2015	Utilization of Artificial Bee colony algorithm as a learning strategy for training spiking neuron for solving pattern recognition problems.	Artificial Bee Colony	ABC is come out to be useful alternative for the adjustment of synaptic weights of spiking neuron model. Spiking neuron model has provided acceptable results in the task of pattern recognition.
Morteza Behnam and Hossein Pourghassem	2015	A new method is introduced for prediction and detection of seizure attack.	Firefly algorithm; Back propagation neural network	Accuracy of 86.8% is obtained with offline mode and 85.7 % by means of predicted signals with 3.12 seconds as prediction time

Wan-Chen Huang	2012	For the reduction of input vector dimension for decrement of memory storage burden, hybridization of HMM and BPNN	Hidden Markov Model and Back propagation neural network	The system identification rate is 93 % and faster response is obtained
Khalid Saeed and Mohammad Kheir Nammous	2017	Dealt with Toeplitz matrices with the minimal eigen values with Neural network types for speech recognition	Radial Neural network and probabilistic neural network	95.82 % accuracy is achieved when used with radial neural network and accuracy of 98.72 % is obtained for probabilistic neural networks
Pedro O. Domingos, Fernando M. Silva, Hor´acio C. Neto	2005	Description of the implementation of the reconfigurable hardware of ANN (Artificial Neural Network) that features online supervised learning	Back Propagation Neural network	Versatile, Fast and efficient architecture for the neural network processing
Omaima Al-Allaf, Abdelfatah A Tamimi	2016	Pattern recognition NN (neural network) for Iris recognition system.	FFBPNN; CFBNN; FitNet and LVQNet	Lowest MSE value, highest PSNR value and high recognition ratio

3. PROPOSED WORK

In this section, the work being proposed is described and below figure 4 shows the process of proposed methodology.

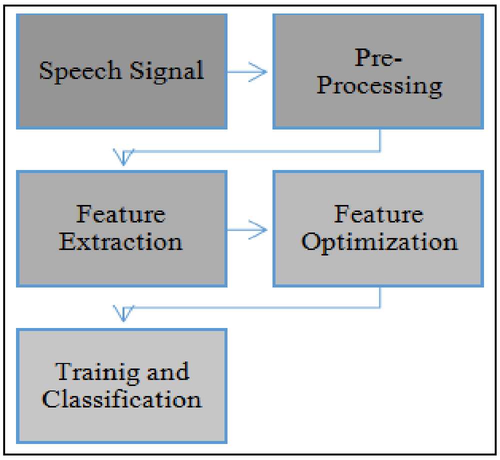

Figure 4: Block Figure of Methodology

The user performs user interpreted commands like Stop, Left, Right, Backward and Forward. A speech signal is recorded for 2 sec interval. The voice signals uses various processing methods like MFCC feature extraction in which feature optimization will be done using artificial bee colony algorithm and classification using feed forward back propagation neural network.

The process of the execution in form of algorithms is shown below with ABC algorithm, MFCC algorithm and feed forward back propagation neural network.

3.1 Proposed optimization with ABC Algorithm

1. Upload dataset for Training

2. Select Case (B, F, L, R and S)

3. Choose Noise Type

    A: Without Noise

    B: White Gaussian Noise (WGN)

    C: Adaptive WGN

If user=1 (Without Noise)

4. Speech_signal=load (Speech Data)

$$\mathbf{Speech\_MFCC_{features\{i\}}} = \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \mathbf{mfcc}(\ \mathbf{Speech\_signal})$$

Initialize ABC Algorithm

Define - Employed bee

- Onlookers bee and

- Scouts bee

Set objective function: $f(fit) = \begin{cases} 1, & fs < ft \\ 0, & fs \geq ft \end{cases}$

$$\mathbf{Opitmized\_MFCC\{i\}} = \sum_{\boldsymbol{i}=1}^{\boldsymbol{r}} \sum_{\mathbf{j}=1}^{\mathbf{c}} \mathbf{ABC}(\ \mathbf{MFCC_{features}}, \mathbf{fs}, \mathbf{ft})$$

where fs is selected value and ft is threshold value

else if user=2 (WGN Noise)

Speech_signal_WGN=load (Speech Data)

$$\mathbf{Speech\_WGN\_MFCC_{features\{i\}}} = \sum_{i=1}^{n} \mathbf{mfcc(\ Speech\_signal\_WGN)}$$

$$\mathbf{Opitmized\_WGN\_MFCC\{i\}} = \sum_{i=1}^{r} \sum_{j=1}^{c} \mathbf{ABC(\ MFCC_{features}, fs, ft)}$$

where fs is selected value and ft is threshold value

else if user=3 (AWGN Noise)

Speech_signal_AWGN=load (Speech Data)

$$\mathbf{Speech\_AWGN\_MFCC_{features\{i\}}} = \sum_{i=1}^{n} \mathbf{mfcc(\ Speech\_signal\_AWGN)}$$

$$\mathbf{Opitmized\_AWGN\_MFCC\{i\}} = \sum_{i=1}^{r} \sum_{j=1}^{c} \mathbf{ABC(\ MFCC_{features}, fs, ft)}$$

where fs is selsted value and ft is threshold value

end (if)

Where B, F, L, R and S are Back, Forward, Left, Right, and Stop respectively are the type of speech by speaker. Speech_signal is the speech data which is uploaded by users

3.2 Feature Extraction Using MFCC Algorithm

Firstly Initialized parameters

Tw = 25(analysis frame duration (ms))

Ts = 10 (analysis frame shift (ms))

Alpha = 0.97(preemphasis coefficient)

R = [300 3700] (frequency range to consider)

M = 20 (number of filter bank channels)

C = 13 (number of cepstral coefficients)

L = 22 (cepstral sine lifter parameter)

hamming=((N)(0.54-0.46*cos(2*pi*[0:N-1].'/(N-1))))

$$\mathbf{MFCC_{features\{i\}}} = \sum_{i=1}^{n} \mathbf{mfcc(\ Signal, fs, Tw, Ts, Alpha, hamming, R, M, C, L)}$$

Where Signal is the speech data which is uploaded by users

3.3 FBPNN Algorithm

Load Opitmized_MFCC_Data

Trainingdata = Opitmized_MFCC_Data

Initialize FBPNN

Generate group of data = group

Set iteration = 100

for i = 1: iteration

Weight = Opitmized_MFCC_Data(i)

Hidden_Layer = [25, 25, 25] (tansig)

Net_algo = trainrp

Generat Net structure of FBPNN (net)

Net = train (net, Trainingdata, group)

end (for)

Save Net as a Training Data and simulate with test data and find appropriate results for proposed work with feed forward back propagation neural network.

3.4 BPNN Algorithm

Load Opitmized_MFCC_Data

Trainingdata = Opitmized_MFCC_Data

Initialize BPNN

Generate group of data = group

Set iteration = 100

for i = 1: iteration

Weight = Opitmized_MFCC_Data(i)

Hidden_Layer = [25, 25, 25] (tansig)

Net_algo = trainrp

Generat Net structure of BPNN (net)

Net = train (net, Trainingdata, group)

end (for)

Save Net as a Training Data with back propagation neural network and simulate with test data and find appropriate results for proposed work.

4. SIMULATION RESULTS

In this section, the simulation results and analysis of proposed work is described. In results, the power of utilization of MFCC, ABC Algorithm and feed forward back propagation neural network for speech oriented wheelchair application is demonstrated. It is much more difficult to recognize speech in presence of noise. Proposed work is tested on various types of noises like WGN, AWGN etc. Due to noise, speech recognition becomes difficult, so, we are using ABC Algorithm for feature optimization to optimize the MFCC feature sets for the training. There are two sections in the proposed work, first is training section and second one is the testing section. The steps involved in the training and testing are same like feature extraction from the speech signal using the MFCC feature extraction, feature optimization using the ABC optimization algorithm and for the classification two types of different classifier are used namely, feed forward back propagation neural network and back propagation neural network.

In the simulation, speech signal are uploaded and then the extraction of the MFCC feature using the given formula takes place.

$$\text{Speech\_MFCC}_{\text{features}\{i\}} = \sum_{i=1}^{n} \text{mfcc}(\text{ Speech\_signal})$$

Where Speech_signal is the uploaded signal and it may be normal signal and noisy signal. Speech recognition with noisy signal is a big task so in proposed work, we have divided the simulation in three different phases.

1. Signal without noise
2. Signal with WGN noise
3. Signal with AWGN noise

For the all types of signal, we have extracted the feature and then optimized them to enhance the feature set using the ABC optimization algorithm. After the optimization, we trained the proposed work with different types of artificial intelligence algorithms to check the accuracy of proposed work. There are two types of artificial intelligence algorithms used in proposed work, that are:

1. Feed forward back propagation neural network
2. Back propagation neural network

In the training phase, we have used the set of 25 hidden layers with tan sigmoid transfer function to train the input feature data. After the training of the proposed work, we have tested the simulation with a test speech signal and repeat of process for training in testing phase. By using the performance metrics like precision rate, recall rate, accuracy, sensitivity and specificity, we can compare the proposed work with different artificial intelligence algorithms. The performance metrics are defined as:

$$Precision\ Rate = \frac{Tp}{Tp + Fp}$$

$$Recall\ Rate = \frac{Tp}{Tp + Fn}$$

$$Sensitivity\ Rate = \frac{Tp}{P}$$

$$Specificity\ Rate = \frac{Tp}{N}$$

$$Accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn}$$

Where, Tp is true positive, Fp is false positive, Tn is true negative, Fn is false negative P is Fn+Tp and N is Fp+Tn. In this section, we have analyzed the obtained results of proposed work for Speaker and speech recognition system. The experimental results have confirmed our expectations by giving good values in terms of measuring metrics like precision rate, recall rate, accuracy, sensitivity and specificity. At the last, result and analysis section shows the comparison of performance parametersThe reliability of classification depends on feature extraction and training process. Therefore, feature extraction and training process should be better. To find out the better feature set, we have used ABC optimization algorithm.

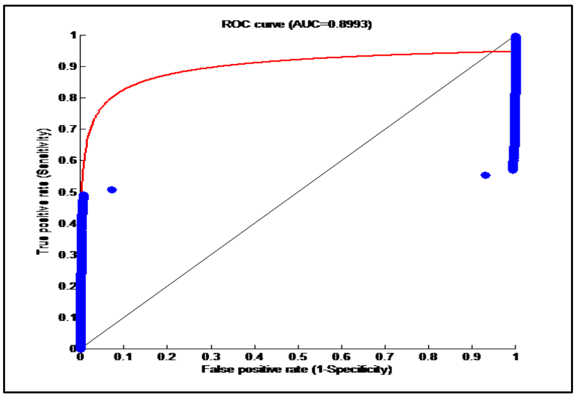

Figure 5: ROC Curve for Proposed Work

ROC (Receiver Operating Characteristics) is a graphical method for comparing two empirical distributions. In this research work, true positive and false negative parameters have been taken. Graphics are always useful as a method of summarizing data and often they can reveal aspects that are not obvious by other means, and of course, this is ROC's greatest strength.

Table 2: Proposed Metrics Result Analysis

Parameters	FBPNN	BPNN
True positive	0.944	0.928
False positive	0.443	0.452
True negative	0.444	0.463
False negative	0.504	0.452
Precision rate	0.68	0.654
Recall rate	0.87	0.784
Accuracy	96.23	91.56
Sensitivity	0.652	0.653
Specificity	0.500	0.503

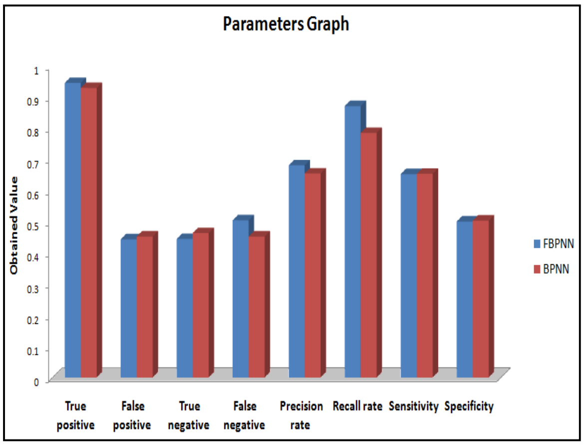

Figure 6: Parameters comparison of Proposed Work

Figure 5 and table 2 represents the calculated parameters of proposed work using MFCC feature extraction with feed forward back propagation neural network and with back propagation neural network. Blue color bar graph represents parameters of proposed work using MFCC feature extraction with feed forward back propagation neural network and red color bar graph represents parameters of proposed work using MFCC feature extraction with back propagation neural network. In the figure, all calculated parameters are shown like precision rate, recall rate, sensitivity and specificity is better with feature extraction using MFCC and feed forward back propagation neural network technique for the classification of speech and speaker.

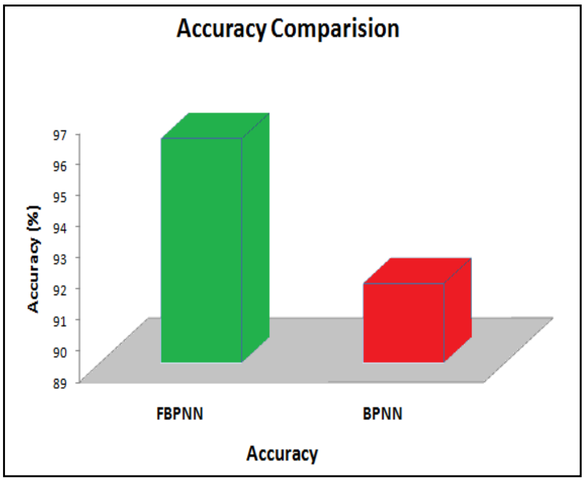

Figure 7: Accuracy comparison of Proposed Work

Above figure represents the accuracy comparison of proposed work with feature extraction using MFCC with feed forward back propagation neural network technique and with back propagation neural network technique and we achieve approximately 97% accuracy for speech and speaker recognition system using feed forward back propagation neural network.

Table 3: Comparison of proposed work based on iterations

No. of Iterations	MFCC with FBPNN	MFCC with BPNN
1	96.18	91.67
2	97.83	89.89
3	96.11	90.41
4	96.87	91.17
5	95.97	91.38

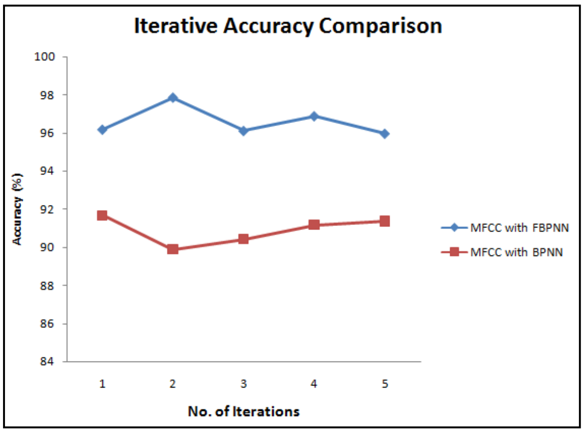

Figure 8: Iterative Accuracy comparison of Proposed Work

Figure 8 and Table 3 shows the comparison for proposed between different types of classifier based on the iterations. Above figure is obtained on the basis of several experiments with speech from different speakers. In the graph, blue line shows the proposed technique based on feature extraction using MFCC with feed forward back propagation neural network and red line shows the proposed technique based on feature extraction using MFCC with back propagation neural network. The average accuracy for proposed work with FBPNN is 96.59% and average accuracy for proposed work with BPNN is 90.91%. So, it is concluded that the proposed system has more accuracy and is reliable.

5. CONCLUSION

In proposed work, we have presented that speech as well as speaker recognition system with MFCC feature extraction technique is helpful in achieving more accuracy of the proposed recognition system. In the proposed work, a comparative research is presented using MFCC feature extraction with feed forward back propagation neural network and with back propagation neural network. To be specific, we have found that optimization and feature extraction are very important as well as difficult steps in any pattern recognition system. In proposed work, we have extracted more useful feature set from speech signal using MFCC technique, feature optimization using ABC optimization algorithm and for the training and classification of data, feed forward back propagation neural network is used, feed forward neural network is used for suitable training in case of optimized signal as well as noisy signal. The experimental results analyzed that proposed method using MFCC with FBPNN provides good results having values true positive 0.944, false positive 0.443, True negative 0.444, false negative 0.504, Precision rate 0.68, and Recall rate 0.87, Accuracy 96.59%, Sensitivity 0.652 and Specificity 0.500.

REFERENCES

[1] R. Chauhan, Y. Jain, H. Agarwal and A. Patil, "Study of implementation of Voice Controlled Wheelchair," *3rd International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, pp. 1-4, 2016.

[2] R. C. Simpson and S. P. Levine, "Voice control of a powered wheelchair," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 10, no. 2, pp. 122-125, 2002.

[3] C. Aruna, A. D. Parameswari, M. Malini and G. Gopu, "Voice recognition and touch screen control based wheel chair for paraplegic persons," *International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)*, Coimbatore, pp. 1-5, 2014.

[4] J. H. L. Hansen and T. Hasan, "Speaker Recognition by Machines and Humans: A tutorial review," in *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 74-99, 2015.

[5] H. J. Hsieh, J. H. Jheng, J. s. Lin and J. w. Hung, "Linear prediction filtering on cepstral time series for noise-robust speech recognition," *IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, Nantou, pp. 1-2, 2016.

[6] I. Patel and Y. S. Rao, "Speech Recognition Using Hidden Markov Model with MFCC-Subband Technique", *International Conference on Recent Trends in Information, Telecommunication and Computing*, Kochi, Kerala, pp. 168-172, 2010.

[7] G. Jhawar, P. Nagraj and P. Mahalakshmi, "Speech disorder recognition using MFCC", *International Conference on Communication and Signal Processing (ICCSP)*, Melmaruvathur, pp. 0246-0250, 2016.

[8] U. Sharma, "Identification of emotion from speech signal", *3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi pp. 2805-2807, 2016.

[9] Vani H. Y and M. A. Anusuya, "Hilbert Huang transform based speech recognition", *Second International aConference on Cognitive Computing and Information Processing (CCIP)*, Mysore, pp. 1-6, 2016.

[10] S. Mendiratta, N. Turk and D. Bansal, "Automatic speech recognition using optimal selection of features based on hybrid ABC-PSO", *International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, pp. 1-7, 2016.

[11] V. Anoop and P. V. Rao, "Adaptive noise cancellation in speech signals using optimization algorithms", *International Conference on Communication and Signal Processing (ICCSP)*, Melmaruvathur, pp. 0027-0031, 2016.

[12] Vazquez, Roberto A., and Beatriz A. Garro, "Training spiking neural models using artificial bee colony", *Computational intelligence and neuroscience,* 2015.

[13] Devi, J. Sirisha, and Srinivas Yarramalle, "Multi objective optimization problem resolution based on hybrid ant-bee colony for text independent speaker verification", *International Journal of Modern Education and Computer Science,* 2015.

[14] Saeed, Khalid, and Mohammad Nammous, "Heuristic method of Arabic speech recognition", *Proc. IEEE 7th Int. Conf. DSPA,* 2005.

[15] Khwaja, Mohammed Kamal, et al., "Robust phoneme classification for automatic speech recognition using hybrid features and an amalgamated learning model", *International Journal of Speech Technology,* pp. 895-905, 2016.

[16] Domingos, Pedro O., Fernando M. Silva, and Horacio C. Neto., "An efficient and scalable architecture for neural networks with backpropagation learning",

Field Programmable Logic and Applications, International Conference on IEEE, 2005.

[17] AL-Allaf, Omaima N. Ahmad, Shahlla A. AbdAlKader, and Abdelfatah Aref Tamimi, "Pattern recognition neural network for improving the performance of Iris recognition system", *Journal of Scientific and Engineering Research,* 2013.

[18] W. C. Huang, "A Method to Combine HMM and BPNN on Speech Recognition", *International Conference on Machine Learning and Cybernetics*, pp. 1899-1902, 2007.

[19] Wan-Chen Huang, "A study of appling BPNN to robot speech interface", *International Conference on Machine Learning and Cybernetics*, Xian, pp. 1681-1685, 2012.

[20] M. Behnam and H. Pourghassem, "Power complexity feature-based seizure prediction using DNN and firefly-BPNN optimization algorithm", *22nd Iranian Conference on Biomedical Engineering (ICBME)*, Tehran, pp. 10-15, 2015.