

INCREASING THE PERFORMANCE OF SPEECH RECOGNITION SYSTEM BY USING DIFFERENT OPTIMIZATION TECHNIQUES TO REDESIGN ARTIFICIAL NEURAL NETWORK

^[1]SHILPI SHUKLA, ^[2]MADHU JAIN, ^[3]R.K. DUBEY

^[1]Department of Electronics and Communication Engineering, Mahatma Gandhi Mission's College of Engineering & Technology, Noida, India.

^[2]Department of Electronics and Communication Engineering, Jaypee Institute of Information Technology, Noida, India

^[3]Department of Electrical Engineering, Central University of Haryana, Mahendergarh, India

^[1]shilpi.avighna@gmail.com, ^[2]ermadhu2003@gmail.com, ^[3]rajeshk_dubey@yahoo.com

ABSTRACT

Speech recognition has a high complexity and a broad range of applications, since it has to predict the word under many types of distortions. This paper aims to compare the performance of different optimization techniques like genetic algorithm, particle swarm optimization and artificial bee colony for optimizing the different hidden layers and neurons of the hidden layers of artificial neural network, for maximum speech recognition accuracy. The features of the input speech signals are extracted using amplitude modulation spectrogram. The outcome demonstrates that the accuracy of artificial bee colony is 95.3% and it performs better when compared with the other optimization techniques.

Keyword: *Amplitude Modulation Spectrogram, Artificial Neural Network, Artificial Bee Colony, Speech Signal, Particle Swarm Optimization*

1. INTRODUCTION

Speech processing has become very popular to interface with electronic gadgets [1-2]. In speech signal processing, the input speech signal is converted into its specific form called features that can be used by learning algorithms for the recognition of the speech signal. Speech quality is a major contributor to the end user's perception of quality of service. As networks become more heterogeneous, complex and new technologies interoperate with legacy equipment, identifying the root cause of voice quality problems can be a challenging task [3-8]. For telephone service providers the evaluation and assurance of speech quality is a critical parameter.

In this work, the input speech signal is pre-emphasized by pre-emphasis filters then processed through first-order filters for smoothing

the signal [9]. The detection of speech emotion is particularly useful for applications which require usual man-machine trades like web films, and computer tutorial applications [10-11]. The speaker recognizable proof framework can be executed by watching the voiced/unvoiced segments or by analyzing the speech energy appropriation. Such frameworks can be partitioned into two fundamental advances: feature extraction and speaker classification [12]. A model adaptation technique leaves the perceptions unchanged and rather refreshes the model parameters of the recognizer to be more illustrative of the speech [13-21]. The amplitude modulation spectrogram (AMS) design has been considered as a novel approach for classification. The AMS patterns can provide a dependable separation between voiced speech and noise as well as

between unvoiced speech and noise [22]. A mainstream approach in speech innovation is artificial intelligence (AI) [23]. The classification of speech signal is imperative phenomenon in speech recognition process. In [24] Artificial neural network (ANN) system is utilized for classification. The ANN has been characterized from multiple points of view by a few researchers. The neurons of ANN are trained using input–output data sets [25], where, the isolated active words of speech are converted into text. For this, various words spoken by different persons are considered as the inputs. The features of input speech signals are extracted using AMS. The extracted input features are fed as the input to the proposed ANN for training and testing. Various optimization techniques have used to increase the efficiency of designs by choosing best solution among various viable solutions in different domains [26-41]. In [34] hybrid algorithms are used in breaking the hard electronic cipher text for linear and non-linear system. For optimizing hidden layer and neurons, various optimization techniques are utilized such as genetic algorithm (GA) [35], particle swarm optimization (PSO) [36-40] and artificial bee colony (ABC) [41].

This work demonstrates the use of ABC optimization algorithm for the purpose of redesigning the ANN structure and using it for increasing the speech recognition accuracy. Here, the ANN structure is redesigned by using existing Levenberg-Marquardt (LM) training algorithm [42] to retrieve optimal prediction rate accuracy. In the last, the result of ABC optimization method is also compared with other optimization algorithms like GA and PSO.

The organization of this paper is as follows. The proposed methodology related to speech recognition is given in Section 2. In Section 3 brief description of ANN is given. Section 4 deals with the brief description of optimization algorithms (ABC, PSO and GA) used in this paper. Section 5 describes the results, while, Section 6 concludes the paper.

2. PROPOSED METHODOLOGY

In this work the input isolated speech signal is recognized using ANN. Different words spoken by different persons (in ambient conditions) represent the normal speech signal and they are taken as the input speech signal. The features of input speech signals are then extracted by using AMS and these features are used as input

to ANN for training and testing. ANN is utilized here for predicting the isolated word. For training and testing LM algorithm is used. The optimization techniques GA, PSO and ABC are used to optimize the neurons and the number of hidden layers. The complete block diagram of proposed work is shown in figure 1. This work is implemented in the working platform matrix laboratory (MATLAB 2015a).

The distinctive words are recorded by the different people (70 males, 50 females and 25 children) in ambient conditions. These words are used as input speech signal. In the current work AMS technique is used for extracting the features of the input speech signal. The AMS technique uses modulation spectrogram instead of spectral envelope as the information in the envelope is robust against noise and room reverberations. The three dimensional coefficients of AMS i.e. time, acoustic and modulation frequencies are obtained by firstly taking the short time fourier transform (STFT) of the input speech signal and then the envelope of the spectrogram is taken by squaring the magnitude of the complex values. The Bark scale decomposition is used for decaying the frequency scale into a set of critical bands and then second STFT is used for the analysis of the long-term spectral envelope of each sub-band.

3. ARTIFICIAL NEURAL NETWORKS

The artificial neural network includes three layers, for example, the input layer, hidden layer and the output layer in figure 2. The neurons of the various layers are connected to each other by synaptic weights. The synaptic weights are continuously updated during the learning or training process until an optimum solution is obtained.

The fundamental function of hidden neurons is given in equation (1).

$$Q_f = \sum_{j=1}^N P_i \times w_{ij} \quad (1)$$

where Q_f is a basics function, P_i is the input, w_{ij} is an input layer weight and i is a number of input. The activation function of hidden neurons is evaluated with equation (2).

$$\tan sig(Q_f) = \frac{2}{(1 + \exp(-2 * Q_f)) - 1} \quad (2)$$

The LM algorithm is utilized to train and predict the improved artificial network structure design. It is a very popular optimization algorithm which uses both the gradient descent method and Gauss –Newton method for finding the optimum value.

The gradient vector is represented in equation (3).

$$p = \begin{Bmatrix} \frac{\partial S}{\partial w_1} \\ \frac{\partial S}{\partial w_2} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial S}{\partial w_n} \end{Bmatrix} \quad (3)$$

Where S is the error of the network for that model and w suggests the weights. The Jacobian matrix obtained is given in equation (4).

$$JC = \begin{Bmatrix} \frac{\partial S_1}{\partial w_1} & \frac{\partial S_1}{\partial w_2} & \dots & \frac{\partial S_1}{\partial w_n} \\ \frac{\partial S_2}{\partial w_1} & \frac{\partial S_2}{\partial w_2} & \dots & \frac{\partial S_2}{\partial w_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial S_m}{\partial w_1} & \frac{\partial S_m}{\partial w_2} & \dots & \frac{\partial S_m}{\partial w_n} \end{Bmatrix} \quad (4)$$

In this work for ANN structural design, different optimization techniques: GA, PSO and ABC are utilized.

4 OPTIMIZATON ALGORITHMS

4.1 Artificial Bee Colony Optimization

This optimization algorithm was proposed by Karaboga [43] and it is based on the intelligent foraging behaviours of swarms of honey bees. It consists of employed bees and unemployed bees. An unemployed bee comprises of onlookers bees (OB) and scout bees (SB). Based on each solution the neural network structure is designed and in this structure, features are given as input and then the output is predicted. Employed bees

(EB) are used to find the new solution in defined solution i , by the following equation (5)

$$t_{ij} = u_{ij} + \Phi_{ij}(u_{ij} - u_{kj}) \quad (5)$$

Where, j is chosen randomly between $[1,C]$ where C is the no. of dimensions, Φ_{ij} is distributed randomly between $[-1, 1]$ and ij demonstrates the record position and $k \in \{1, 2, \dots, \text{size}(i)\}$ is also decided randomly. The critical limitation in this procedure is that $k \neq i$. The probability (p_i) is calculated for the u_{ij} according to the equation (6).

$$p_i = \frac{F_i}{\sum_{i=1}^N F_i} \quad (6)$$

Where, F_i is the fitness function. The above equation (6) finds the probability rate of fitness function after the greedy selection process. Based on the probability rate estimation, the solution is arranged and viewed as the initial solution for the onlooker bee section. Figure 3 gives the pseudo code of ABC algorithm.

4.2 Particle Swarm Optimization

In PSO, the single solution in the search space is called a particle. The fitness function given by equation (6) is optimized and the pbest (best solution) and gbest (global best of any particle in the population) values are updated in every iteration.

The velocity and positions of the particle are updated by the following equations (7-8)

$$x[i+1] = x[i] + a1 * r1 * (pbest[i] - p[i]) + a2 * r2 * (gbest[i] - p[i]) \quad (7)$$

$$p[i+1] = p[i] + x[i] \quad (8)$$

where $x[i]$ is the particle velocity
 $p[i]$ is the current particle (solution)
 $r1$ and $r2$ are random numbers between $(0,1)$ and $a1$ & $a2$ are the learning factors.

4.3 Genetic Algorithm Optimization

GA [26, 35 and 36] is a global search examination technique for finding out the true or

approximate solution for an optimization problems.

The steps involved in GA:

- 1) A random initial population is generated.
- 2) GA then creates a new population by using individuals in the current generation to generate next population.
- 3) For generating new population
 - a) Fitness value is computed by scoring each member of the current population. This is called raw fitness score.
 - b) Expected values are then calculated by scaling the raw fitness scores to convert it into more usable range of values.
 - c) Members called parents are selected based on their expectation.
 - d) Elite individuals are selected from the individuals having lower fitness values in the current population
 - e) Next children are produced from parents either by making random changes to a single parent that is mutation or by crossover that is combination of the vector entries of a pair of parents.
 - f) Current population is then replaced with the children from next generation.
- 4) GA then stops after meeting the stopping criteria.

5. RESULTS AND DISCUSSION

In the speech recognition process different speech signals from diverse persons are considered shown in Table 1 and the ANN classifying technique is used to predict the text. For training purpose, LM algorithm is performed and for optimizing hidden layer and neurons three optimization algorithms are applied named as PSO, GA and ABC.

Figure 4 shows network structures based on three algorithms, hidden layers and neurons are optimized for better recognition accuracy. For ABC algorithm 375 input layers, 2 hidden layers and 1 output layer are considered. In hidden layer 1 the neurons are 23 and in hidden layer 2 the neurons are 21.

For GA in hidden layer 1 there are 7 neurons and in hidden layer 2 the neurons are 20. For PSO the input and output layer is general and in hidden layer 1 the neurons are 8 and in hidden layer 2 the neurons are 30.

The comparison of the algorithms is done on the basis of sensitivity, specificity and accuracy given by equation (9-11)

$$Sensitivity = \frac{TP}{TP + FP} \quad (9)$$

$$Specificity = \frac{TN}{TN + FP} \quad (10)$$

$$Accuracy = \frac{(TP + TN)}{TP + TN + FP + FN} \quad (11)$$

Where TP is true positive, TN is true negative, FP is false positive and FN is false negative.

Table 2 gives the comparison based on classification time consumption for these three algorithms. The comparison graph is plotted with performance matrix such as accuracy, sensitivity and specificity in Table 3.

It is concluded from Table 2-3 that the ABC algorithm used here has maximum accuracy, specificity and sensitivity and takes minimum classification time as compared to GA and PSO algorithms.

6. CONCLUSION

In this work the recognition of the input speech signal is done using AMS as the feature extraction technique. The features extracted are then the input to ANN where LM algorithm is used for training. The existing structure of ANN is optimized using ABC, PSO and GA. The outcome clearly demonstrates that the proposed method of speech recognition is superior to different existing methods. This method decreases the complexity of the calculation and offers good recognition result along with reduction in time consumption. The accuracy value of Artificial Bee Colony is 95.3%, while contrasted with other algorithms like PSO and GA. In future, the strategy can be connected to other speech recognition systems for increasing the efficiency.

REFERENCES

- [1] C.Vimala , V.Radha, "A Review on Speech Recognition Challenges and Approaches", in *World of Computer Science and*

- Information Technology Journal (WCSIT)*, vol.2, no.1, pp.1-7, 2012.
- [2] V. Grancharov, D. Y. Zhao, J. Lindblom, W. Bastiaan Kleijn “Low- Complexity, Nonintrusive Speech Quality Assessment”, in *IEEE Transaction on Audio, Speech, and Language Processing*, vol.14,no.6,pp.1948-1956,Nov.2006.
- [3] T. H. Falk, Q. Xu, W. Chan “Non-Intrusive GMM Based Speech Quality Measurement”, in *Proc. of the International Conference on Acoustics, Speech and Signal Processing*, vol. no. 9, pp.125-128,2005.
- [4] J. Holub, M. Street, O. Tomiska “A Novel Non-Intrusive Voice Transmission Quality Measurement Algorithm based on Wavelet Transform,”, *Wireless Telecommunications Symposium, Pomona, CA*, pp.1-4, Apr. 2006.
- [5] H.Hermansky, “Perceptual linear prediction (PLP) analysis of speech”, *Journal of Acoustical Society of America*, vol. 87, no. 4, pp. 1738–1752, Apr. 1990.
- [6] T. Dresler , J. Holub, R. Smid, “Voice Transmission Quality Measurement based on Wavelet Transform”, in *XVII IMEKO World Congress, Dubrovnik, Croatia*, pp. 233, June 2003.
- [7] J. Lewalle “Tutorial on continuous wavelet analysis of experimental data”, *Technical report, Syracuse University*, Apr. 1995.
- [8] M. Gupta, M. Jain, B. Kumar, “Novel Class of Stable Wideband Recursive Digital Integrators and Differentiators,” *IET Signal Processing*, vol.4, iss.5, pp.560–566, 2010.
- [9] K. Kumar, R.K Aggarwal, A. Jain, "An Analysis of Speech Recognition Performance Based upon Network Layers and Transfer Functions”, *International Journal of Computer Science, Engineering and Applications*, vol.1, No.3, pp.11-20, 2011.
- [10] M. O. Elfahal, M. E. Mustafa , R. A. Saeed, “Automatic Spoken Language Recognition for Multilingual Speech Resources”, *Journal of Theoretical and Applied Information Technology*, vol. 96. No. 24, 2018.
- [11] M. Gupta, M. Jain, B. Kumar, “Recursive Wideband Digital Integrator and Differentiator,” *International Journal of Circuit Theory and Applications (IJCTA)*, vol.39, issue.7, pp.775–782, 2011.
- [12] F .Yan, Y .Zhang, J. Yan, “A Sub-Band-Based Feature Reconstruction Approach for Robust Speaker Recognition”, *EURASIP Journal on Audio, Speech, and Music Processing*, pp. 1-13, July 2014.
- [13] T May, S van de Par , A Kohlrausch, “Noise-robust speaker recognition combining missing data techniques and universal background modeling”, *IEEE Trans. Audio, Speech, Lang. Process*, vol 20 no. 1, pp. 108- 121, 2012.
- [14] J. Li, L. Deng, Y. Gong , R. Haeb-Umbach, “An Overview of Noise-Robust Automatic Speech Recognition”, *IEEE/ACM Transactions on Audio, Speech, And Language Processing*, vol. 22, no. 4, pp.745-777, Apr.2014.
- [15] M. Krawczyk, T. Gerkmann, “STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement,” *IEEE Transactions on Audio Speech and Language Processing*, vol. 22, no. 12, pp. 1931–1940, Dec 2014.
- [16] A. Khalil , M. Elnaby , E.M. Saad , Y. Al-nahari, N.Al-Zubi, M. El, “Efficient speaker identification from speech transmitted over Bluetooth networks”, *International Journal of Speech Technology* , June 2014.
- [17] O. Abdel-Hamid, A. Mohamed, H. Jiang, L .Deng, G. Penn , D.Yu, “Convolutional Neural Networks for Speech Recognition”, *IEEE/ACM Transactions on Audio, Speech, And Language Processing*, vol. 22, no. 10, pp.1533-1545, Oct 2014.
- [18] J. González, A. Peinado, N .Ma, A.Gómez , J. Barker, “MMSE-based missing-feature reconstruction with temporal modeling for robust speech recognition”, *IEEE Trans. Audio, Speech, Lang. Process*, vol. 21 no.3, pp.624–635, 2013.
- [19] M. Ayadia, M. Kamel , F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases", *Pattern Recognition*, vol.44, No.3, pp.572–587, 2011.
- [20] K. Daqrouqa, T. A. Tutunjiba, "Speaker identification using vowels features through a combined method of formants, wavelets, and neural network classifiers”, *Applied Soft Computing*, vol.27, pp.231-239, 2015.
- [21] M. Seltzer, D. Yu ,Y .Wang," An Investigation of Deep Neural Networks for Noise Robust Speech Recognition", in *Proc. of IEEE International Conference on*

- Acoustics, Speech and Signal Processing (ICASSP)*, pp.7398-7402, 2013.
- [22] X. Ma , Z.Weidong, “AMS Based Spectrum Subtraction Algorithm with Confidence Interval Test”, in *Proc. of 7th Asian-Pacific Conference on Medical and Biological Engineering*, pp.389-391, 2008.
- [23] Md. Salam, M. Dzulkifli , S. Sheikh, Malay, “Isolated Speech Recognition using Neural Network: A Work in Finding Number of Hidden Nodes and Learning Parameters”, *The International Arab Journal of Information Technology*, vol.8, no.4, pp.364-371, 2011.
- [24] R. Pankaj, K. Sushil , R. Shweta, “Speech Recognition using Neural Network”, *International Journal of Computer Applications*, pp.11-14,2015.
- [25] D. Gülin , H.S Murat, “Speech recognition with artificial neural networks”, *Digital Signal Processing*, vol.20, pp.763-768, 2010.
- [26] M. Jain, M. Gupta, N. Jain, “Linear Phase Second Order Recursive Digital Integrators and Differentiators”, *Radioengineering*, vol. 21, no. 2, pp. 712 – 717, 2012.
- [27] M. Jain, M. Gupta, N. Jain, “Analysis and Design of Digital IIR Integrators and Differentiators using Minimax and Pole, Zero and Constant Optimization Methods”, *ISRN Electronics*, vol. 2013, pp. 1 – 14, 2013.
- [28] M. Gupta, M. Jain , B. Kumar, “Wideband Digital Integrator and Differentiator,” *IETE Journal of Research*, vol. 58, issue 2, pp. 166 - 170, 2012.
- [29] M. Jain, M. Gupta, N. Jain, “The Design of the IIR Differentiator and its Application in Edge Detection”, *Journal of Information Processing Systems*, vol. 10, issue. 2, pp. 223 - 239, 2014
- [30] M. Jain, M. Gupta, N. Jain, “Design of Half Sample Delay Recursive Digital Integrators using Trapezoidal Integration Rule,” *International Journal of Signal & Imaging Systems Engineering*, vol. 9, issue. 2, pp. 126 - 134, 2016
- [31] M. Gupta, M. Jain , B. Kumar, “Wideband digital integrator”, in *Proc. of IEEE International Conference on Multimedia, Signal Processing and Communication technologies (IMPACT)*, pp. 107-109, March 2009.
- [32] M. Jain, M. Gupta , N. Jain, “A new fractional order recursive digital integrator using continued fraction expansion”, in *Proc. of 2010 IEEE India International Conference on Power Electronics (IICPE)*, pp. 1-6, January 2011.
- [33] M. Jain, M. Gupta , B. Kumar, “Design of a novel Digital Integrator”, in the *Proc. of IEEE International Conference on Aerospace Electronics, Communications and Instrumentation (ASECI)*, pp. 191-193, January 2010.
- [34] M. Alabbas, H. Abdulkareem , A. kareem, “Hybrid artificial bee colony algorithm with multi-using of simulated annealing algorithm and its application in attacking of stream cipher systems” *Journal of Theoretical and Applied Information Technology* 97(1) , pp23-33 ,Jan 2019.
- [35] K.S. Tang, K.F. Man, S. Kwong , Q. He, “Genetic algorithms and their applications”, *IEEE Signal Processing Magazine*, vol. 13, issue 6, Nov 1996.
- [36] R. Hassan, B. Cohanin, O. Weck , “ A Comparison of Particle Swarm Optimization and the Genetic Algorithm”, *Structural Dynamics and Materials Conference, Structures, Structural Dynamics, and Materials and Co-located Conferences*, Texas , 2005.
- [37] J. Kennedy, R. C. Eberhart, “Particle swarm optimization”, in *Proc. IEEE international conference on neural networks* vol. IV, pp.1942-1948.IEEE service center, Piscataway, NJ, 1995.
- [38] R. C Eberhart, J. Kennedy, “A new optimizer using particle swarm theory” in *Proc. of the sixth international symposium on micro machine and human science* pp. 39-43. IEEE service center, Piscataway, NJ, Nagoya, Japan, 1995.
- [39] R. C Eberhart, Y. Shi, “Particle swarm optimization: developments, applications and resources”. in *Proc. congress on evolutionary computation 2001 IEEE service center*, Piscataway, NJ. Seoul, Korea, 2001.
- [40] H. A. Habeeb, H. A. Hussein, M. H Abdulameer, “A New Human Face Authentication Technique Based on Median-Oriented Particle Swarm Optimization and Support Vector Machine”, *Journal of Theoretical and Applied Information Technology* vol. 96. No. 01 , 2019
- [41] A. Kumar, D. Kumar ,S. K. Jarial , “A Review on Artificial Bee Colony

- Algorithms and Their Applications to Data Clustering”, *Cybernetics and Information Technologies*, vol 17, no 3,2017
- [42] M.N Nazri , K. Abdullah, M. Z R., “A New Levenberg Marquardt Based Back Propagation Algorithm Trained with Cuckoo Search”, *Procedia Technology*, vol.11, pp.18-23, 2013
- [43] D. Karaboga , “An idea based on honey bee swarm for numerical optimization”, *Technical Report TR06*, Engineering Faculty, Computer Engineering,2006

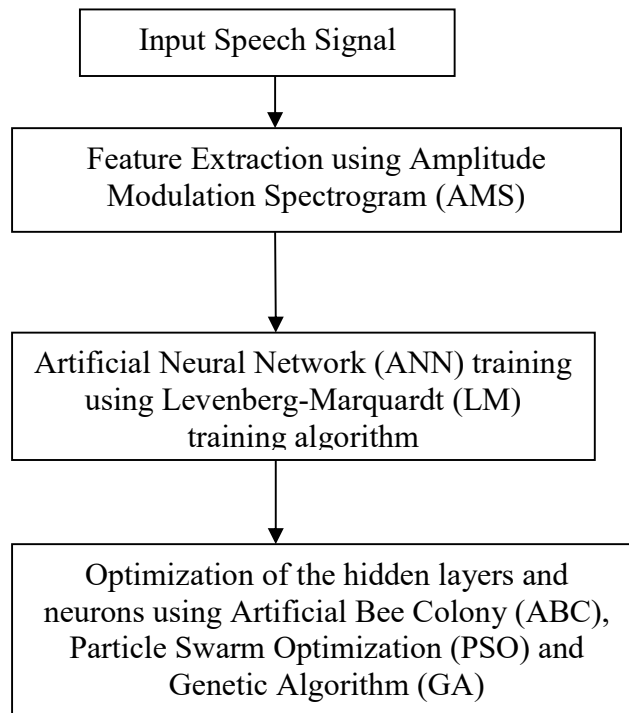


Fig.1 Overall block diagram

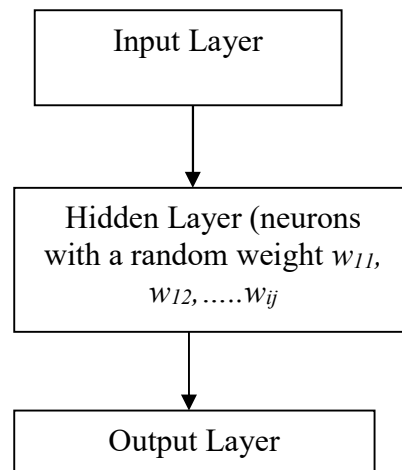


Fig.2 Neural network structure

Notations used:

FS - Food Sources

NS – New Solution

iter – iteration

max_iter – maximum iterations

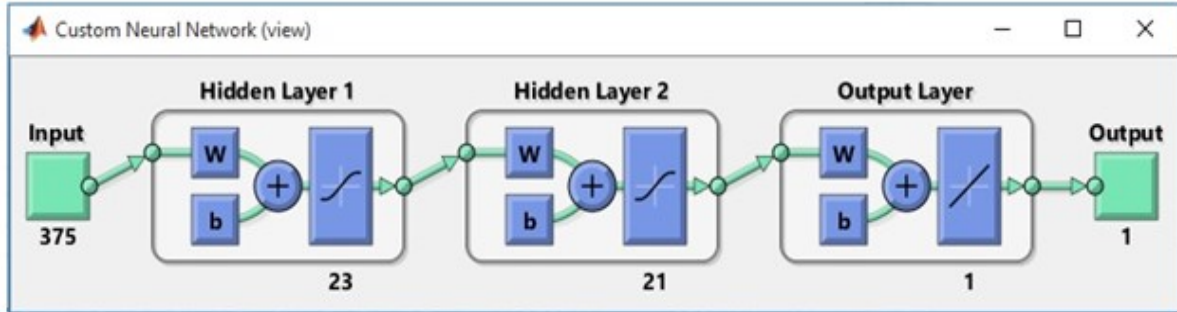
R – Random number

bsn – best solution

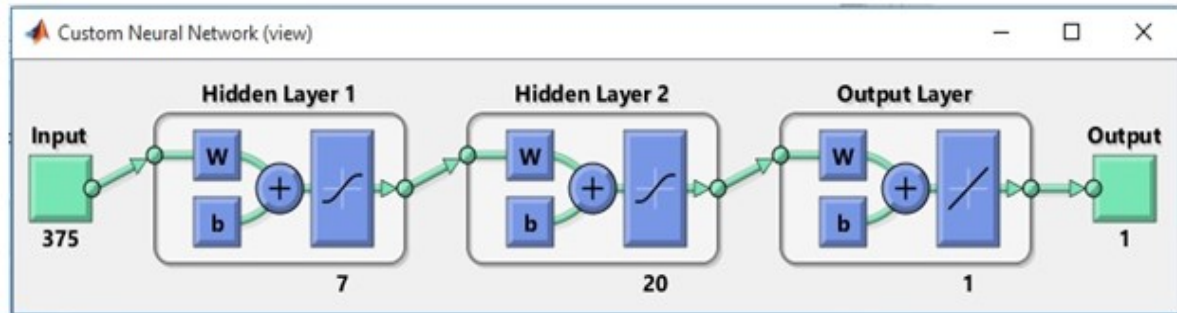
Inputs: Max Cycles, Colony Size, limit

```
1: Initialize FS
2: Evaluate the initialized FS
3: iter =1
4: Set the initial FS to the best  $S_n$  solutions
5: while iter <  $max\_iter$  do
6: Generate NS using EB
7:     Evaluate NS & apply greedy selection process
8:     Calculate probability using eq. (6)
9:     Generate the NS using OB
10:    Evaluate NS & apply greedy selection process
11:    NS generated and abandoned solutions determined randomly using SB
12:    Save best solution
13:   iter=iter+1
14: End while
15: Return  $bsn$ 
Output:  $bsn$ 
```

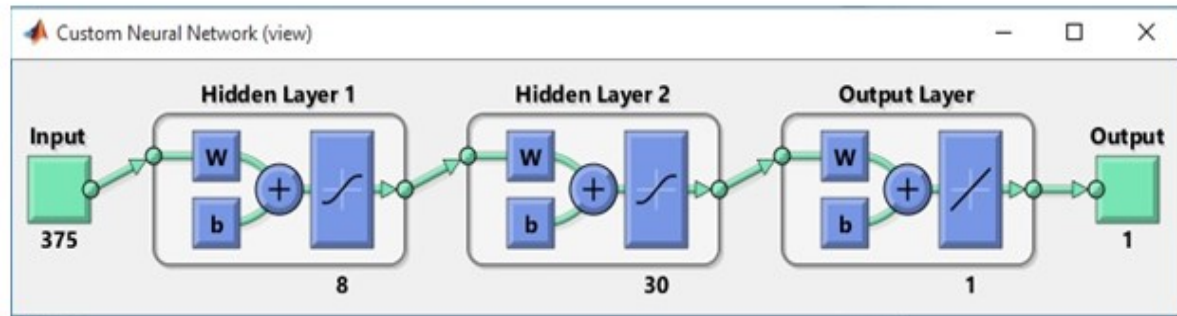
Fig.3 Pseudo code of ABC algorithm



a) ABC Neural Network Structure



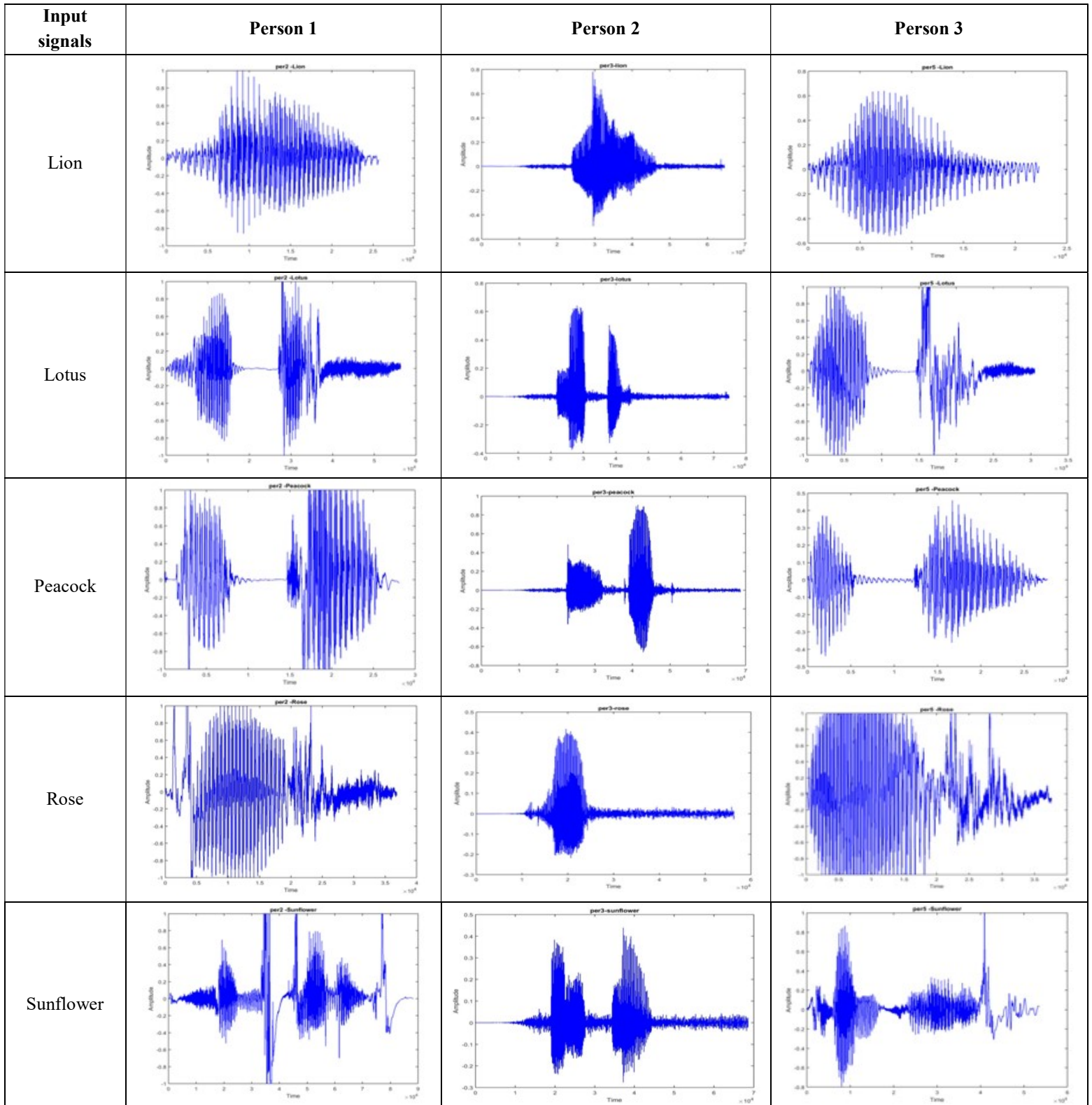
b) GA Neural Network Structure



c) PSO Neural Network Structure

Fig.4 Neural network structure for ABC, GA and PSO algorithms

Table.1 Sample of different person's input signal data



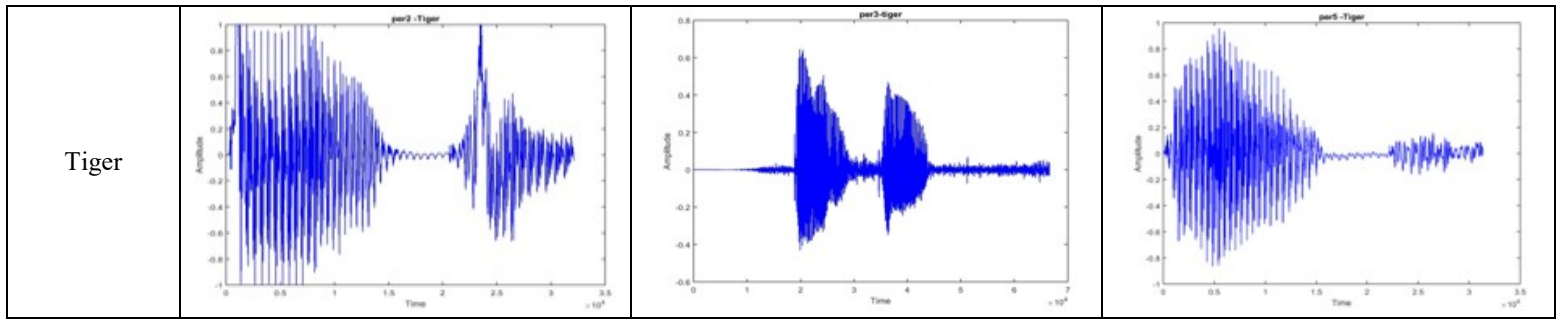


Table 2 Comparison of Classification Time consumption

| Algorithm | Time consumption (in seconds) |
|-----------|-------------------------------|
| ABC | 12.2 seconds |
| PSO | 13.5 seconds |
| GA | 15 seconds |

Table 3. Comparison Results of different algorithms

| Optimizing Algorithm | ANN-ABC % | ANN-PSO % | ANN-GA % |
|----------------------|-----------|-----------|----------|
| Accuracy | 95.3 | 93.3 | 90 |
| Specificity | 97.7 | 96.7 | 94.3 |
| Sensitivity | 88.3 | 87.7 | 87.2 |