

**Review Article**

**RECOGNIZING SPEECH SIGNAL-TO-TEXT USING DEEP NEURAL NETWORK AND OPPOSITION BASED MONARCH BUTTERFLY OPTIMIZATION (DNN-OMBO)**

N.P. Lavanyakumari<sup>1</sup>, M.P.Rao<sup>2</sup>, P.S.V. Subbarao<sup>3</sup>,

<sup>1</sup>Assistant Professor, Department of Physics, Andhra University.

<sup>2</sup>Professor (RTD) Department of Physics, Andhra University. [raomp17@gmail.com](mailto:raomp17@gmail.com)

<sup>3</sup>Associate Professor of Physics, Andhra University. [raopsvs@rediffmail.com](mailto:raopsvs@rediffmail.com)

Received: 13.12.2019

Revised: 16.01.2020

Accepted: 18.02.2020

**Abstract**

The goal of this research work is to identify isolated speech signal using Deep Neural Networks (DNN) technique. For this, different verses enunciated by diverse folks are deliberated as input speech signs. The characteristics of these signs are mined by applying Mel Frequency Cepstral Coefficients (MFCC). The mined features are then inputted to the DNN for training. This paper uses the Opposition based Monarch Butterfly Optimization (OMBO) algorithm for optimizing the weight values of DNN. The performance of the proposed DNN-OMBO algorithm has been compared with the optimization algorithms MBO, PSO, and GA. Experimental results show that DNN-OMBO achieves the highest accuracy with the lowest convergence time when compared to other algorithms.

**Keywords:** Speech Recognition, Feature Extraction, Training, Deep Neural Networks (DNN), Monarch Butterfly Optimization (MBO), Mel Frequency Cepstral Coefficients (MFCC).

© 2019 by Advance Scientific Research. This is an open-access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)  
DOI: <http://dx.doi.org/10.31838/jcr.07.04.145>

**INTRODUCTION**

The goal of Automatic Speech Recognition (ASR) is the text of human language into spoken lyrics. It is a very perplexing mission as human language signals are extremely flexible owing to several utter characteristics, diverse talking graces, in exact ecological sounds, and so on. ASR, furthermore, wants to plot varied-length speech signals into separate length orders of lyrics or phonetic codes. It can be recognized that Hidden Markov Models (HMMs) have been very fruitful in managing varied length orders as well as demonstrating the sequential conduct of speech signals by means of an order of states; each one is related to specific possibility dissemination of explanations [1][2].

Numerous approaches are utilized to arrange AVSR schemes. Several of them are entered on the possible proto types like the ones obtained from the HMM; several others are built through diverse Neural Network enter constructions. There are also heaps of thoughts linking the stated methods to apply more influential AVSR schemes. NNs afford numerous effectual outlines in all phases of the AVSR schemes, together with the characteristic abstraction, information synthesis, and language demonstrating. Diverse styles of NN are used for this principle, particularly after the summary of the new profound learning approaches that have completed several efficacies for additional fast and fruitful employment of the DNNs [8,9].

ASR schemes change the human language or acoustic signs to script. Automatic speech identification is utilized in brainy usages like interactive voice response (IVR) schemes, verbal file recovery, and transcription tools. We want a dependable speech recognition scheme for these usages. Prominent developments have arisen in specific automatic speech identification usages. For instance, discriminative training methods comprising least grouping fault preparation, extreme joint evidence assessment, and least phone fault preparation have been established [3].

Utterer identification is the connected area for the researchers of speech handling. Utterer identification is the responsibility of differentiating distinct from the properties of speech models.

Utterer identification is a vibrant device for countless uses viz. admittance control and customer safety defense. The speech signal transfers vital data viz. message content, language, speaker individuality, speaker reaction, speaker nature, and so on. Latest progresses in the vocal investigation have authorized the picturing of voiced fold dynamics associated with specific vocal ailments. Identification is usually more problematic when words are huge or have numerous similar-sounding lyrics. The aim of this work is to overcome this drawback and identify the utterer. The model factors are frequently educated on MFCC centered front-end parameterization of speech signs. The MFCC is a wide spread feature in Automatic Speech Recognition (ASR) and is following incidental stationery (non-signal reliant) handling procedures [4].

Signs initiating from a similar speech source generally seem contrary because of a variation of audio sound effects. In speech identification, these audio effects bring about in congruities amid proficient speech identification facsimiles and input speech—furthermore, characteristic augmentation targets to unswervingly eliminate the impact of sound from tainted characteristic trajectories. Of late, DNNs have been utilized for toning reverberant speech to its anechoic form [5][14].

**Monarch Butterfly Optimization (OMBO) Algorithm**

So as to mark the relocation conduct of sovereign butter flies addresses numerous developmental issues, the relocation conduct of sovereign butter flies can be flawless into the subsequent rubrics.

1. The entire sovereign butter flies are only positioned in Terrestrial 1 or Terrestrial 2. That is to say, sovereign butter flies in Terrestrial 1 and Terrestrial 2 make up the entire sovereign butterfly populace.
2. Every kid sovereign butterfly discrete is created by a relocation worker from emperor butterfly in Terrestrial 1 or in Terrestrial 2.

3. So as to retain the populace affected, an ancient sovereign butterfly will terminate once a kid is created. In the MBO technique, this can be done by substituting its parent with recently created one if it has improved appropriateness as associated with its parent. Conversely, the newly produced one is accountable to be cast-off if it does not display enhanced suitability regarding its parent. Under this situation, the parent is retained whole and unaffected.
4. The sovereign butterfly individuals with the finest appropriateness transfer habitually to the subsequent cohort, and they cannot be altered by any workers. This can assure that the excellence or the efficiency of the sovereign butterfly populace will never worsen with the increase of peers [6].

#### RELATED WORKS

Kuniaki Noda et al [7] familiarized a connectionist-HMM scheme for sound-robust AVSR. Initially, a profound denoising auto encoder is used for obtaining sound-robust acoustic characteristics. Then, a CNN is used to remove pictorial topographies from raw aperture zone images. By making the preparation data for the CNN as sets of original images and the equivalent phoneme tag outputs, the system is skilled to envisage phoneme tags from the consistent aperture zone input images. Lastly, a multi-stream HMM is used for assimilating the attained acoustic and pictorial HMMs autonomously proficient with the relevant topographies.

Darryl Stewart et al [8] have offered the maximum weighted stream posterior (MWSP) exemplary as a healthy and effectual rivulet assimilation technique for video language identification in situations, where the acoustic or audio-visual rivulets may be exposed to unidentified and time-varying exploitation. A momentous benefit of MWSP is no need for any precise dimensions of the sign in either rivulet to compute suitable rivulet weights in the course of identification, and as such it is modality-liberated. This also means that MWSP accompaniments can be utilised alongside several of the other methods that have been suggested in the collected works for this issue.

Negin Najkar et al. [9] have suggested a lively coding that is substituted by a hunting technique that is centered on the PSO procedure. The chief notion is attentive on engendering a preliminary populace of division trajectories in the result hunt space and augmenting the place of sections by an appraising method. Numerous approaches are familiarized and assessed for the depiction of elements and their consistent effort organizations. Besides, two-division plans are discovered. The first technique is the standard division, which attempts to exploit the possibility function for every challenging audio model distinctly. In the succeeding technology, a worldwide division knotted amid numerous models, and the scheme attempts to enhance the possibility by means of a mutual knotted division.

Yanmin Qian et al. [10] have suggested an unverified learning outline to use these untranscribed data for audio exhibiting. At that juncture, so as well guide the data cohort, ail evidence is familiarized into GAN organizations, and the conditional GAN is used: two diverse circumstances are discovered, comprising the audio state of every speech edge and the unique harmonizing spotless speech of every speech edge. With the amalgamation of precise circumstance evidence into data generation, these provisional GANs can offer factual tags openly, which can be utilized for later audio demonstrating. In the course of the audio model teaching, these descriptive tags are pooled with the lenient tags, which make the exemplary improved.

#### PROPOSED METHODOLOGY

##### Overview

The main objective of this research work is to recognize speech signals using Deep Neural networks (DNN) and the OMBO algorithm. In this process, different words spoken by different persons representing the normal speech signal are deliberated as input. The characteristics of input language are mined by using MFCC. The mined characteristics are considered as input to DNN for training and testing. DNN is utilized for predicting the isolated word speech into text. The optimal weight values for DNN are configured by applying an OMBO optimization algorithm to further enhance the performance. The proposed technique is compared with other optimization-techniques MBO, PSO, and GA.

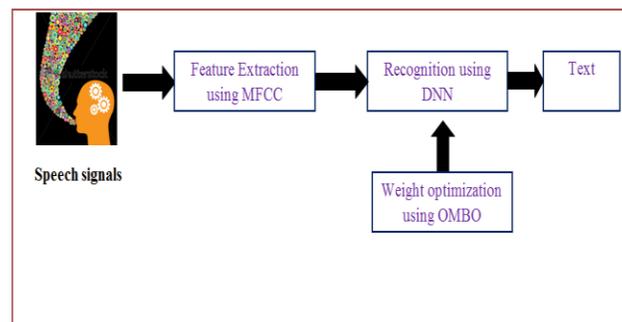


Figure 1: Block diagram of the proposed technique

#### Input Signal

The various people record the dissimilar voice, and all are talked similar words like a lion, lotus, peacock, rose, sunflower, and tiger. It is a powerful speech, which is recorded from the ambient condition, utilized as the input signal, and extracts the features in the below section.

#### Feature Extraction

The feature extraction procedure includes an inspection of the speech sign. The MFCC is utilized for removing the structures of speech sign.

#### Mel Frequency Cepstral Coefficients (MFCC)

In speech recognition related assignments, MFCC is a standout amongst the best feature representations, commonly utilized in automatic speech and speaker recognition and by means of a mesh store investigation, the coefficients are acquired. The general process of the MFCC feature computation is outlined in Figure 2.

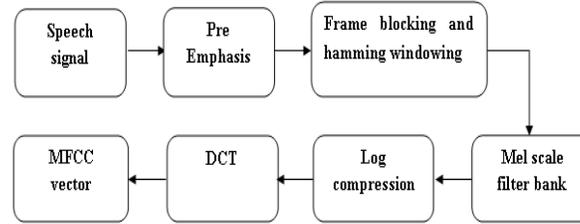


Figure 2: MFCC process

### Pre-Weighting

Pre-weighting is the development to intensify the extent of occurrences as for the extent of diverse incidences.

The SS is primarily pre-weighted by an initial order FIR mesh with pre-weighting constant  $\beta$ . The initial order FIR mesh transmission purpose in the z field is,

$$F(z) = 1 - \beta z^{-1} \quad (1)$$

The pre-weighting s coefficient  $\beta$  lies within the series  $0 \leq \delta \leq 1$ .

Eq. (2) and (3) are utilized for training and testing

$$e(v_i^{r'}) = \rho(v_i^{r'}) - \beta \rho(v_i^{r'} - 1) \quad (2)$$

$$e(v_j^{t'}) = \rho(v_j^{t'}) - \beta \rho(v_j^{t'} - 1) \quad (3)$$

### Frame Blocking and Hamming Windowing

Presently, the successive frames being detached by  $f_A$  samples and the pre-weighted signal is jammed into borders of  $f_S$  models (edge dimension). If the  $k^{th}$  edge of speech is  $x_k(v_i^{r'}), x_k(v_j^{t'})$  and there are  $K$  frames inside the whole speech signal, at that point

$$x_k(v_i^{r'}) - \rho(f_{A'} + v_i^{r'}), \quad 0 \leq v_i^{r'} \leq f_{A'} - 1 \quad (4)$$

In windowing, every one of the exceeding edges is increased with a pretence window so as to retain steadiness of the sign to decrease the sign disruptions towards the starting and finish of the edges.

Hamming window is calculated as,

$$y(n) = x(n) \times w(n) \quad (5)$$

where  $w(n)$  is the function given by

$$w(n) = 0.54 - 0.46 \cos\left(2\pi \frac{n}{N}\right) \quad 0 \leq n \leq N \quad (6)$$

### Mel scale Filter Bank (MSFB)

To change every period area edge of  $f_S$  models into regularity area, the filter bank examination is viably applied. The filters taken in general are commonly known as MSFB. In this manner,

the following equation is utilized to assess the Mel for the pre-specified frequency  $f$  in HZ:

$$F(Mel) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \quad (7)$$

### Logarithmic Compression

Presently, the mesh productivities acquired from mesh panel examination is trodden by the logarithmic operation. The  $f_a^{th}$  mesh logarithmically trodden throughput is conveyed as,

$$X_{f_a^{(ln)}} = \ln(X_{f_a}), \quad 1 \leq f_a \leq f_{A'} \quad (8)$$

### Discrete Cosine Transformation (DCT)

This is the procedure to convert the log Mel range into period area utilizing DCT. At that point, DCT is connected to the filter outcomes of a particular speech frame.

The  $K^{th}$  MFCC coefficient is computed as,

$$y(k) = w(k) \sum_{n=1} x(n) \cos\left(\frac{\pi}{2N} (2n-1)(k-1)\right) \quad k = 1, 2, \dots, N \quad (9)$$

Where

$$w(k) = \begin{cases} 1 & k = 1 \\ \sqrt{\frac{2}{N}} & 2 \leq k \leq N \end{cases} \quad (10)$$

The features extracted from MFCC are non-linear so that those features divided into 10-blocks. The values filled in the blocks; average values in the blocks utilized for further process.

### Deep Learning Neural Network (DNN)

A DNN is a network with a fixed level of intricacy and with diverse layers. DNN uses a complex technical exemplary for managing the data in an erratic mode. DNN with plentiful layers typically combines the characteristic removal and organization procedure into a signal learning body. These kinds of NN have attained achievement in multifaceted areas for the documentation of designs in contemporary ages. In common, DNN has an input layer for the fresh descriptors  $X_l, L$  concealed layers, and an throughput layer for the data extrapolation [15]. A sketch of the suggested DNN is exposed in Figure 3.

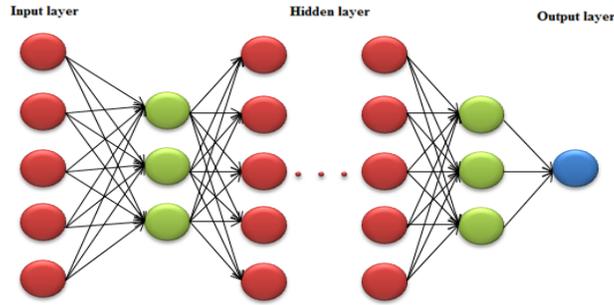


Figure 3: DNN architecture

The pseudo code clarifying the steps of DNN is given underneath:

- load the teaching and analysis group of CKD dataset
- create the classifier using tf.contrib.learn.DNN Classifier Google collection centered on the selected manual arrangement, i.e., amount of concealed layers, initiation function, No. of learning stages, neuron sum for making up the concealed layer;
- Fit the exemplary using the fit function;
- Calculate the exactness in the preparation group using evaluate() function;
- Calculate the extrapolation in the analysis group using the predict() function;
- Evaluate the cataloging outcomes in the analysis group using the misperception matrix;
- Authenticate the organization outcomes on the complete CKD dataset.

For optimizing the weights in DNN, the OMBO algorithm is employed for which the flowchart is shown in below Figure 4.

#### Opposition based Monarch Butterfly Optimization (OMBO)

MBO is a conservative metaheuristic procedure suggested by Wang [6] in 2015. It is enthused by the performance of sovereign butterflies via relocation. In MBO, every discrete sovereign butterfly venerates and approves in two terrains. The dwellings of the sovereign butterflies are efficient in two specific habits: the relocation worker and the butterfly-adjusting worker.

Lin Sun et al. [11] have industrialized a new MBO procedure entered on opposition-based learning (OBL) and random local perturbation (RLP). Yanhong Feng [12] offered opposition-based knowledge and displayed its expediency in precise developmental issues. The for emozstgoal in this plan is to produce an opposition-based result for primarily produced arbitrary results to trap ideal resolution around the crook.

Hui Hu et al. [13] have presented the notion of a self-adaptive plan as an enhanced MBO procedure. In this work, the self-adaptive method enthusiastically alters the butterflies in terrestrial 1 and 2. Consequently, the populace cope in subpopulation 1 and 2 are animatedly altered as the procedure progresses in an undeviating method.

#### Initialization

The solutions are generated based on weight, and the solution length determined as feature extraction. The solution range from -100 to 100, here the solution length determined as 11 (MFCC-10).

The function to ascertain the opposition solution with the initial solution is as follows.

$$O_i = x + y - I_i \quad (11)$$

$I_i \in I_1, I_2, I_3 \dots NP$ , these are preliminary results arbitrarily produced antagonism based result where  $x$  and  $y$  are least and extreme values.

#### Fitness Computation

The following eqn. is considered to evaluate the fitness of the previously generated solutions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

#### Relocation Worker

Let  $G$  denotes the generation of worker. The relocation worker is reliant upon to appraise the sovereign butterfly relocation amid Terrestrial 1 and Terrestrial 2, on which sovereign butterflies frame subpopulations 1 and 2 completely. In the beginning, the number of sovereign butterflies in Domains 1 and 2 is deliberated as  $NP1 = \text{ceil}(p * NP)$  and  $NP2 = NP - NP1$ , autonomously. Where  $NP$  denotes the aggregate number of sovereign butterflies,  $p$  denotes the proportion of sovereign butterflies in Terrestrial 1, and  $\text{ceil}(x)$  rounds  $x$  to the adjacent number more or less equal to  $x$ . In a similar method, relocation worker is organized as

$$x^{t+1}_{i,k} = \begin{cases} x^t r1, k & | r \leq p \\ x^t r2, k & | r > p \end{cases} \quad (13)$$

Where  $x^{t+1}_{i,k}$  signifies the  $k^{\text{th}}$  component of  $x_i$  at  $G=t+1$ . Fundamentally,  $x^{t+1}_{r1,k}$  validates the  $k^{\text{th}}$  component of  $x_{r1}$  at  $G=t$ , and  $x^{t+1}_{r2,k}$  signifies the  $k^{\text{th}}$  component of  $x_{r2}$  at  $G=t$ . The present  $G$  is signified by  $t$ . Sovereign butterflies ( $r1$  and  $r2$ ) are designated arbitrarily from subpopulation 1 and subpopulation 2. The ailment in constant ( $r$ ) is identified as

$$r = \text{rand} * \text{peri} \quad (14)$$

Where  $\text{peri}$  validates the relocation periodedge and is fixed to 1.2 in the chief MBO algorithm and  $\text{rand}$  is an arbitrary constant, acquired from an unvarying dispersal.

#### Butterfly Regulating Worker

This supervisor is utilized to inform the spaces of the sovereign butterflies in subpopulation 2. It is rationalized, as shown below:

$$x^{t+1}_{j,k} = \begin{cases} x^t \text{best}, k & | \text{rand} \leq p \\ x^t r3, k & | \text{rand} > p \end{cases} \quad (15)$$

Where  $x^{t+1}_{j,k}$  indicates the  $k^{\text{th}}$  element of  $x_j$  at  $G=t + 1$ ;  $x^t_{\text{best},k}$  authenticates the  $k^{\text{th}}$  element of  $x_{\text{best}}$  at  $G=t$ , which indicates the premium sector of the sovereign butterflies in fields 1 and 2. By then  $x^t_{r3,k}$  indicates the  $k^{\text{th}}$  element of  $x_{r3}$  at  $G=t$ ; the sovereign butterfly  $r3$  is nominated arbitrarily from subpopulation 2. If

$rand > p$ , there has extra power. The region of the butterfly is furthermore reorganised by means of Levy flight if  $rand > BAR$

$$x^{t+1}_{i,k} = x^t_{j,k} + \alpha \times (dx - 0.5) \quad (16)$$

Where the BAR denotes the butterfly-altering rate. In the instance that BAR is lesser than arbitrary value, the  $k^{th}$  component of  $x_j$  at  $G=t+1$  is rationalized, where  $\alpha$  is the premium aspect, as displayed in eq. (17).

$$\alpha = S_{max} / t^2 \quad (17)$$

Where  $S_{max}$  is the suprememarchpace. In Eq. (18),  $dx$  is the marching paceof butterfly  $j$  that is deliberated by Levy flight.

$$dx = Levy(x^t_j) \quad (18)$$

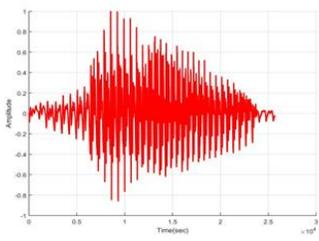
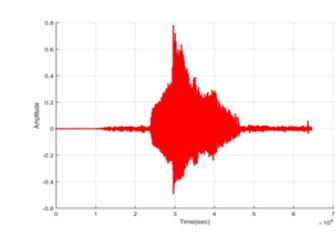
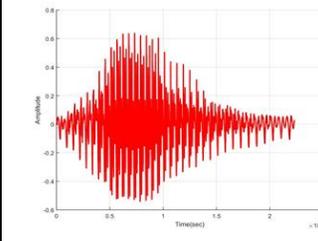
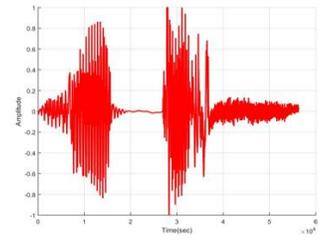
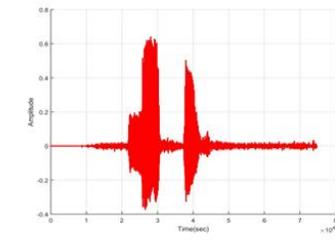
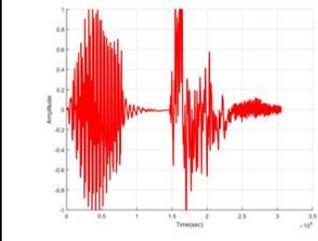
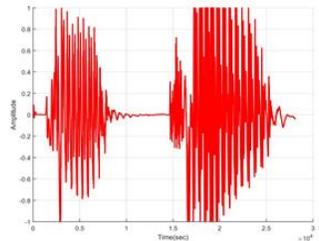
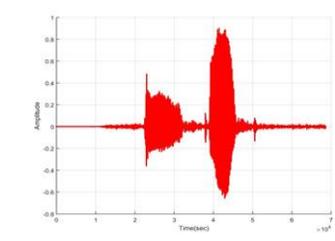
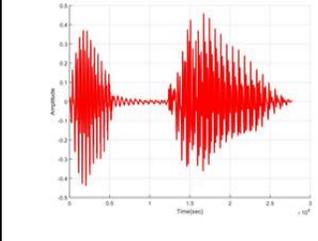
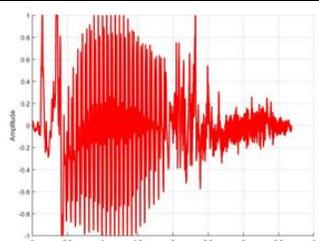
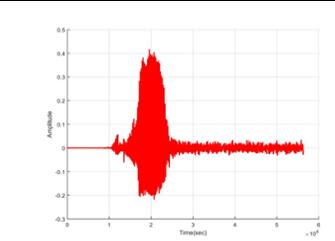
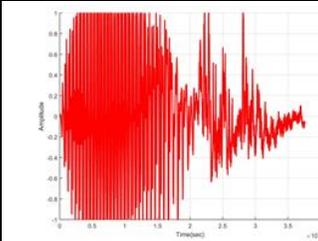
Lastly, the newly produced butterfly with the greatest suitability is substituting with its parent and inspired to the subsequent cohort; similarly, it is removing to endure the populace cope as it is.

## RESULTS AND DISCUSSION

In the speech identification procedure, several speech signs from varied people such as lion, lotus, peacock, rose, sunflower, and tiger are considered, and the DNN categorizing method is applied to expect the script. For weight optimization, four optimization algorithms are applied, namely, OMBO, MBO, PSO, and GA. For these algorithms, the performance comparison graph is plotted with respect to metrics Accuracy, Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), False Positive Rate (FPR), False Negative Rate (FNR) and False Discovery Rate (FDR).

Table 1 portrays diverse persons input signals data, and Table 2 displays features extracted from the signals.

Table 1: Different persons input signal data

Input signals	Person 1	Person 2	Person 3
Lion			
Lotus			
Peacock			
Rose			

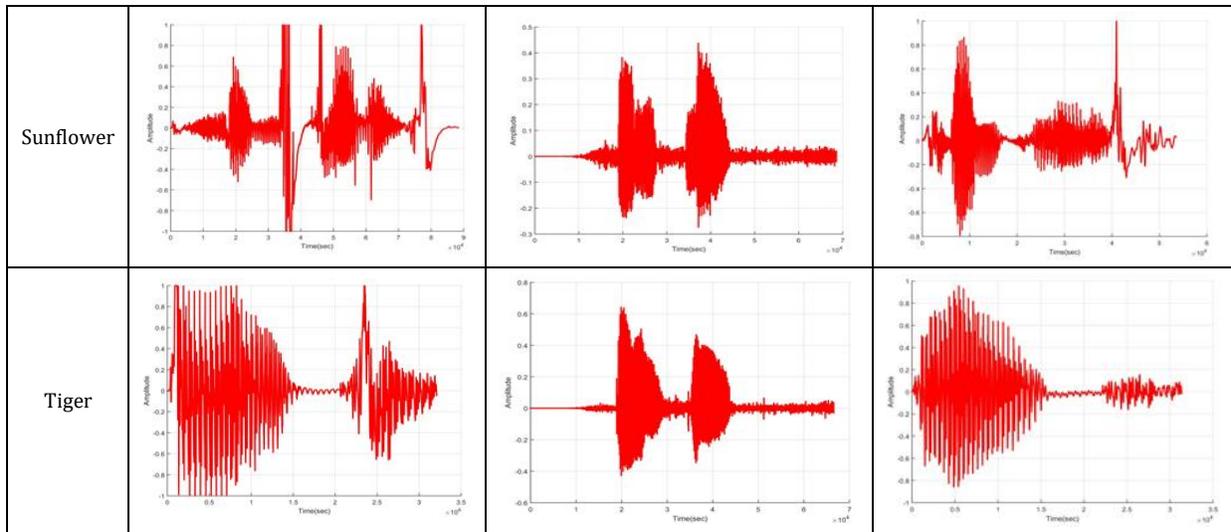


Table 2: Features extracted from the signals

Persons	Signals	Features									
		MFCC									
Person 1	Lion	7.336	7.520	6.397	4.543	4.530	4.241	5.117	6.107	6.375	5.623
	Lotus	6.948	5.934	4.867	5.086	1.931	4.179	5.149	5.667	3.869	4.236
	Peacock	7.050	7.472	8.159	6.462	4.279	4.010	3.215	6.029	4.857	4.623
	Rose	4.854	6.471	5.474	4.771	5.075	6.665	7.373	7.386	5.526	5.057
	Sunflower	3.601	3.419	4.909	5.654	4.211	5.760	5.523	4.350	5.260	3.971
Person 2	Tiger	4.887	4.165	5.136	7.151	8.894	5.533	4.093	6.187	5.193	4.534
	Lion	1.112	2.702	2.917	4.682	5.424	4.269	4.823	4.503	3.687	3.719
	Lotus	0.863	3.067	3.655	6.822	4.089	3.987	3.781	3.655	3.567	3.307
	Peacock	0.822	3.148	3.188	2.521	3.099	3.579	3.692	3.986	3.306	3.243
	Rose	0.734	3.060	4.280	6.153	5.780	3.870	3.561	3.156	4.029	3.860
Person 3	Sunflower	1.287	3.007	4.253	3.715	3.824	5.957	4.336	3.758	3.111	3.373
	Tiger	1.510	3.472	4.347	3.714	4.807	3.925	5.374	3.671	4.073	4.000
	Lion	5.498	5.347	5.994	6.067	5.853	5.199	4.395	5.257	5.952	6.371
	Lotus	5.471	7.112	6.603	4.747	4.407	6.781	7.203	5.672	4.704	3.475
	Peacock	9.237	7.926	5.474	6.865	6.340	4.011	5.554	4.253	3.915	3.872
Person 3	Rose	5.683	7.960	8.710	7.377	5.927	7.127	7.415	5.270	5.787	7.037
	Sunflower	5.110	5.896	3.390	5.542	4.001	5.044	4.473	4.073	6.407	5.601
	Tiger	5.330	5.624	5.604	5.143	6.369	4.766	4.276	5.832	6.105	3.200

Figure 4 and 5 exhibits the performance of applied optimization algorithms for various standard measures.

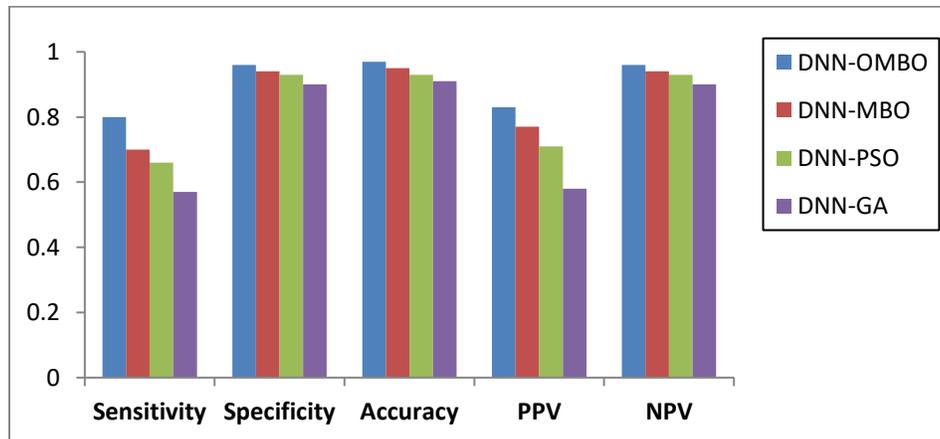


Figure 4: Performance Accuracy of optimization algorithms

Figure 4 shows the performance comparison graph of the optimization algorithms w.r.t Sensitivity, Specificity, Accuracy, PPV, and NPV metrics.

As seen from Figure 4, DNN-OMBO attains an accuracy of 97% followed by DNN-MBO with 95%, DNN-PSO with 93%, and lastly, DNN-GA with 91%.

Regarding Sensitivity measure, DNN-OMBO attains 80%, followed by DNN-MBO with 70%, DNN-PSO with 66%, and finally DNN-GA with 57%.

Regarding the Specificity measure, DNN-OMBO attains 96%, followed by DNN-MBO with 94%, DNN-PSO with 93%, and finally DNN-GA with 90%.

Regarding PPV measure, DNN-OMBO attains 83%, followed by DNN-MBO with 77%, DNN-PSO with 71% and finally DNN-GA with 58%.

Regarding NPV measure, DNN-OMBO attains 96%, followed by DNN-MBO with 94%, DNN-PSO with 93% and finally DNN-GA with 90%.

Hence DNN-OMBO has superior performance compared to the other algorithms, whereas DNN-GA has the least performance.

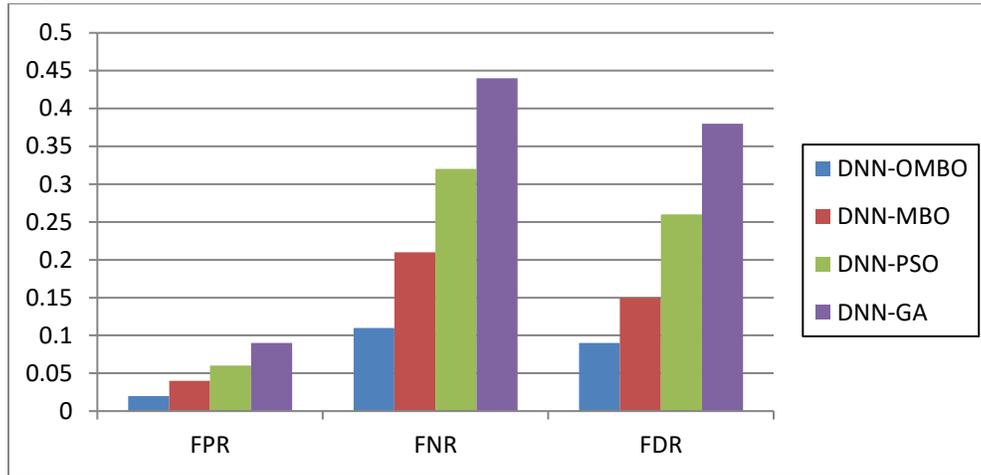


Figure 5: Performance comparison of optimization algorithms w.r.t FPR, FNR, and FDR

Figure 5 shows the performance comparison graph of the optimization algorithms w.r.t FPR, FNR, and FDR metrics.

As seen from the figure, the error rate of prediction for DNN-OMBO is least since it has the highest accuracy (as shown in Figure 4). It has an FPR of 0.02, FNR of 0.11, and FDR of 0.09.

Followed by DNN-OMBO, DNN-MBO has an FPR of 0.04, FNR of 0.21, and FDR of 0.15. In the case of DNN-PSO, it has an FPR of 0.06, FNR of 0.32, and FDR of 0.26. Since DNN-GA has the least accuracy, it has the highest prediction error with FPR of 0.09, FNR of 0.44, and FDR of 0.38.

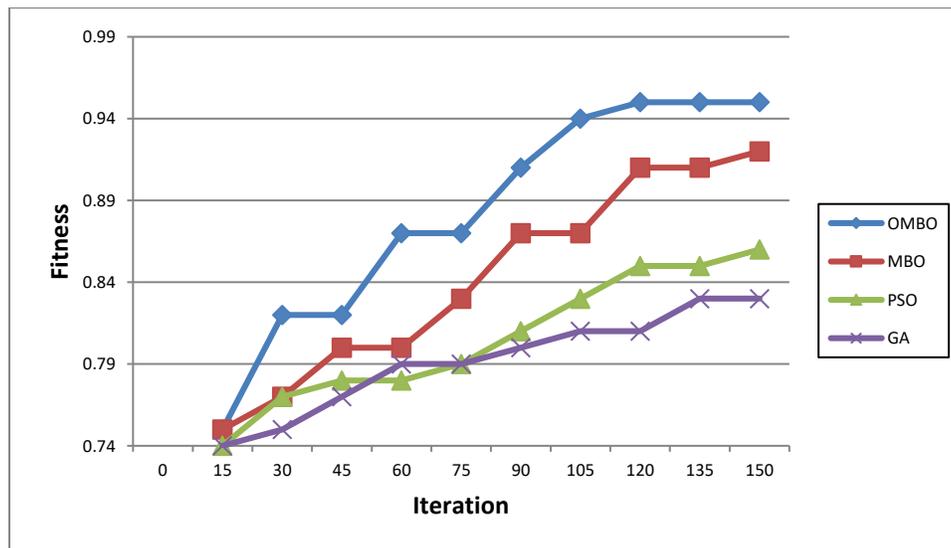


Figure 6: Convergence of optimization algorithms

Figure 6 shows the convergence performance of optimization algorithms to evaluate their accuracy of fitness functions.

Incorporation of Opposition based learning technique enhances the performance of OMBO to achieve the highest fitness results compared to other algorithms. OMBO achieves 95% fitness in

150<sup>th</sup> iteration, followed by MBO with 91%, PSO with 85% and finally, GA with 83%.

Regarding convergence performance, OMBO converges in 100<sup>th</sup> iteration, MBO converges in 120<sup>th</sup> iteration, PSO and GA converge on 135<sup>th</sup> iteration.

## CONCLUSION

In this article, OMBO algorithm is used for optimizing the weight values of DNN. The performance of the proposed DNN-OMBO algorithm has been associated with the optimization algorithms MBO, PSO, and GA. For algorithms, the comparison graph is plotted based on the performance metrics accuracy, sensitivity, specificity PPV, NPV, FPR, FNR, and FDR.

## REFERENCES

1. Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu, "Convolutional Neural Networks for Speech Recognition", *IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, VOL. 22, NO. 10, OCTOBER 2014.
2. Mohammad Hasan Rahmani, Farshad Almasganj\*, Seyyed Ali Seyyedsalehi, "Audio-visual feature fusion via deep neural networks for automatic speech recognition", Elsevier, *Digital Signal Processing* 82(2018), pp:54-63,2018.
3. Toktam Zoughi, Mohammad Mehdi Homayounpour, Mahmood Deypir, "Adaptive Windows Multiple Deep Residual networks for Speech Recognition", Elsevier, *Expert Systems With Applications*, 2019.
4. Aditi Chaturvedi, Priyanka Singh. "The Unheard Pain of Cancer Patients." *Systematic Reviews in Pharmacy* 4.1 (2013), 14-19. Print. doi:10.4103/0975-8453.135832
5. Bhatia, L.C., Naik, R.H.Clinical profile of acute myocardial infarction in elderly patients(2013) *Journal of Cardiovascular Disease Research*, 4 (2), pp. 107-111. DOI: 10.1016/j.jcdr.2012.07.003
6. Gai-Ge Wang, Suash Deb and Zhihua Cui, "Monarch butterfly optimization", Springer, *Neural Comput & Applic*, 2015.
7. Kuniaki Noda, Yuki Yamaguchi, Kazuhiro Nakadai, Hiroshi G. Okuno and Tetsuya Ogata, "Audio-visual speech recognition using deep learning", Springer, *Appl Intell*, DOI 10.1007/s10489-014-0629-7,2014.
8. Nam Xuan Vo, Thuy Van Ha, Usa Chaikledkaew. "The Quality of Life - A Systematic Review Orientation to Establish Utility Score in Vietnam." *Systematic Reviews in Pharmacy* 8.1 (2017), 92-96. Print. doi:10.5530/srp.2017.1.16
9. Negin Najkar, Farbod Razzazi and Hossein Sameti, "A novel approach to HMM-based speech recognition systems using particle swarm optimization", Elsevier, *Mathematical and Computer Modelling* 52 (2010), pp: 1910-1920,2010.
10. Yanmin Qian, Hu Hu, Tian Tan, "Data Augmentation using Generative Adversarial Networks for Robust Speech Recognition", Elsevier, *Speech Communication*,2019.
11. Lin Sun, Suisui Chen, Jiucheng Xu and Yun Tian, "Improved Monarch Butterfly Optimization Algorithm Based on Opposition-Based Learning and Random Local Perturbation", *Hindawi Complexity*, Volume 2019, Article ID 4182148, 20 pages,2019.
12. Firas Hassan, Salam Abd AlQadeem Mohammed, Anil Philip, Ayah Abdul Hameed, Emad Yousif. "Gold (III) Complexes as Breast Cancer Drug." *Systematic Reviews in Pharmacy* 8.1 (2017), 76-79. Print. doi:10.5530/srp.2017.1.13
13. Hui Hu, Zhaoquan Cai, Song Hu, Yingxue Cai, Jia Chen and Sibao Huang, "Improving Monarch Butterfly Optimization Algorithm with Self-Adaptive Population", *Algorithms* 2018, 11, 71; doi:10.3390/a11050071,2018.
14. S. Elavarasi, G. Suseendran, "Automatic Robot Processing Using Speech Recognition System" *Advances in Intelligent Systems and Computing, Data Management, Analytics and Innovation*, Vol.1, October.2019, pp. 185-195, [https://doi.org/10.1007/978-981-32-9949-8\\_14](https://doi.org/10.1007/978-981-32-9949-8_14)
15. S. Elavarasi, G. Suseendran "Design and Implementing Hybrid Semi Neural Network for Speech Recognition", *Journal of Advanced Research in Dynamical and Control Systems*, Vol.11(6), July, 2019 pp.683-694