

An automated speech recognition and feature selection approach based on improved Northern Goshawk optimization

Santosh Kumar Suryakumar^{1,2}, Bharathi S. Hiremath¹, Nageswara Guptha Mohankumar³

¹School of Electronics and Communication Engineering, REVA University, Bengaluru, India

²Department of Electronics and Communication Engineering, Sri Venkateshwara College of Engineering, Bengaluru, India

³Department of Computer Science and Engineering, Sri Venkateshwara College of Engineering, Bengaluru, India

Article Info

Article history:

Received Dec 12, 2022

Revised Feb 13, 2023

Accepted Mar 10, 2023

Keywords:

Automatic speech recognition

Feature selection

K-nearest neighbour

Northern Goshawk optimization

Opposition based learning

ABSTRACT

Automatic speech recognition (ASR) approach is dependent on optimal speech feature extraction, which attempts to get a parametric depiction of an input speech signal. Feature extraction (FE) strategy combined with a feature selection (FS) approach should capture the most important features of the signal while discarding the rest. FS is a crucial process that can affect the pattern classification and recognition system's performance. In this research, we introduce a hybrid supervised learning using metaheuristic technique for optimum FE and FS termed Northern Goshawk optimization (NGO) and opposition-based learning (OBL). Pre-processing, feature extraction and selection, and recognition are the three steps of the proposed technique. The pre-processing is done first to lessen the amount of noise. In the FE stage, we extract features. The OBL-NGO method is used to pick the best collection of extracted characteristics. Finally, these optimised features are utilised to train the k-nearest neighbour (KNN) classifier, and the matching text is shown as the output based on these optimised characteristics of the provided input audio signal. The system's performance is outstanding, and the suggested OBL-NGO is best suited for ASR, according to the testing data.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Santosh Kumar Suryakumar

School of Electronics and Communication Engineering, REVA University

Bengaluru, India

Email: reachsun@gmail.com

1. INTRODUCTION

Speech recognition (SR) technologies are becoming more powerful these days. Speech recognition is a method of detecting a person's uttered words using data in the speech signal [1]. It processes the incoming audio input to make it compatible with various recognition software. Automated speech recognition (ASR) is a term used to describe the process of translating human voice into text using a computer [2]. The usage of machine learning and pattern recognition in SR technologies has skyrocketed in recent years [3]. The task of speech recognition is incredibly difficult for a computer system to do. This is due to the fact that people speak in a variety of ways, resulting in complicated speech signals that must be handled by automated speech signals. ASR technologies must be used to deal with issues. There are numerous main areas of research in voice recognition for the current improvement of spoken language frameworks [4], [5].

Pattern matching is used in current speech recognition frameworks. Hidden Markov models are used as a classifier, which is normally visible. Hidden markov model (HMM)-based voice recognition has progressed significantly, and it can currently achieve high recognition accuracy [6]. For acoustic signals, there exist a variety of parametric representations. Mel-frequency cepstrum coefficients (MFCC) is the most often used among them. There have been several documented collaborations with MFCC, notably in the area of

improving recognition accuracy. Mel frequency cepstral coefficients can be widely used techniques for feature extraction.

Strategies that make advantage of information in the periodicity of speech signals may be utilised to circumvent this issue, despite the fact that speech includes aperiodic content. It has been discovered that feature selection techniques and classification methods play a significant influence in the detection of stuttering occurrences. We employed the feature selection (FS) [7] idea instead of feature extraction to determine the best characteristics that impact the log files to overcome the disadvantages in supervised learning task. FS is a popular supervised learning classification task for extracting useful subsets from large datasets. The raw log files may contain useless and superfluous values in a real-time setting. It has an impact on the predictive model's performance. Filter [8], wrapper [9], and embedded [10] are the three kinds of FS. There are no learning algorithms linked with the filter approach in the prediction model. Correlation [11], Chi-square test [12], analysis of variance (ANOVA) [13], and other popular approaches are only a few examples. The wrapper methods can communicate with the dataset's features. To discover the best outcomes, a learning algorithm is utilised. Particle swarm optimization [14], genetic algorithm (GA) [15], and whale optimization [16], for example, are well-known approaches. The embedded methods are a hybridization of both filter and wrapper. The popular embedding approaches in the literature include least absolute shrinkage and selection operator (LASSO) [17] and ridge regression [18]. However, we applied the metaheuristic (MH) optimization strategy for the suggested approach to determine the best features.

A specific model in this system, support vector machines (SVM) [19], has been detailed by Zhang *et al.* [20] and how it may be used to medium to large vocabulary voice recognition applications. The shape of the joint feature spaces was a critical part of SVMs. The characteristics are obtained using context-dependent generative models, such as hidden Markov models. First, the retrieved characteristics are a function of the utterance's segmentation. A new training procedure was suggested that allowed for the use of universal Gaussian priors in the big margin requirement. Wu *et al.* have built on prior work by using the methodology of second-order cone programming to investigate a convex optimization approach. The second-order cone program (SOCP) approach greatly outperforms the gradient descent method in the test, according to the experimental data [21]. Hirayama *et al.* [22] devised a technique for identifying mixed dialect utterances with multiple dialect language models. Maximization of recognition likelihoods and integration of recognition outcomes were our two key strategies [22]. The influence of MFCCs, energy, formant, and pitch-related variables on boosting the performance of emotion identification systems was explored by Gharavian *et al.* The normalised values of formants were employed as supplemental features to compensate for the influence of mood on recognition rate [23].

The Northern Goshawk optimization (NGO) is a raptor that optimises its hunting strategy [24]. This approach is used by the northern goshawk to select and attack its prey, following which it chases the animal in a pursuit. However, no optimization approach based on the behaviour of northern goshawks has been developed to our knowledge. The researchers took advantage of this information gap in supervised learning by developing a unique optimization approach based on mathematical modelling of the northern goshawk's hunting techniques. NGO is a programme that replicates northern goshawk hunting behaviour. It also achieves better results than regular MH algorithms in areas like as exploration, exploitation, finding local optima, and avoiding premature convergence. Finding the optimal subset of features in FS is challenging, especially when using wrapper-based approaches, despite the advantages of NGO that have been outlined above. This is due to the fact that a learning algorithm (such as a classifier) must be used to assess the chosen subset at each optimisation stage. Thus, we provide a solution to the supervised learning FS issue based on a hybridization of the NGO and opposition based learning (OBL) concept [25] in order to reduce the number of assessments.

The following are the study's main inferences: i) To test our supervised learning strategy, we employed three distinct types of datasets with k-nearest neighbour classifier; ii) Hybrid OBL-NGO supervised learning-based metaheuristic approach is used to find the optimal results; and iii) The suggested supervised learning-based model evaluated in terms of precision, recall, F1-score, classification accuracy and converging ability.

2. PRELIMINARIES

2.1. Initialization

The hunting style of the northern goshawk is separated into two stages, the first of which is a high-speed chase after spotting the prey, and the second of which is a brief tail chase after spotting the victim. The suggested NGO, which is a population-based algorithm, has northern goshawks as searcher members. The population members in the search space are randomly initialised. In the NGO technique is utilised to calculate the population matrix using (1).

$$X = \begin{bmatrix} x_{1,1} & \dots & x_{1,d} & \dots & x_{1,m} \\ x_{i,1} & \ddots & x_{i,d} & \ddots & x_{i,m} \\ x_{N,1} & \dots & x_{N,d} & \dots & x_{N,m} \end{bmatrix} \quad (1)$$

The values collected for the objective function may be expressed as a vector using (2).

$$V(X) = \begin{bmatrix} V_1 = V(X_1) \\ \vdots \\ V_i = V(X_i) \\ \vdots \\ V_N = V(X_N) \end{bmatrix} \quad (2)$$

Where V_i is the objective function (OF) value acquired by the i th suggested solution and V is the vector of achieved OF values. The value of the OF is used to determine which choice is the best. The lower the OF value, the better the suggested solution in minimization difficulties, whereas the higher the OF value, the better the proposed solution in maximisation issues.

2.2. Exploration

The Northern Goshawk chooses a prey at random during the early phase of hunting and attacks it quickly. It improves the NGO's exploration capacity because to the random picking of prey in the search space. In (3)–(5) are utilised to model the concepts presented in the first phase quantitatively.

$$P_i = X_k \quad (3)$$

$$X_{ij}^{new,p1} = \begin{cases} x_{i,j+r}(p_{i,j-1} x_{i,j}) \\ x_{i,j+r}(x_{i,j-p_{i,j}}) \end{cases} \quad (4)$$

$$xi = \begin{cases} X_{ij}^{new,p1}, V_i^{new,p1} < V_i \\ X_i, V_i^{new,p1} \geq V_i \end{cases} \quad (5)$$

Where P_i is the i th northern goshawk's prey position, V_i is the objective function value, and k is a random value [0-1].

2.3. Exploitation

After being attacked by a northern goshawk, the victim tries to run. As a result, the northern goshawk maintains a tail-and-chase pattern when pursuing prey. Simulating this behaviour increases the algorithm's exploitation potential for local search of the search space (6)–(8) is utilised to mathematically represent the concepts presented in the second phase.

$$X_{ij}^{new,p2} = x_{i,j} + R(2r - 1)x_{i,j} \quad (6)$$

$$R = 0.02(1 - \frac{t}{T}) \quad (7)$$

$$i = \begin{cases} X_{ij}^{new,p2}, V_i^{new,p2} < V_i \\ X_i, V_i^{new,p2} \geq V_i \end{cases} \quad (8)$$

2.4. Opposition based learning

An arbitrary number of prior knowledges are used by MH algorithms to generate the initial population of random search agents. The approach iteratively modifies the location of random search agents to find the optimal solution to the optimisation problem at hand. As the optimisation process only goes in one direction, the resulting solution may not be optimal, and the approach may enter local optima. Tizhoosh *et al.* [25] devised the OBL approach in 2005 to address these concerns, which substantially increases the convergence ability of MH algorithms. The opposite search agent which is calculated by (9).

$$\bar{r} = lb + ub - r \quad (9)$$

3. IMPLEMENTATION OF PROPOSED METHODOLOGY

The goal of this study is to demonstrate an efficient recognizer that uses voice recognition techniques to create a human-machine interaction. Pre-processing, feature extraction (FE), and optimal FS for recognition are among the phases in this suggested approach. Pre-processing is carried out to improve the efficiency of the feature extraction procedure. FE is extracting the characteristics of a spoken stream. A feature selection technique is built in order to pick the most optimum collection of extracted characteristics. Finally, these best qualities are used to text recognition. Figure 1 depicts the suggested automatic speech recognition system's process flow.

3.1. Pre-processing

The noise in the incoming voice signals will affect the recognition process. As a result, the first and most important stage in SR is the pre-processing of speech data, which is done to eliminate undesired waveforms and speed up the recognition process. In this pre-processing, a Gaussian filtering is used to remove noise that is related to the spectral subtraction standard. This reduces noise by evaluating the noiseless signal and then comparing it to the original signals. The mathematical representation of the Gaussian filtering's given in the (10).

$$p(k) = \exp\left(\frac{-y^2}{2y\sigma^2}\right) \quad (10)$$

The signal was subjected to a high-pass finite impulse response (FIR) filter with unique component values. The method is completed by recognising the signal's end points and removing the silence. Finally, during the pre-processing stage, a noise-free signal is generated without sacrificing the original data that will be used in the feature extraction procedure.

3.2. Feature extraction

The speech signal is analysed as part of the feature extraction (FE) process. The spectrum analysis approach is used to extract voice signal features. The translation of the input data into a set of features is referred to as feature extraction. Evaluate the following aspects in this study: compare the standard and normal voice signals; peak frequency modulation; MFCC; tri spectral features; and discrete wavelet transform (DWT). In FE, normal speech signals are compared to the standard speech signal. For the individual speech signal, a frequency range of threshold is subsequently determined using the (11). The input signal is compared with the standard signal after feature extraction, which has the most influence on the output signals is determined using the mathematical function in (12). The development of increasing the amplitude of frequencies in relation to the magnitude of other frequencies is known as pre-emphasis. This can be determined using (13).

$$C(k) = S(s) - P(s) \quad (11)$$

$$f(k) = \begin{cases} P + 1 \\ 0 \text{ Otherwise} \end{cases} \quad (12)$$

$$E(z) = 1 - \beta z^{-1} \quad (13)$$

The goal of this section is to use tri-spectral analysis to classify speech signals. In most cases, the voice signals are captured first, and then the tri spectrum properties are analysed. Tri-spectrum statistics are used to identify output speech signals that are not closely related to the resulting input speech signal. This can be determined using the (14). Then we need to calculate the frequency response for the input signal and complex integration of the frequency response is given in the (15)-(16).

$$nft = \max(nft, 2^A) \quad (14)$$

$$T \text{ spec} = X_f * (X_f * X_f) * \text{hankel}(X_{fc}) \quad (15)$$

$$T \text{ spec} = \text{FFTshift}(T \text{ spec}) \quad (16)$$

Each level represents a lower frequency band with a coarser resolution, whereas higher frequency bands represent higher frequency bands. Continuous and discrete transforms are the two primary types of transforms. On the low-pass band, the DWT repeatedly applies a two-channel filter bank (with down sampling).

3.3. Feature selection

We must choose the best set of features for voice signal identification from these extracted features. Since it is a supervised classification task, k-nearest neighbour (KNN) classifier is used to assess the predictive model. The reduced feature set is divided into two divisions using the 10-fold cross-validation (CV) method: training and testing. When a 10-fold CV is utilised, the chance of overfitting the prediction model is lowered. The solution representation and the evaluation function are two critical parts of the optimization issue that must be addressed while constructing any optimizer. To enhance the efficacy of the supervised learning optimization method, the OBL technique in NGO produces an optimal solution for the OF in both directions at the same time. The new model's findings give much superior outcomes in overcoming the problems associated with traditional NGO. The suggested supervised learning algorithm's fitness function is constructed using (17). Subsequently, for the provided input signal, the system generates recognised text. The proposed approach is shown in the Figure 1.

$$Fitness = \alpha R(D) + \beta \frac{|R|}{|N|} \quad (17)$$

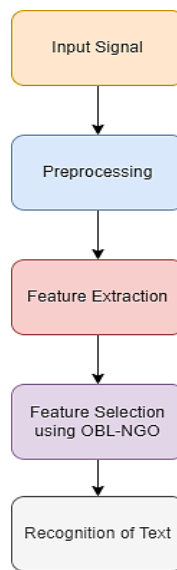


Figure 1. Proposed approach

4. RESULT AND DISCUSSION

4.1. Datasets

We considered three types of datasets in this proposed method: LibriSpeech, CHiME-5 and AISHELL-1. The first dataset corpus is a library of audiobooks from the LibriVox project that totals over 1,000 hours. The overview of the datasets is presented in the Tables 1-3.

Table 1. Overview of LibriSpeech corpus dataset

Subset	Hours	Per-spkr mins	Female	Male	Total
Dev-clean	5.4	8	20	20	40
Dev-clean	5.4	8	20	20	40
Dev-other	5.3	10	16	17	33
Test-other	5.1	10	17	16	33
Train-clean-100	100.6	25	125	126	251
Train-clean-360	363.6	25	439	482	921
Train-clean-500	496.7	30	564	602	1166

Table 2. Overview of CHiME-5 dataset

Dataset	Parties	Speakers	Hours	Utterances
Train	16	32	40:33	79980
Dev	2	8	4:27	7440
Eval	2	8	5:12	11028

Table 3. Overview of AISHELL-1 dataset

Age range	Speakers	Male	Female
16-25	316	140	176
26-40	71	36	35
>40	13	10	3

4.2. Discussion

For each dataset, the suggested model is performed using the Python environment with 100 epochs. Figure 2(a) is LibriSpeech corpus, Figure 2(b) is CHiME-5, Figure 2(c) is AISHELL-1, depicts the prediction model's error rate throughout the procedure. The recommended model's ability to converge to global minima is revealed by the decrease in error rate as the model goes through each aeon. A classifier like KNN is used to evaluate the proposed approach's capacity to forecast. After 25 epochs, the AISHELL-1 dataset begins to converge, whereas the other two datasets begin to converge after 35 epochs. As a result, the average convergence ability of a model is 20 epochs. Furthermore, the performance of the Librispeech datasets is not significantly different from that of the other datasets. It demonstrates that the model is not overfitted. Figure 3(a) is LibriSpeech corpus, Figure 3(b) is CHiME-5, Figure 3(c) is AISHELL-1 depicts the KNN classifier's prediction capabilities. During training and testing, the recommended technique provides accuracy values between [0.68, 0.95] and [0.55, 0.82]. The recommended model has a stronger proclivity to improve accuracy in the early stages. The difference between trading and testing accuracy should be narrowed. The gap in the AISHELL-1 dataset is relatively large when compared to other datasets [range: 0.55 to.70]. Figure 4(a) is LibriSpeech corpus, Figure 4(b) is CHiME-5, and Figure 4(c) is AISHELL-1, depicts the receiver operastic charecteristic (ROC) curve for the recommended model. Based on the higher ROC value for datasets, the recommended approach may assign a larger chance to a randomly selected genuine positive sample than a negative sample on average. The component selection is verified using the KNN classifier. The wider area under the ROC curve (ROCAUC) curve of the recommended model shows that the qualities it picked can provide significant confidence in knowledge discovery and decision making. Furthermore, a smaller set of well selected characteristics can yield more accurate findings than the entire set of features in the input data. Table 4 summarises and compares the performance assessments of the prediction models for the specified feature subsets in order to validate the feature subsets. The table shows that the suggested model had the highest classification accuracy across all datasets.

Table 4. Validation of selected features

Dataset	Precision	Recall	F1-score	Accuracy
LibriSpeech corpus	0.79	0.93	0.89	0.98
CHiME-5	0.72	0.81	0.77	0.91
AISHELL-1	0.91	0.97	0.85	0.96

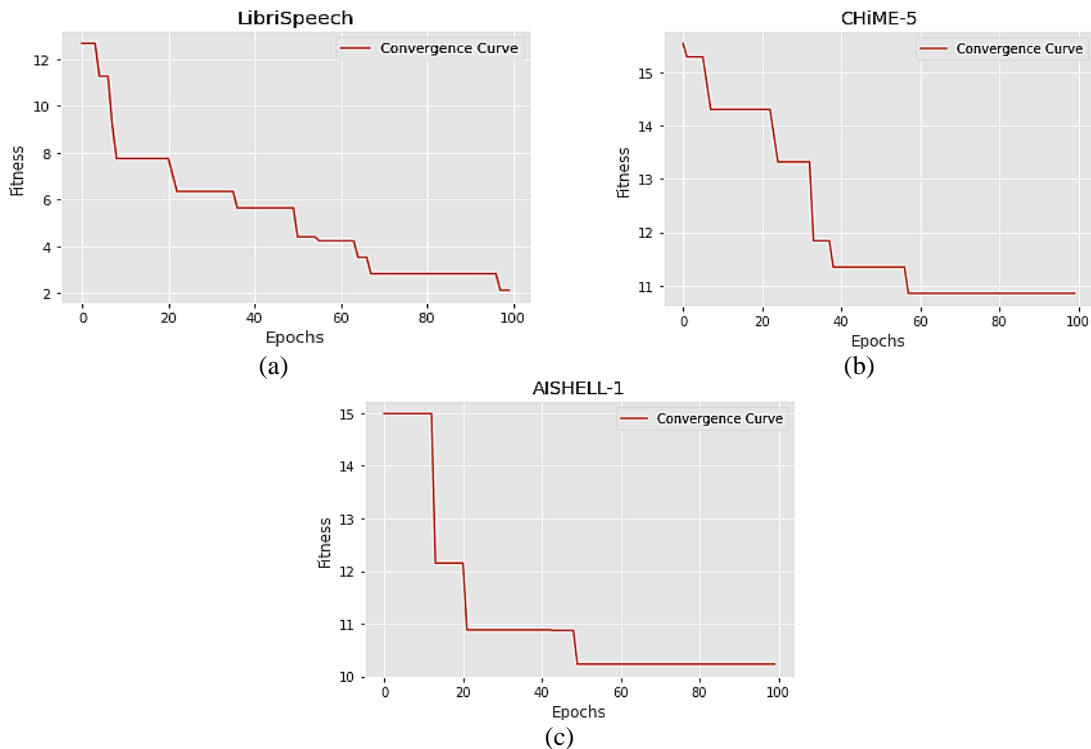


Figure 2. Converging ability: (a) LibriSpeech corpus, (b) CHiME-5, and (c) AISHELL-1

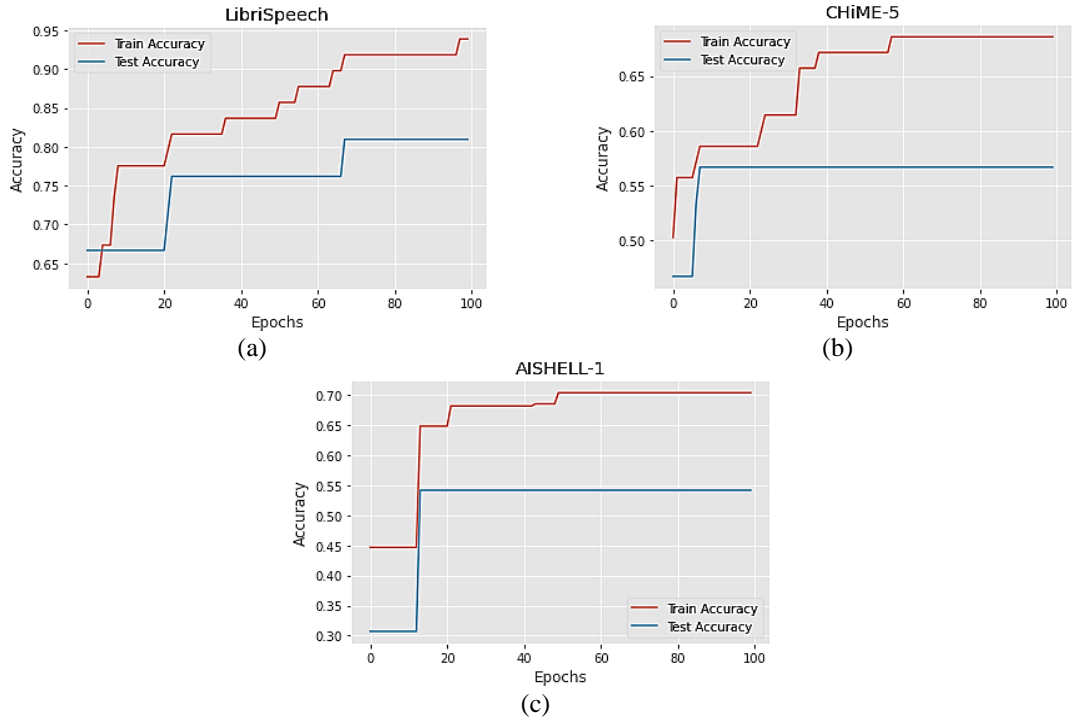


Figure 3. Train and Test accuracy of suggested model: (a) LibriSpeech corpus, (b) CHiME-5, and (c) AISHELL-1

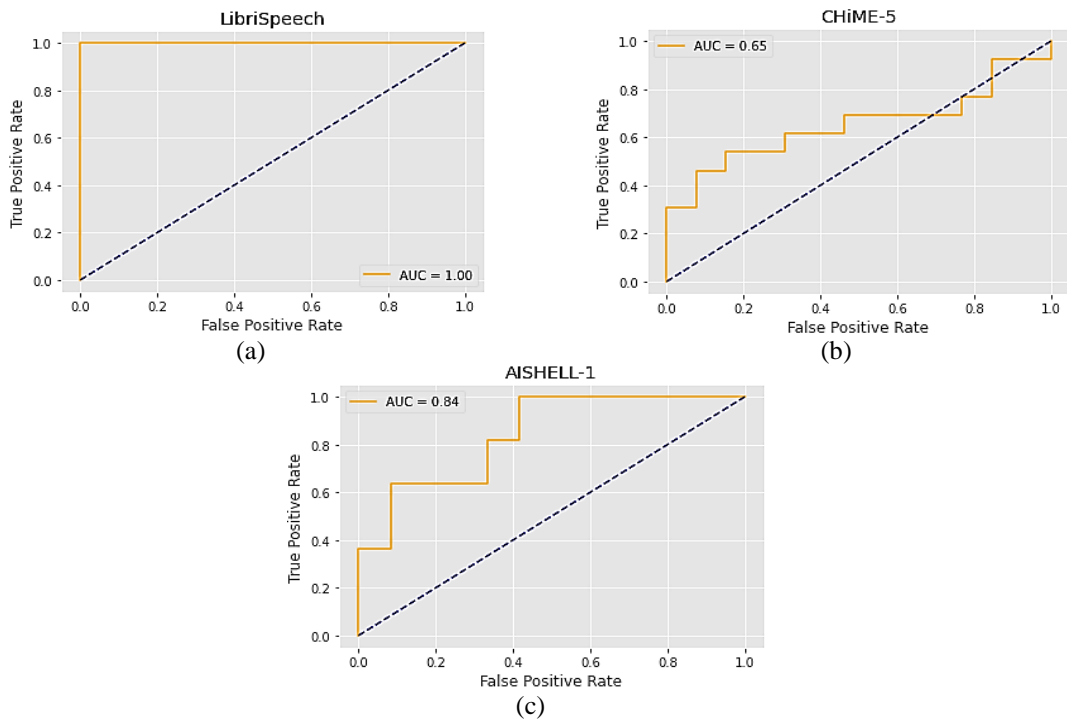


Figure 4. ROC score: (a) LibriSpeech corpus, (b) CHiME-5, and (c) AISHELL-1

5. CONCLUSION

This paper suggests a unique supervised learning based optimal approach for speech recognition sensing based on feature selection and improved OBL-NGO optimization method. Feature extraction is

commonly utilised in the literature in conjunction with typical ML classifiers. However, in order to detect speech recognition, the suggested model is employed to pick the ideal features utilising an OBL-NGO optimization strategy. Since it is using the supervised learning classification task, in terms of convergence ability, training and testing accuracy, precision, recall, and F1-score, the proposed model's performance is evaluated based on three benchmark datasets. Except for the CHiME-5 dataset, the remaining datasets excel in all assessment measures, according to the results. The proposed approach may be implemented using a variety of ML classifiers, and a deep learning environment will be used in the future.

ACKNOWLEDGEMENTS

Authors acknowledge the support from REVA University for the facilities provided to carry out the research.




REFERENCES

- [1] L. Kerkeni, Y. Serrestou, K. Raoof, M. Mbarki, M. A. Mahjoub, and C. Cleder, "Automatic speech emotion recognition using an optimal combination of features based on EMD-TKEO," *Speech Communication*, vol. 114, pp. 22–35, 2019, doi: 10.1016/j.specom.2019.09.002.
- [2] T. Tuncer, S. Dogan, and U. R. Acharya, "Automated accurate speech emotion recognition system using twine shuffle pattern and iterative neighborhood component analysis techniques," *Knowledge-Based Systems*, vol. 211, 2021, doi: 10.1016/j.knsys.2020.106547.
- [3] Z. T. Liu, M. Wu, W. H. Cao, J. P. Xu, and G. Z. Tan, "Speech emotion recognition based on feature selection and extreme learning machine decision tree," *Neurocomputing*, vol. 273, pp. 271–280, 2018, doi: 10.1016/j.neucom.2017.07.050.
- [4] J. Ancilin and A. Milton, "Improved speech emotion recognition with Mel frequency magnitude coefficient," *Applied Acoustics*, vol. 179, 2021, doi: 10.1016/j.apacoust.2021.108046.
- [5] O. Atila and A. Şengür, "Attention guided 3D CNN-LSTM model for accurate speech based emotion recognition," *Applied Acoustics*, vol. 182, 2021, doi: 10.1016/j.apacoust.2021.108260.
- [6] Y. B. Singh and S. Goel, "A systematic literature review of speech emotion recognition approaches," *Neurocomputing*, vol. 492, pp. 245–263, 2022, doi: 10.1016/j.neucom.2022.04.028.
- [7] G. Kou, P. Yang, Y. Peng, F. Xiao, Y. Chen, and F. E. Alsaadi, "Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods," *Applied Soft Computing Journal*, vol. 86, 2020, doi: 10.1016/j.asoc.2019.105836.
- [8] J. Biesiada and W. Duch, "Feature selection for high-dimensional data - A pearson redundancy based filter," *Advances in Soft Computing*, vol. 45, pp. 242–249, 2007, doi: 10.1007/978-3-540-75175-5_30.
- [9] M. Kadhum, S. Manaseer, and A. L. A. Dalhoum, "Evaluation feature selection technique on classification by using evolutionary ELM wrapper method with features priorities," *Journal of Advances in Information Technology*, vol. 12, no. 1, pp. 21–28, 2021, doi: 10.12720/jait.12.1.21-28.
- [10] T. Windeatt, R. Duangsoithong, and R. Smith, "Embedded feature ranking for ensemble MLP classifiers," *IEEE Transactions on Neural Networks*, vol. 22, no. 6, pp. 988–994, 2011, doi: 10.1109/TNN.2011.2138158.
- [11] M. A. Hall, "Correlation-based feature selection for machine learning," *Univ. Waikato*, vol. 19, no. April, 1999.
- [12] X. Jin, A. Xu, R. Bie, and P. Guo, "Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3916 LNBI, pp. 106–115, 2006, doi: 10.1007/11691730_11.
- [13] W. S. Albaldawi and R. M. Almuttairi, "Hybrid ANOVA and LASSO methods for feature selection and linear support vector, multilayer perceptron and random forest classifiers based on spark environment for microarray data classification," *IOP Conference Series: Materials Science and Engineering*, vol. 1094, no. 1, p. 012107, 2021, doi: 10.1088/1757-899x/1094/1/012107.
- [14] M. E. H. Pedersen and A. J. Chipperfield, "Simplifying particle swarm optimization," *Applied Soft Computing Journal*, vol. 10, no. 2, pp. 618–628, 2010, doi: 10.1016/j.asoc.2009.08.029.
- [15] R. Dhanalakshmi, P. Parthiban, K. Ganesh, and T. Arunkumar, "Genetic algorithm to solve multi-period, multi-product, bi-echelon supply chain network design problem," *International Journal of Information Systems and Supply Chain Management*, vol. 2, no. 4, pp. 24–42, 2009, doi: 10.4018/jisscm.2009062902.
- [16] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016, doi: 10.1016/j.advengsoft.2016.01.008.
- [17] R. Muthukrishnan and R. Rohini, "LASSO: A feature selection technique in predictive modeling for machine learning," *2016 IEEE International Conference on Advances in Computer Applications, ICACA 2016*, pp. 18–20, 2017, doi: 10.1109/ICACA.2016.7887916.
- [18] S. Paul and P. Drineas, "Feature selection for ridge regression with provable guarantees," *Neural Computation*, vol. 28, no. 4, pp. 716–742, 2016, doi: 10.1162/NECO_a_00816.
- [19] E. Tuba, I. Strumberger, T. Bezdan, N. Bacanin, and M. Tuba, "Classification and feature selection method for medical datasets by brain storm optimization algorithm and support vector machine," *Procedia Computer Science*, vol. 162, pp. 307–315, 2019, doi: 10.1016/j.procs.2019.11.289.
- [20] S. X. Zhang and M. J. F. Gales, "Structured SVMs for automatic speech recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 3, pp. 544–555, 2013, doi: 10.1109/TASL.2012.2227734.
- [21] D. Wu, H. Jiang, and Y. Yin, "Large-margin estimation of hidden markov models with second-order cone programming for speech recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 6, pp. 1652–1664, 2011, doi: 10.1109/TASL.2010.2096213.
- [22] N. Hirayama, K. Yoshino, K. Itoyama, S. Mori, and H. G. Okuno, "Automatic speech recognition for mixed dialect utterances by mixing dialect language models," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 23, no. 2, pp. 373–382, 2015, doi: 10.1109/TASLP.2014.2387414.




- [23] D. Gharavian, M. Sheikhan, and F. Ashoftehd, "Emotion recognition improvement using normalized formant supplementary features by hybrid of DTW-MLP-GMM model," *Neural Computing and Applications*, vol. 22, no. 6, pp. 1181–1191, 2013, doi: 10.1007/s00521-012-0884-7.
- [24] M. Dehghani, S. Hubalovsky, and P. Trojovsky, "Northern Goshawk optimization: A new swarm-based algorithm for solving optimization problems," *IEEE Access*, vol. 9, pp. 162059–162080, 2021, doi: 10.1109/ACCESS.2021.3133286.
- [25] H. R. Tizhoosh, "Opposition-based learning: A new scheme for machine intelligence," *Proceedings - International Conference on Computational Intelligence for Modelling, Control and Automation, CIMCA 2005 and International Conference on Intelligent Agents, Web Technologies and Internet*, vol. 1, pp. 695–701, 2005, doi: 10.1109/cimca.2005.1631345.

BIOGRAPHIES OF AUTHORS






Santosh Kumar Suryakumar    got his under-graduation degree in 2005 from SJC Institute of Technology (SJCIT) Chickabalur in Electronics and Communication Engineering branch. He completed Master's in Digital Communication from M.S. Ramaiah Institute of Technology Bengaluru in the year 2009. He is currently pursuing Doctoral Degree on the Speech Signal processing REVA University, Bengaluru, India. He can be contacted at email: reachsun@gmail.com.



Dr. Bharathi S. Hiremath    has more than 29 years of experience in teaching. She has published 38 reputed papers in the high indexed journals. Her research interests are Image Processing, Video processing, Antenna design, VLSI, Device Modelling and Sensors. She can be contacted at email: bharathish@reva.edu.in.



Dr. Nageswara Gupta Mohankumar    has 20 years of experience in teaching and 13 years of experience in research. He received B.E. and M.E. Computer Science and Engineering from Kumaraguru College of Technology and completed his Doctorate in PSG College of Technology under Anna University, Chennai. His area of interest includes business intelligence, service-oriented computing, user interface design, aquaponics, RPA and robotics. He can be contacted at email: mngupthasvce@gmail.com.