



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Special Issue 2, December 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.569

LPC based BFO with Radial Basis Function Neural Network for Speaker Recognition

P S Subhashini Pedalanka¹, Dr M. Satya Sai Ram², Dr Duggirala Sreenivasa Rao³

Associate Professor, Department of E.C.E, R.V.R & JC College of Engg., Guntur, India¹

Research scholar, Department of E.C.E, JNTUH, Hyderabad, India¹

Professor, Department of E.C.E, R.V.R & JC College of Engineering, Guntur, India²

Professor, Department of E.C.E, JNTUH, Hyderabad, India³

ABSTRACT: Speaker recognition is essential in the field of authentication and surveillance to validate the identity of the user using the extracted features of the audio speech signal. A novel LPC-based Bacterial Foraging Optimization (BFO) with Radial Basis Function (RBF) for identification of exact speaker is proposed in this paper. The speaker recognition function can be done with two modules, namely Training and Testing. Initially throughout the training step, the Liner Predictive Coefficients (LPCs) of each speech sample are derived by pre-processing the audio speech signal. Then the features are classified towards the target speaker using RBFNN. Bottleneck features for abstract representation are extracted in RBFNN and the dimensionality by Principal Component Analysis (PCA) is further decreased. To improve the classification accuracy, the features are optimized with Bacterial Foraging optimization (BFO) algorithm. Finally, the probability score for each speaker is generated to identify the speaker. The proposed LPC based BFO using Bottleneck features (LBFOB) method is evaluated with TIMIT data corpus. The improved performance is presented with regard to Equal Error Rate (EER).

KEYWORDS: Speech processing, speaker recognition, LPC extraction, RBF neural network, and Bacterial foraging optimization, scoring

I. INTRODUCTION

Speaker recognition plays a major role in communication and surveillance area. This recognition process was evaluated by matching the training data with the test data. In recent years number of experiments were done to improve this recognition approach. The existing approaches outcomes include some trouble causing factors they are linear channel distortion, reverberation, and additive noise [1]. The speed enhancement and speaker recognition are still considered as a challenging task due to the presence of non-stationary noise and reverberation [2]. The classifier modeling and feature extraction of audio are the two important components in speaker recognition [23].

The best generally used feature in recognition of speakers was LPC. The obtained LPC features are calculated for the entire training set samples and they are stored for speaker recognition. LPC feature computes both the training and testing set samples [5]. Bacterial algorithms on optimizing foraging (Passino, 2002; Biswas et al., 2007) have been implemented to increase the way natural selection appears to eradicate species with low capacity to find and digest survival food [11]. The experimental results illustrate that BFO model is better than other traditional algorithms. This BFO system has therefore been integrated with the RBFNN for optimum classification performance [12].

The existing methods introduce a number of speaker recognition algorithms. None of these existing approaches provides an algorithm to increase the precision of the input signals. This article proposes a BFO algorithm based on LPC to improve speaker recognition accuracy. In recent works the amount of input data was reduced to improve the accuracy and perform number of iterations to identify the appropriate speaker. But in this approach, the issues regarding the accuracy was removed by injecting LPC based BFO. The entire unpredictable covariance matrix was determined by this LPC in single iteration. Finally, the perfectly optimized result was obtained by the BFO algorithm.

The proposed methods outline is as follows. The related work for speaker recognition was deliberated briefly in Section 2. Section 3 designates the proposed LBFOB approach. The evaluation for results and performance of proposed LBFOB approach was discussed in Section 4. Finally, section 5 concludes the proposed work.

II. RELATED WORK

Fred Richardson et.al, 2015, [16], presented a new method for speaker recognition (SR) and language recognition (LR) using DNN. The improved performance obtained by DNN in automatic speech recognition (ASR) encourage it to employ its application in speaker and language recognition. The important benchmarks provided for both SR and LR was DAC13 (Domain Adaptation Challenge 2013) and LRE11 (language recognition evaluation 2011). In this method, both the integrated DNN and baseline contains the i-vector classifier. The tandem features or score fusion was used by DAC13 SR to reduce the error rates. The tandem features provide better performance in SR, but in LRE11 performance, the score fusion does not provide any significant improvements.

Zhaofeng Zhang, et al, 2015, [18], introduced the DNN with de-noising autoencoder (DAE) based dereverberation and for recognition of speaker, bottleneck features are considered. In this, the distance-talking set was selected for speaker identification. This method introduce the multichannel least mean squares (MCLMS) method to reduce the error rates. The corresponding error rate obtained by this method for bottleneck feature is 21.4% and for the auto encoder feature the error rate was 47.0%. Furthermore, the performance of this method was enhanced by fusing the DAE-based de-reverberation and the DNN-based bottleneck feature. The two features used by DNN train Gaussian mixture model (GMM) for recognition of speakers. The likelihood features obtained from both the DAE and bottleneck enhances the speaker recognition performance.

III. SPEAKER RECOGNITION WITH LBFOB

Speaker identification is the process by which the person can be identified from voice characteristics. It is the way that the well-trained voice of a particular person is translated. It may also be used along with safety procedure to authenticate or validate the speaker identity. The identification of speakers is based on extraction and classification of speakers. The complete design of the work is seen in Fig 1.

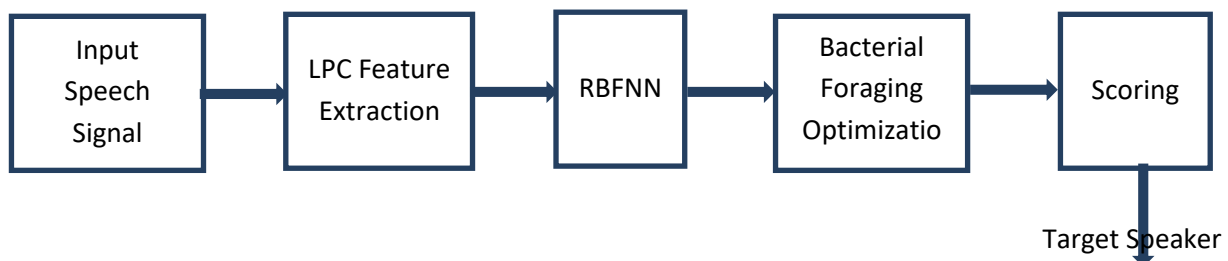


Fig 1: Proposed Speaker Recognition system Block diagram

The proposed architecture divides the speaker recognition task into two modules. They are LPCfeature extraction and speaker identification. Initially, the feature vector from the speech signal is obtained as LPC feature. The input for the neural network is LPC feature vector which optimizes the features at the output layer and they are scored to identify the specific speaker.

A. Preprocessing

The main purpose of this parameterization is to remove appropriate information by eliminating redundant information from speech waveforms. This process is performed frame by frame and each frame is converted to a single N-dimensional feature vector. Within each frame, the number of samples is greater than the value of N. This provides the requisite data for the backend device processing, in which the volume of data is reduced. The audio input is converted into a vector sequence for extraction,

$$X = [x_1, x_2, \dots, x_k \dots],$$

here k denotes frame index, x_k denotes N-dimensional vector.

Linear Predictive Coding Linear predictive coding (LPC) is a tool used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model. It is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters

The LPC method is quite close to the FFT. The envelope is calculated from a number of formants or poles specified by the user. The formants are estimated removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. The removing process is called inverse filtering, and the remaining signal is called the residue. In this analysis first convert each frame of p+1 autocorrelations into LPC parameter set by using Durbin's method.

B. RBF and Bacterial Foraging Optimization (BFO) based classification

RBF is a state of the art technique which achieves better recognition performance in speaker recognition. RBF is considered with multiple non-linear or linear hidden layers that represents data in encoded form. The main concept of RBF is activating the current output layer to the input of next hidden layer. Identification capability is enhanced by using large number of hidden layers. The relationship between the input layers and the first hidden layer is given by,

$$a_1 = F(W_1 x + b_1) \quad (3)$$

Where W_1 and b_1 are weight, bias matrix of first layer and the activation function gaussian is denoted as F(.). The logistics function special case can be defined as,

$$F(x) = e^{-x^2} \quad (4)$$

Here, x denotes the input of activation. The mapping between the current and next hidden layer after getting the first hidden layer is given as,

$$a_l = F(W_l a_{l-1} + b_l), l = 2, \dots, L \quad (5)$$

Where L signifies total quantity of layers, a_{l-1} represent the first layer, F(.) represent the activation function.

For speaker recognition, G(.) is applied in output layer. Hence the output of RBF is represented by,

$$\hat{y} = G(a_L) \quad (6)$$

Where, G(.) is the softmax function and for the label y.

For RBF training, the training data $X = [x_1, \dots, x_i, \dots, x_n]$ and the output labels are $Y = [y_1, \dots, y_i, \dots, y_N]$ where, N denotes the total number of training samples. The cost function is denoted as,

$$C(Y, \hat{Y}; X, \theta) = \frac{-1}{NJ} \sum_{i=1}^N \sum_{j=1}^J [y_{i,j} \log \hat{y}_{i,j}] \quad (7)$$

Where, $\hat{Y} = [\hat{y}_1, \dots, \hat{y}_i, \dots, \hat{y}_N]$ denotes the DNN-RBF output, $\hat{y}_{i,j}$ and $y_{i,j}$ is the jth element of \hat{y}_i and y_i respectively. The value of R(W) is calculated as,

$$R(W) = \sum_l \|W_l\|_F^2 \quad (8)$$

Where, $\rho(A)$ represents the penalty sparsity of the hidden layer output, $\|\cdot\|_F^2$ is the Frobenius norm, η and γ denotes the controlling coefficients.

i) Bottleneck feature extraction and dimensionality reduction

The bottleneck features (BFs) from the hidden layer of neural network are extracted which contains fewer units than other hidden layers. This layer creates the abstract features which is suitable for speaker recognition in DNN-RBF. BFs are generally used in auto encoders and to predict input features, a neural network is trained. The output of hidden layers can be used as the features for further processing. Nonlinear transformations are represented for preprocessing the speech signal. The dimensionalities of bottleneck features are required to be reduced due to its large size. Principal component analysis (PCA) is a technique which reduce the original variable p by q number of derivative variables. Principal components are denoted as the linear combinations of original variables with reduced size than the original variables. Due to its simplicity, it is used in dimensionality reduction of features.

When using large set of features, measurements of these features are required to be processed. It is possible to lessen the features when much of them are available. Consider the vector x of p variables and n measurements. For reducing the dimension from p to q , PCA detects the linear combinations $a_1'x, a_2'x, \dots, a_q'x$ which is known as principal components. It has maximum variance and they are being correlated with other $a_k'x$. The Eigen vectors a_1, a_2, \dots, a_q of the covariance matrix S correspond to q largest Eigen value are used to solve the given maximization problem. This value provides variance for their principal components and the proportion of total variance. The first q principal components are denoted as the ratio between q Eigen values and p original variables.

ii) Bacterial Foraging optimization algorithm

A BFO algorithm consists of multiple bacteria that consists optimization solutions and consist of three processes: chemotaxis, replication and elimination dispersal.

In chemotaxis, a random-directed bacterium represents a fall and a bacterium in the same direction indicates a sprint. All bacteria are next sorted, and only half of the fittest bacteria survive during the reproduction cycle. The remaining bacteria are then split into two similar bacteria, which form new bacteria. Finally, a bacterium is probabilistically chosen to move to a different random location in the search area during the elimination / dispersion procedure. While this action preserves the diversity of the output, the optimization cycle may be interrupted and thus carried out after many stages of the reproduction process[28].

C. Scoring

The work proposed is implemented as a framework level so that each frame is fed to the network and for each target speaker a class posterior likelihood is achieved. This makes MBFOB suitable for real-time applications as the decision concerning the target speaker is made in each frame. In the optimization process, the input from the past system is combined with the new frame. The probability and the definition of target speakers are defined as

$$P(y) = \frac{e^y}{\sum_{k=1}^K e^y} \quad (9)$$

If the output layer is indicated, K is the total class number and the rear of the target speaker is described as $P(y)$. Using the output layer, you can make flexible decisions on the target speaker. The target speaker is highly scored based on the performance criterion by using this probability measure.

IV. PERFORMANCE ANALYSIS

The TIMIT data corpus assesses the proposed LBFOB method for the recognition of speakers. Information about the data corpus and data from the extraction and description of the function are presented in this section. Evaluation of performance measures such as EER, DCF, Cavg, precision, recall and comparison of test results with the current methods.

A. Corpus

For the proposed LBFOB speaker recognition function, the TIMIT read corpus is used. It contains a total of 6300 phrases, 10 phrases, spoken by 630 speakers from eight major United States dialect regions. For each

utterance, the TIMIT corpus includes a 16-bit 16 kHz waveform file. The entire archive is marked as files for training and research. Seven audio files for training are used for each speaker and three files for testing.

B. Parameter selection

The speech is analyzed with a 20-ms Hamming window with a fixed frame rate of 10-ms to extract features. The LPC feature vector dimension is says the number of units in input layer for RBF+BFO, and the output layer with 64 units are used. There are 6 hidden layers and the bottleneck layer has been used with 64 units. For the purpose of comparison, the work is simulated with other approaches like multilayer perceptron, back propagation, and Radial basis function.

C. Equal Error Rate (EER)

The EER is well-defined where the false acceptance rate (FAR) is equal to the false rejection rate (FRR). This measure is considered for predicting the accuracy of the speaker recognition. Parameters such as FAR and FRR are tested as follows.

$$FAR = \frac{\text{No.of false accep tan ce}}{\text{No.of identification attempt}} \quad (10)$$

$$FRR = \frac{\text{No.of false rejection}}{\text{No.of identification attempt}} \quad (11)$$

Both these parametersevaluates the number of incorrect acceptance and number of incorrect rejection. The EER value of various features included in number of existing and proposed approaches are depicted in table 1, table 2 and table 3.

Table 1: Speaker identification approaches with EER comparison

Feature extraction method	EER comparison for Existing Speaker Recognition approaches		
	RBF	BP	MLP
LPC	0.0265	0.055	0.0722

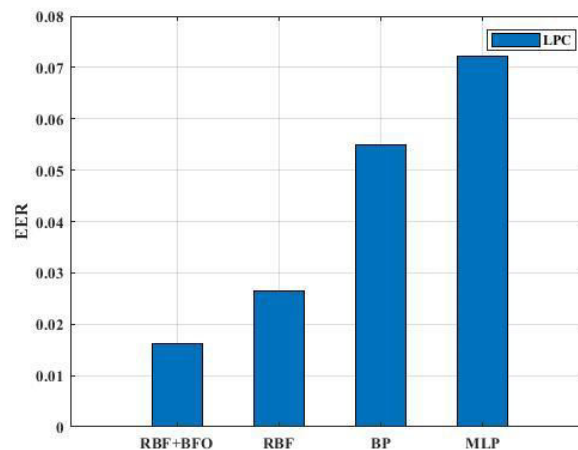
The EER of this RBF with LPC provide minimum error rate value.

Table 2 : EER comparison for Proposed **RBF+BFO** Speaker Recognition Approach

Feature extraction methods	EER comparison for Proposed RBF+BFO Speaker Recognition
LPC	0.0162

Table 3:EER comparison for **Existing** and **Proposed** Speaker Recognition

Feature extraction methods	EER comparison for Existing and Proposed Speaker Recognition approaches			
	RBF+BFO	RBF	BP	MLP
LPC	0.0162	0.0265	0.055	0.0722

**Figure 2.** Comparison graph for the EER values

The obtained results of this proposed approach LPC-RBF+BFO is found to be superior than the other three existing approaches.

The results of this proposed performance factors provide better results for speaker recognition. The proposed LPC-RBF+BFO speaker recognition system achieve high accuracy and better performance than other methods.

E. Accuracy

It determines the system ability for the accurate detection of speaker. The expression for accuracy is shown in Equation. (12),

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

The accuracy of the methodology proposed was much higher than the three other approaches currently in use. As a result, the time for this proposed approach is popular. The precision and time-consuming results of this proposal and of the three other existing methods are shown below and are shown in Table 4.

Table 4. Comparison results for the accuracy of proposed and existing methods

LPC based Speaker recognition systems	Accuracy (%)
RBF+BFO	98.04
RBF	97.4
BP	96.15
MLP	94.34

The accuracy shown in table 4, indicates, the proposed methods performance is found to be better than the existing method. The comparison graph for accuracy of this proposed and some other existing methods are highlighted in Figure 3.

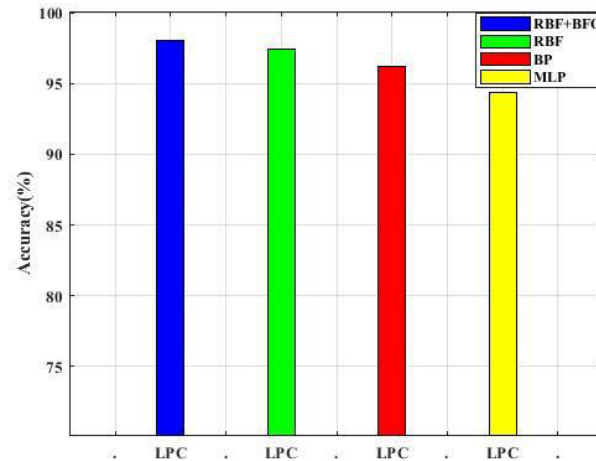


Figure 3. Comparison results for accuracy

The accuracy value of this proposed LPC-RBF+BFO for speaker recognition is found to be better than the other methods. The obtained values indicate that our proposed method has higher performance due to the involvement of BFO algorithm.

V. CONCLUSION

This paper suggests an LBFOB solution based on LPC vector elimination and RBF with BFO, for the identification of speakers. The speech utterance from the TIMIT data corpus is preprocessed to obtain LPC feature vectors. RBF is used for the purpose of classifying the speaker and the feature vectors in the output layers are optimized with Bacterial Foraging optimization. Finally, the scores for each speaker are calculated to identify the speaker. For speaker recognition, both TIMIT data sets are taken into account. Different output metrics like EER and accuracy are used to test the proposed speaker recognition technique. The execution time of this proposed method is found to be lesser than the other existing methods. The experimental findings are contrasted with other current methods and it shows the efficiency of our approach.

REFERENCES

1. Kim, C. and Stern, R.M., 2016. Power-normalized cepstral coefficients (PNCC) for robust speech recognition. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 24(7), pp.1315-1329.
2. Vincent, E., Watanabe, S., Nugraha, A.A., Barker, J. and Marxer, R., 2017. An analysis of environment, microphone and data simulation mismatches in robust speech recognition. *Computer Speech & Language*, 46, pp.535-557.
3. Mannepalli, K., Sastry, P.N. and Suman, M., 2016. MFCC-GMM based accent recognition system for Telugu speech signals. *International Journal of Speech Technology*, 19(1), pp.87-93.
4. Wang, K., An, N., Li, B.N., Zhang, Y. and Li, L., 2015. Speech emotion recognition using fourier parameters. *IEEE Transactions on Affective Computing*, 6(1), pp.69-75.
5. Borde, P., Varpe, A., Manza, R. and Yannawar, P., 2015. Recognition of isolated words using Zernike and MFCC features for audio visual speech recognition. *International journal of speech technology*, 18(2), pp.167-175.
6. Singer, E. and Reynolds, D.A., 2015. Domain mismatch compensation for speaker recognition using a library of whiteners. *IEEE Signal Processing Letters*, 22(11), pp.2000-2003.
7. Gonzalez-Dominguez, J., Lopez-Moreno, I., Moreno, P.J. and Gonzalez-Rodriguez, J., 2015. Frame-by-frame language identification in short utterances using deep neural networks. *Neural Networks*, 64, pp.49-58.

8. Cumani, S., Laface, P., Cumani, S. and Laface, P., 2017. Nonlinear i-vector transformations for PLDA-based speaker recognition. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 25(4), pp.908-919.
9. Miao, Y., Zhang, H. and Metze, F., 2015. Speaker adaptive training of deep neural network acoustic models using i-vectors. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 23(11), pp.1938-1949.
10. Liu, Z., Wu, Z., Li, T., Li, J. and Shen, C., 2018. GMM and CNN hybrid method for short utterance speaker recognition. *IEEE Transactions on Industrial Informatics*, 14(7), pp.3244-3252.
11. Kar, A.K., 2016. Bio inspired computing—A review of algorithms and scope of applications. *Expert Systems with Applications*, 59, pp.20-32.
12. Cai, Z., Gu, J., Wen, C., Zhao, D., Huang, C., Huang, H., Tong, C., Li, J. and Chen, H., 2018. An Intelligent Parkinson's Disease Diagnostic System Based on a Chaotic Bacterial Foraging Optimization Enhanced Fuzzy KNN Approach. *Computational and mathematical methods in medicine*, 2018.
13. Fayek, H.M., Lech, M. and Cavedon, L., 2017. Evaluating deep learning architectures for Speech Emotion Recognition. *Neural Networks*, 92, pp.60-68.
14. Jia, F., Lei, Y., Lin, J., Zhou, X. and Lu, N., 2016. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, 72, pp.303-315.
15. Ghahabi, O. and Hernandez, J., 2017. Deep learning backend for single and multisession i-vector speaker recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(4), pp.807-817.
16. Richardson, F., Reynolds, D. and Dehak, N., 2015. Deep neural network approaches to speaker and language recognition. *IEEE Signal Processing Letters*, 22(10), pp.1671-1675.
17. Stafylakis, T., Alam, M.J. and Kenny, P., 2016. Text-dependent speaker recognition with random digit strings. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 24(7), pp.1194-1203.
18. Zhang, Z., Wang, L., Kai, A., Yamada, T., Li, W. and Iwahashi, M., 2015. Deep neural network-based bottleneck feature and denoising autoencoder-based dereverberation for distant-talking speaker identification. *EURASIP Journal on Audio, Speech, and Music Processing*, 2015(1), p.12.
19. ---
20. Zhang, C., Koishida, K. and Hansen, J.H., 2018. Text-independent speaker verification based on triplet convolutional neural network embeddings. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 26(9), pp.1633-1644.
21. Zeinali, H., Sameti, H. and Burget, L., 2017. Text-dependent speaker verification based on i-vectors, Neural Networks and Hidden Markov Models. *Computer Speech & Language*, 46, pp.53-71.
22. Sahidullah, M. and Kinnunen, T., 2016. Local spectral variability features for speaker verification. *Digital Signal Processing*, 50, pp.1-11.
23. Wang, J.C., Wang, C.Y., Chin, Y.H., Liu, Y.T., Chen, E.T. and Chang, P.C., 2017. Spectral-temporal receptive fields and MFCC balanced feature extraction for robust speaker recognition. *Multimedia Tools and Applications*, 76(3), pp.4055-4068.
24. Visalakshi, R., Dhanalakshmi, P. and Palanivel, S., 2016. Analysis of throat microphone using MFCC features for speaker recognition. In *Computational Intelligence, Cyber Security and Computational Models* (pp. 35-41). Springer, Singapore.
25. Chougule, S.V. and Chavan, M.S., 2015. Robust spectral features for automatic speaker recognition in mismatch condition. *Procedia Computer Science*, 58, pp.272-279.
26. Das, S., Biswas, A., Dasgupta, S. and Abraham, A., 2009. Bacterial foraging optimization algorithm: theoretical foundations, analysis, and applications. In *Foundations of Computational Intelligence Volume 3* (pp. 23-55). Springer, Berlin, Heidelberg.
27. Rani, R.R. and Ramyachitra, D., 2016. Multiple sequence alignment using multi-objective based bacterial foraging optimization algorithm. *Biosystems*, 150, pp.177-189.
28. Mavrovouniotis, M., Li, C. and Yang, S., 2017. A survey of swarm intelligence for dynamic optimization: Algorithms and applications. *Swarm and Evolutionary Computation*, 33, pp.1-17.



**Impact Factor:
7.569**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details