An Optimized and Privacy-Preserving System Architecture for Effective Voice Authentication over Wireless Network

Aniruddha Deka, Debashis Dev Misra

***Abstract:** The speaker authentication systems assist in determining the identity of speaker in audio through distinctive voice characteristics. Accurate speaker authentication over wireless network is becoming more challenging due to phishing assaults over the network. There have been constructed multiple kinds of speech authentication models to employ in multiple applications where voice authentication is a primary focus for user identity verification. However, explored voice authentication models have some limitations related to accuracy and phishing assaults in real-time over wireless network. In research, optimized and privacy-preserving system architecture for effective speaker authentication over a wireless network has been proposed to accurately identify the speaker voice in real-time and prevent phishing assaults over network in more accurate manner. The proposed system achieved very good performance metrics measured accuracy, precision, and recall and the F1 score of the proposed model were98.91%, 96.43%, 95.37%, and 97.99%, respectively. The measured training losses on the epoch 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 were 2.4, 2.1, 1.8, 1.5, 1.2, 0.9, 0.6, 0.3, 0.3, 0.3, and 0.2, respectively. Also, the measured testing losses on the epoch of 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 were 2.2, 2, 1.5, 1.4, 1.1, 0.8, 0.8, 0.7, 0.4, 0.1 and 0.1, respectively. Voice authentication over wireless networks is serious issue due to various phishing attacks and inaccuracy in voice identification. Therefore, this requires huge attention for further research in this field to develop less computationally complex speech authentication systems*

***Keyword:** CNN, LSTM, Speaker Authentication, Privacy-Preserving, Phishing Assaults, Wireless Network.*

I. INTRODUCTION

Voice recognition is a critical challenge, particularly in the scenario where high security is required to prevent information over wireless networks. The key issue encountered in voice identification is speaker pronunciation style, state of emotions, and word rate. There are certainly other problems in accurate speech recognition such as environmental noise which also make voice recognition very much difficult[1], [2].

In the modern era, smart phones are playing a significant role in the day-to-day lives of people globally. According to a report by the Gartner Institute, mobile gadgets sales have increased in recent years and most people all times carry smart phones with them and which has been recognized important part of life. Such computing devices i.e., smart phones or tablets help people to store private data, important images, or voice recordings along with confidential information, for instance, bank details or emails on personal smart phones[3]. Sometimes, people usually don't pay the required attention to the confidentiality of such confidential information. In smart phones or similar other computing devices namely tablets, user verification is the initial entry level for maintaining secrecy[4]. There have indeed been developed various applications which help in speech authentication to access smart phones. However, many of the applications are prone to security breaches owing to the various phishing assaults and because of that there are chances of information loss[5].

Speech authentication is a very significant means of user identity verification. The voice-based uses verification system is utilized in multiple fields where high security is required[6]. Some of the fields where speech authorization systems have been utilized include defence, robotics, banking, and many more. There have already been developed multiple techniques and models to save voice prints and process them are frequency estimation, pattern matching protocol, machine learning-based models, decision tree and vector quantization as well as convolution neural network (CNN), hidden Markov Models, matrix representation, etc. A few of the most common and key elements of any speech decoding are feature extraction algorithms and speech classification protocols[7], [8]. During previous research done on speech authentication techniques and state-of-the-art models, multiple speaker identification algorithms have been presented which offer varied degrees of performance metrics for safety purposes. Few of the traditional speech authentication systems have utilized GMM (Gaussian Mixture Model) rooted HMM (Hidden Markov Model)[9], [10]. Though, these existing conventional models and traditional algorithms have certain accuracy limitations and are less resistant to diverse phishing assaults in the modern wireless network environment. These days, owing to the development of Machine learning based techniques namely DNN (Deep Neural Network)[11], SVM (Support Vector Machine)[12], Linear Regression[13], KNN (K-Nearest Neighbours)[14], and multiple speech authentication models are built[15], [16], [17].

Manuscript received on 13 July 2023 | Revised Manuscript received on 18 July 2023 | Manuscript Accepted on 15 September 2023 | Manuscript published on 30 September 2023.
***Correspondence Author(s)**

Dr. Aniruddha Deka, Associate Professor, Department of Computer Science and Engineering, Assam Down Town University, Guwahati (Assam), India. Email: aniruddha.deka@adtu.in, dekaaniruddha@gmail.com, ORCID ID: 0000-0002-1228-232X

Dr. Debashis Dev Misra,* Associate Professor, Department of Computer Science and Engineering, Assam Down Town University, Guwahati (Assam), India. Email: debashis.misra@adtu.in, debashish.dm@gmail.com, ORCID ID: 0009-0002-9838-7241

On the contrary side, due to a large number of datasets and system and network up gradation, there is a key challenge to prevent phishing assaults completely as well as to improve the performance measures of previously developed models that is a key concern to focus more attention towards new framework development. Therefore, for effective speaker authentication, in this work, new privacy-preserving system architecture for user voice identification is presented. The key contribution of the present research work is described as follows.

- The key aim of this research is to develop optimized and privacy-preserving system architecture for effective speaker authentication over a wireless network.
- This proposed system architecture is a hybrid model which includes multiple algorithms ineffective feature extraction and classification for improving the performance metrics.
- The performance of the proposed system is measured both in the training and testing phase and it is to be found that all performance measures of the proposed system are very optimal and accurate.
- The measured accuracy, precision, recall, and F1 score of the proposed speech authentication system are 98.91%, 96.43%, 95.37%, and 97.99%, respectively.

II. LITERATURE REVIEW

In [18], Y. Kang *et al.* proposed a model to enhance the privacy of the user by improving the pre-processing technique. In this work, a new auto-encoder is developed for voice detection. However, this proposed model is insecure and offers large training losses. C. Z. Yang *et al.* in [19], proposed a scheme to prevent Deep Fake assaults on user authorization as means of Lip-based movement analytics. This research shows that a lip-rooted user authorization system may offer pragmatic verification performance. In this proposed approach, a deep convolutional neural network is used and comprises two operational segments i.e., lip-based feature extraction procedure as well as fundamental features retrieval network along with the categorization network. However, this proposed scheme is intricate for implementation in real-time speech authentication systems due to hefty computational complexity.

In [20], H. Park *et al.* presented a user identity verification system that is based on speech synthesis recognition. In the modern era, due to the widespread usage of internet facilities multiple social networking services have been developed and globally utilized by multifarious individuals. Though there are multiple means for accessing varying services over the internet, speech authentication is also one of the secure methods for user verification. As the proposed speech recognition system may verify the user's speech in real-time. However, this approach has less accuracy rate in modern wireless network applications also it has large computing complexity. In [21], K. Khadar Nawas*et al.* proposed another method for speaker identification by utilizing the Random Forest technique. In this work, the researcher developed a user speech recognition model based on the Random Forest technique which is used in the classification process. For the feature extraction procedure, another technique i.e., MFCC was adapted for the features extraction procedure. Based on, the proposed approach performance metrics it is identified that, this model's accuracy is lower as well as classification process takes massive time in huge datasets.

In [22],A. Mittal et al. presented an analysis of automated speaker identification models as well as phishing assaults removing approaches developed in the last decade. In this analysis, it is found that automated speech authentication systems have indeed been improved enough by researchers to accurately recognize the user's speech in multiple applications, as security is the primary concern in the modern era. Though, the susceptibility of such developed models to multifarious indirect as well as direct access assaults deteriorates authentication mechanism power in real-time which is a massive issue still present in developed systems and needs more demand for further research.

S. Debnath*et al.* in [23], proposed a multi-modal authorization system architecture rooted in the audio-visual datasets. In this work, the researcher considered audio-video user authorization rooted in multiple information sources. In this approach, audio identification has been included along with visual identification for security purposes. The user speech as well as face features were extracted individually from the combined audio-visual datasets and utilized the double modality for user verification in real-time. However, this proposed approach has some drawbacks related to the inaccuracy in classification and feature extraction procedure due to the use of conventional protocols and the less resistance against phishing assaults over wireless networks.

In [24],V. Gujral*et al.* proposed an advanced scheme for speech signal processing to authenticate the speaker in the communication system. The proposed user speech recognition method is rooted over phase as well as speech frequency and analysis is done according to the frame-by-frame approach. However, the proposed scheme has some flaws related to the synthetic speech, possibilities of phishing assaults owing to the conventional algorithms and more complexity in implementation in wireless network scenario, and many more.

W. Jiang *et al.* in [25], suggested another deep neural network-based rooted speech identification approach. The researcher developed this deep feature neural model for informative feature extraction as means of heterogeneous-based acoustic features cluster that comprises relevant as well as unrelated information or other redundant datasets which leads to lower emotional identification. In this work, the final classifier was employed as the SVM approach and the dataset was adopted by IEMOCAP. However, this suggested system can give an accuracy of 64% only which is very minimal and need to be improved.

In [26], R. Jage et al. discussed the speech coding techniques for effective transmission of the voice datasets over the wireless network. The speech is a common approach for communication among the various people. In the modern era, there are gigantic requirement of the speech translation because it carry huge valuable information.

However, this intensive work does not provide greater accuracy in user authorization as well as speech signal encoding over the wireless network.

III. METHODOLOGY

A. System Design

Speaker voice recognition is a very crucial job due to pronunciation style, speaking speed, and other parameters. Accurate authentication of the user's voice over a wireless network is very essential to eliminate the chances of phishing assaults and helps in the breach of information. Speech authentication systems are helpful in a variety of applications in the modern era such as remote access to wireless networks, internet transactions as well as IVR-rooted banking system, and many more. As accurate voice authentication is still a major challenging problem because of rapid development in communication infrastructure and errors in the deployment of new networking technology models provide room for phishing assaults. In this research, an optimized and privacy-preserving system architecture for effective speaker authentication over a wireless network is proposed. The key advantage of the proposed model is that it can be applied in multiple applications such as robotics, control system, and forensics for speaker recognition more accurately and detects phishing assaults in very less time.

Figure 1 depicts the proposed system architecture for speaker authorization. The working procedure of the proposed voice authentication system is described as follows. The voice datasets utilized in the proposed system training and testing were taken from multiple datasets i.e., ELSDSR and ASVspoof2021, and real-time generated datasets. The voice dataset input is first pre-processed for performing multiple steps such as noise removal using the median filter. Also, datasets standardization and clustering are performed in the pre-processing phase. The k-fold cross-validation method is opted to divide the entire datasets for training as well as a testing phase. The value of k for data separation was taken 5 and it may call a 5-fold cross-validation. The datasets for training and testing were divided into 70% and 30%, respectively for system performance evaluation. The feature extraction procedure was performed using the enhanced MFCC (Mel-frequency cepstral coefficients) and TQWT (Tunable Q wavelet transform) technique. The MFCC performs the extraction of features via speech signals to utilize them in the identification of specific tasks.

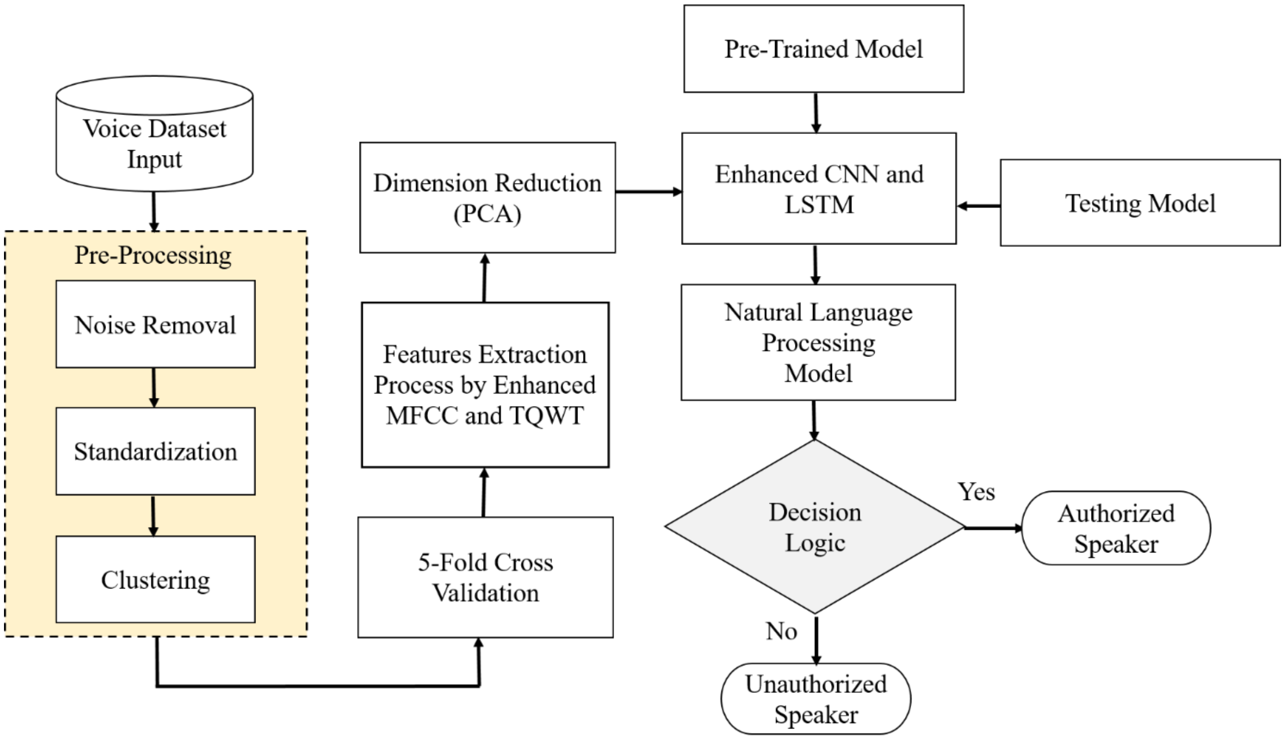

Figure 1: Depicts Proposed System Architecture for Speaker Authorization

TQWT is a method that generates wavelet multi-resolution analysis along with a user-defined Q-factor. The TQWT technique offers a means for accurate signal reconstruction and TQWT factors divide signal energy within multiple sub-bands. The dimension reduction process is performed by the principal component analysis (PCA). Before the classification and speech signal recognition procedure, in the proposed system architecture a pre-trained model i.e., a saved network employed for the training and validation procedure. The enhanced convolution neural network (CNN) and long short-term memory (LSTM) based model have opted for the classification procedure. Final decision logic is done based on the NLP model employed after the classification procedure to recognize the authorized speaker and unauthorized speaker in real-time. The NLP model is employed for accurate verification of the authorized users and prevents any phishing assaults; thereby the proposed system architecture is highly recommended for multiple wireless network applications for speaker authorization in real-time.

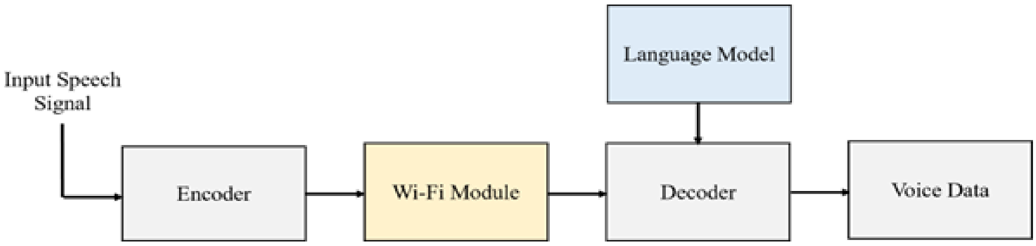

Figure 2: Illustrates the Model for Speech Signal Encoding and Decoding Over Wireless Network

Figure 2 illustrates the model for speech signal encoding and decoding over wireless network. The speech encoder and decoder employed in the suggested model for speech signal transmission over wireless network is based on CELP (Code-Excited Linear Prediction) technique. Speech data encoding methods may support massive amount of the datasets as well as augment information volume that is to be translated over the wireless network. The speech encoding method illustrates the datasets with minimal amount of bits as well as preserves quality of real-time speech. This encoder translates the digitized datasets into codes as well as these converted codes are translated over the wireless network as frames. In this model, for testing of the suggested model a Wi-Fi (Wireless Fidelity) module is employed. On the receiving end a decoder is employed along with a language module which is selected as N-Gram Model. The decoder obtains the encoded frames from the encoder via Wi-Fi for performing the synthesis procedure for real signal reconstruction. The N-Gram Model used with the decoded would offer manifold advantages such as higher speech quality, minimal need of storage, less computationally complex architecture as well as lower channel errors as well as coding latency.

B. Datasets Used

The proposed system training and testing were performed as means of multiple datasets i.e. ELSDSR (English language speech database for speaker recognition)[1], ASVspoof2021[27], and real-time generated dataset for accurate voice authentication during the experimental work. The purpose to select multiple datasets was to test and record the results more accurately.

C. System Configuration

This experimental work has been performed with the help of a personal computing machine which is inbuilt with mentioned system assortment: Intel i7 processor with RAM of 16 GB, Windows 11, and Graphic card RTX 3050 Ti with the operating system of 64-bit. The system training and testing were performed as means of a cloud-based application service namely Google Colab.

D. Data Collection

In this work, the authors used voice samples of both males and females of an average age between 18 to 45 years. From ELSDSR datasets male and female voice samples were taken 20 and 15, respectively. From ASVspoof2021 voice samples of both males and females were taken 22 and 17, respectively. Also, the real-time datasets were adapted and the voice samples for both males and females were 10 and 7, respectively. Figure 3 illustrates the datasets used in the experiment. The entire dataset was divided into training and testing groups i.e., 70% and 30%, respectively, using the 5-fold cross-validation method.

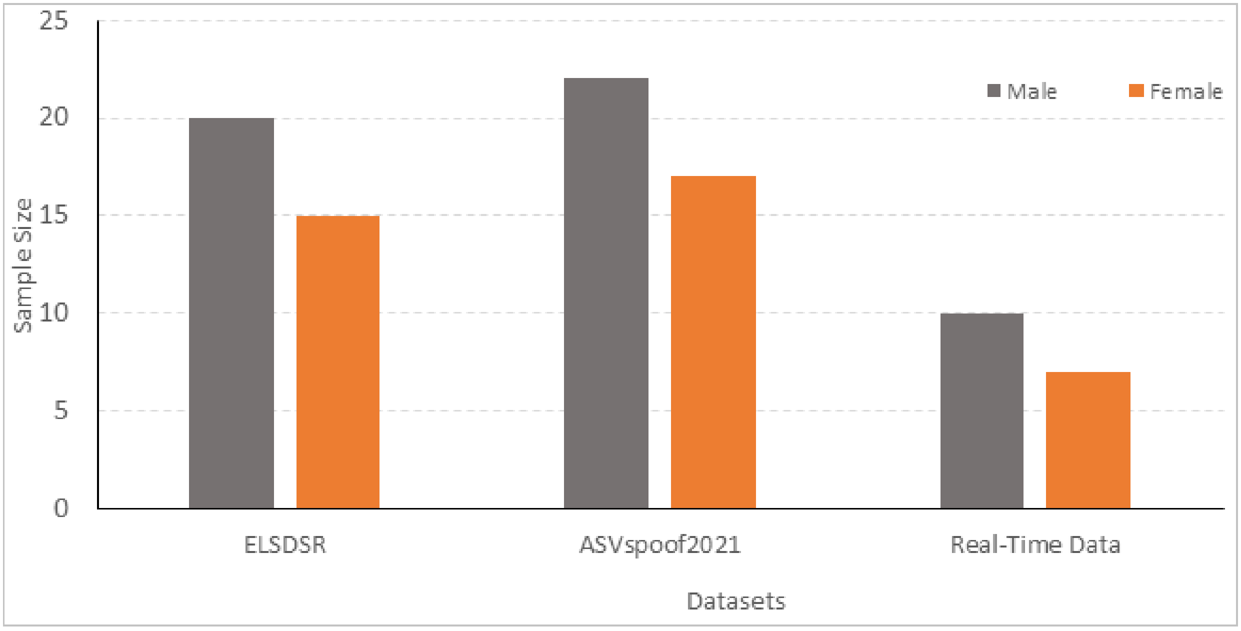

Figure 3: Illustrates the Datasets used in the Experiment

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

E. Performance Evaluation Metrics

This proposed system architecture which is developed for effective speaker authentication over wireless networks was trained and tested by using diverse datasets. The proposed model performance was measured and evaluated as means of multiple metrics such as accuracy, precision, F1 score, and recall as described in this subsection and compared with the existing models for determining the effectiveness of this proposed system architecture.

E. a. Accuracy.

The accuracy metrics depict the ratio of the number of accurately categorized dataset instances over the overall number of dataset instances. The accuracy metrics are described as shown in equation 1.

$$\text{Accuracy} = \frac{\text{TN+TP}}{\text{TN+FP+TP+FN}} \tag{1}$$

E. b. Precision.

The precision metrics may be described as the ratio of the accurately categorized positive sample over overall categorized positive samples. The precision metrics may be described as depicted in equation 2.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

E. c. Recall.

The recall performance metrics may be evaluated as the ratio of accurately classified positive samples over the overall number of positive samples. The recall parameter determines the system's capability for determining the positive samples. The recall performance metrics may be described as shown in equation 3.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

E. d. F1 Score.

The F1 score metrics are utilized for binary categorization model evaluation based on the predictions which may be prepared for positive classes. The F1 score can be evaluated as means of Precision as well as accurate recall values. Thus, the F1 score may be evaluated as precision and recall of both parameters' harmonic mean and assigns equal weight to precision and recall.

$$F1\ score = 2 \times \frac{Precison \times Recall}{Precison + Recall} \tag{4}$$

IV. RESULTS AND DISCUSSION

The speech identification system architecture enables computing units for processing audio signals in textual form or may perform the specified jobs. Advanced build speech authentication system architecture would enable changing the way to establish communication means with the computing units. There are multiple applications where speech authentication systems are playing a very significant role in verifying the users for providing access to specific applications immediately. A few of the most common application areas where speech authentication systems are employed include defense, the telecom industry, banking, and other applications. People's voice is indeed very identical and unique similar to the fingerprint of any individual. Thus, the voice of an individual may be a very important and practical tool for instance to control the accessibility of various devices or applications. Another applicability would be the capability for translating the computing speech in an online scenario. The developed speech authentication systems are still providing limited performance in some aspects and have higher computing costs as well as intricacy in implementation in real-time. To consider such issues about the restricted performance metrics, in this work a novel optimized and privacy-preserving system architecture for effective speaker authentication was developed for improving feature extraction and classification purpose for more accurate speech authentication purpose. This research was conducted using a personal computer equipped with the following hardware components: an Intel i7 CPU, 16 GB of RAM, Windows 11, and an RTX 3050 Ti graphics card running a 64-bit operating system. The system testing phase, as well as the training phase, was carried out using Google Colab, a cloud-rooted application service.

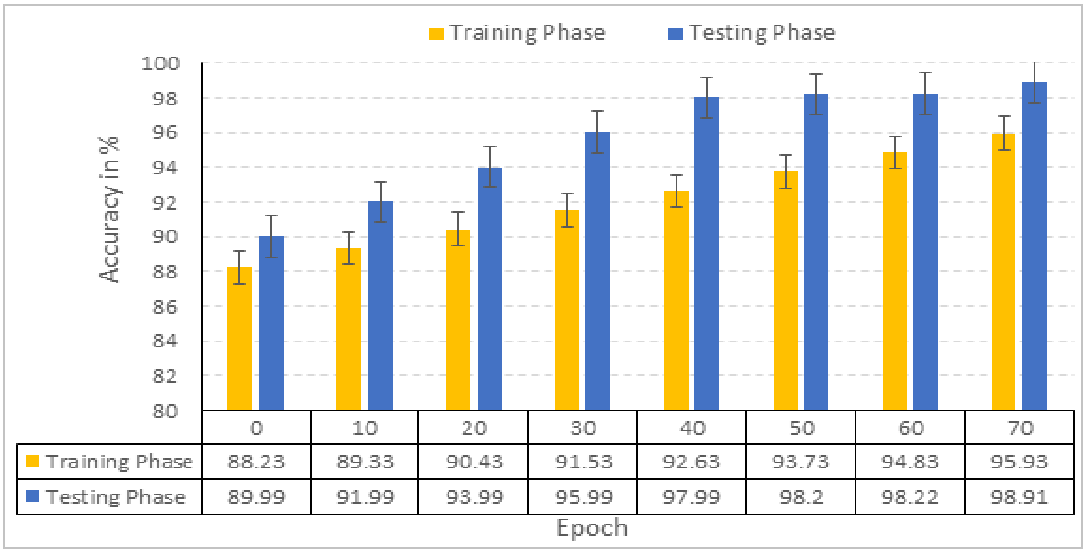

Figure 4: Measured Accuracy of The Proposed System in The Training and Testing Phase

Figure 4 depicts the measured accuracy of the proposed system in the training and testing phase. During the training and testing phase of this proposed system for multiple-speaker voice authentication, the accuracy is recorded on the different epochs. The accuracy in the training phase on epochs 1, 10, 20, 30, 0, 50, 60, and 70 was measured at 88.23%, 89.33%, 90.43%, 91.53%, 92.63%, 93.73%, 94.83%, and 95.93%, respectively. The accuracy in the testing phase on epochs 1, 10, 20, 30, 0, 50, 60, and 70 was measured at 89.99%, 91.99%, 93.99%, 95.99%, 97.99%, 98.2%, 98.22 and 98.91%, respectively.

Table 1: Performance Metrics Comparison of the Proposed System with Existing Systems

Sl.	Systems	Accuracy (%)	Precision	Recall	F1 Score
1	Proposed System	98.91%	96.43%	95.37%	97.99%
2	Raghad et al. [28]	97.83%	×	×	×
3	Z. Hao et al. [29]	93.40%	×	×	×
4	A. Tahseen Ali et al. [30]	97.90%	90.00%	91.00%	90.00%

Table 1 shows the performance metrics comparison of the proposed system with existing systems. The previously developed speech recognition system by Raghad et al. [28]obtained an accuracy, of 97.83%. However, the other performance parameters namely F1 score, recall, and precision are not disclosed in the paper. Another existing paper, presented by Z. Hao et al. [29], disclosed the measured accuracy value of 93.40%, however other performance metrics such as precision, recall, and F1 score are not disclosed. Furthermore, in a paper presented by A. Tahseen Ali et al. [30], the accuracy, precision, recall, and F1 score were measured at 97.9%, 90.00%, 91.00%, and 90.00% respectively. However, in the proposed speech recognition system, the accuracy value, precision value, recall value, and F1 score value, were measure dat 98.91%, 96.43%, 95.37%, and 97.99%, respectively. The proposed system obtained very accurate and improved performance metrics in comparison to the existing approaches.

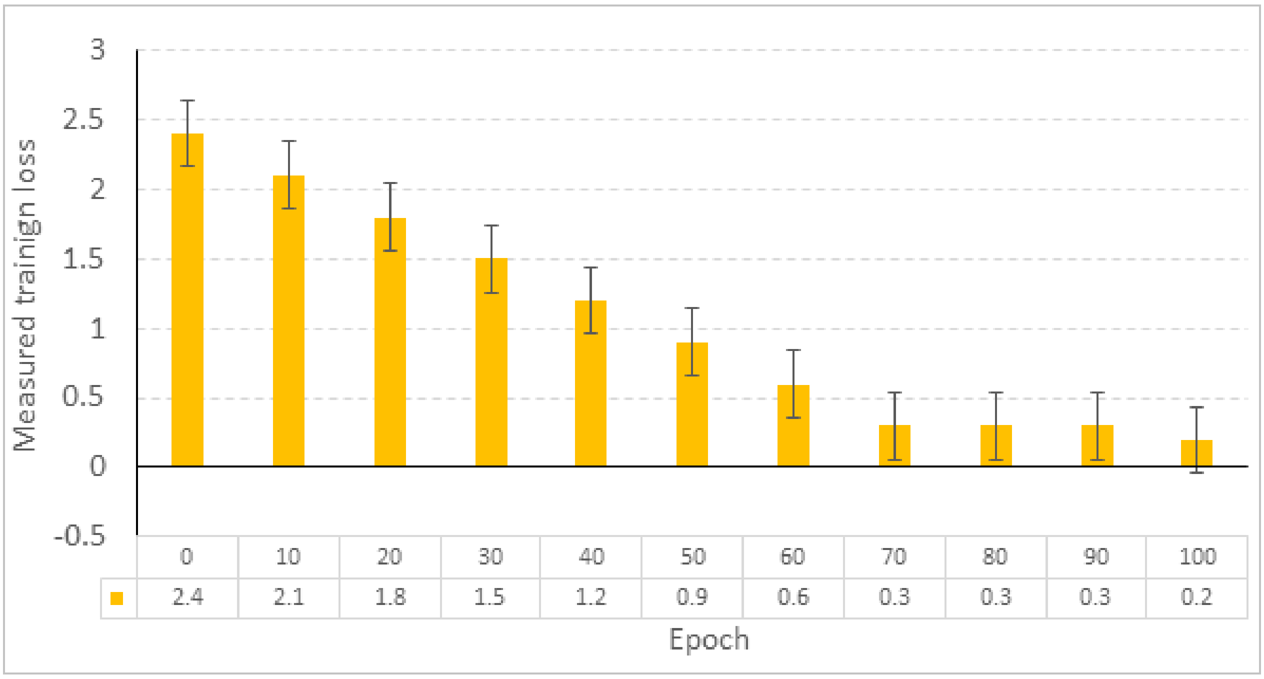

Figure 5: Measured Training Loss of the Proposed System

Figure 5 shows the measured training loss of the proposed system. The measured training losses of the proposed system on epoch 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100, was 2.4, 2.1, 1.8, 1.5, 1.2, 0.9, 0.6, 0.3, 0.3, 0.3 and 0.2, respectively. The reduced training loss depicts that the proposed user speech authentication system losses are effectively decreased on the increasing epoch in the real-time training phase of the system.

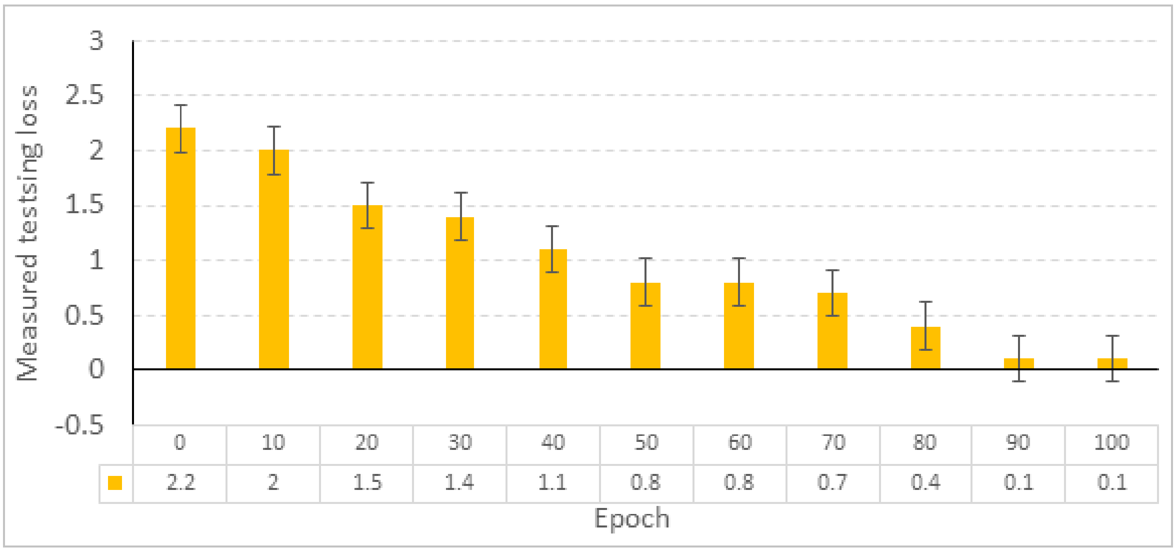

Figure 6: Measured Testing Loss of the Proposed System

Figure 6 shows the measured testing loss of the proposed system. The measured testing losses of the proposed system on epoch 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100, was 2.2, 2, 1.5, 1.4, 1.1, 0.8, 0.8, 0.7, 0.4, 0.1 and 0.1, respectively. The reduced testing loss depicts that the proposed user speech authentication system losses are decreased on the increasing epoch in the real-time testing phase of the system.

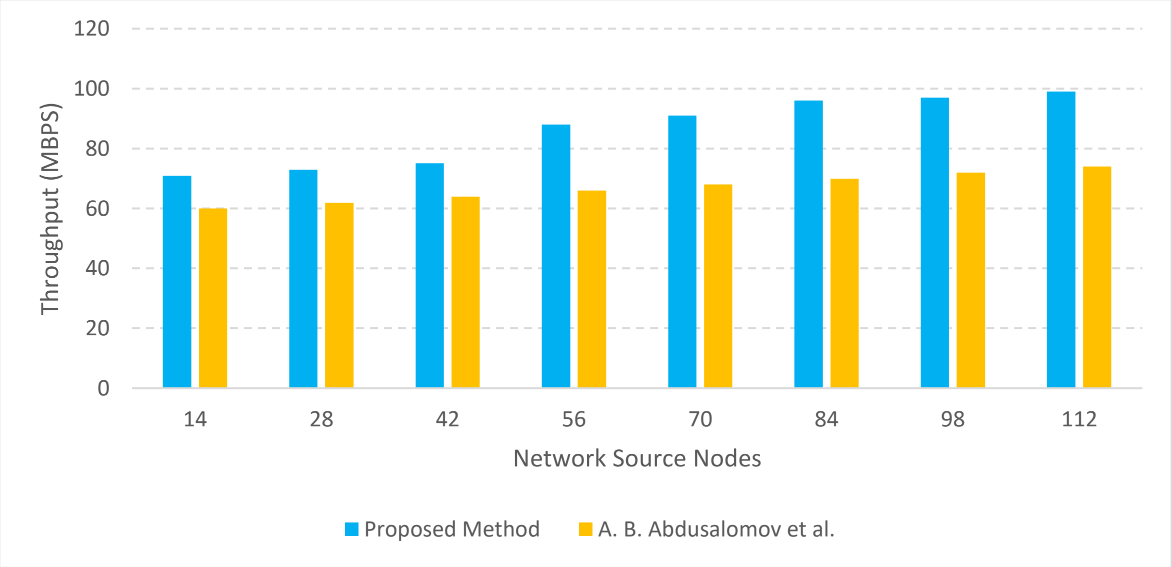

Figure 7: Illustrates the Proposed Model Measured Throughput Over Wireless Network

Figure 7 illustrates the proposed model measured throughput over wireless network. The speech is indeed a distinct kind signal that is very complex and challenging for effective analysis as well as model. In this work encoder and decoder is employed based upon the CELP algorithm to perform the secure communication over the wireless network for increasing the overall network throughput. The measured network throughput of proposed model over the network nodes 14, 28, 42, 56, 70, 84, 98 and 112 is 80, 73, 75, 88, 91, 96, 97, 99. However, the A. B. Abdusalomov et al. [31] model throughput over the network nodes 14, 28, 42, 56, 70, 84, 98 and 112 is 60, 62, 64, 66, 68, 70, 72, and 74 respectively. Hence, the evaluated outcome of the proposed model is improved very significantly in comparison to the previous approach.

V. CONCLUSION

Speech recognition in an accurate manner is a very intricate and challenging problem. The usage of wireless networks for communication purposes has increased in recent years due to rapid wireless network infrastructure development globally. Earlier, many speech recognition approaches had already been developed for accurately performing classification and feature extraction procedures for identifying authorized or unauthorized users in numerous application scenarios in real-time. Though, such a developed scheme has some flaws related to accuracy as well as other low-performance metrics. In this work, a novel optimized and privacy-preserving system architecture for effective speaker authentication over the wireless network has been proposed.

The proposed speech recognition system employed MFCC as well as the TQWT protocol for the informative feature extraction procedure. The classification process is done as means of LSTM and enhanced CNN-based classifier for improved accuracy. The proposed speech authentication system offers optimal performance metrics i.e., accuracy, precision, recall, and F1 score were measured at 98.91%, 96.43%, 95.37%, and 97.99%, respectively. Also, the training, as well as testing validation losses, were measured very optimally. The measured training losses on the epoch of 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 were 2.4, 2.1, 1.8, 1.5, 1.2, 0.9, 0.6, 0.3, 0.3, 0.3, and 0.2, respectively. Likewise, the measured testing losses on the epoch of 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 were 2.2, 2, 1.5, 1.4, 1.1, 0.8, 0.8, 0.7, 0.4, 0.1 and 0.1, respectively.

ACKNOWLEDGMENT

Authors acknowledge the immense help received from the scholars whose articles are cited and included in references to this manuscript. The authors are also grateful to authors/ editors / publishers of all those articles, journals and books from where the literature for this article has been reviewed and discussed.

DECLARATION

Funding/ Grants/ Financial Support	No, I did not receive.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material/ Data Access Statement	Not relevant.
Authors Contributions	All authors have equal participated in this article.

REFERENCES

1. P. Dhakal, P. Damacharla, A. Y. Javaid, and V. Devabhaktuni, “A Near Real-Time Automatic Speaker Recognition Architecture for Voice-Based User Interface,” Mach. Learn. Knowl. Extr., vol. 1, no. 1, pp. 504–520, 2019, doi: 10.3390/make1010031. [CrossRef]
2. A. V. Amrutha, K. H. Anagha, A. Kamal K, and B. Kumaraswamy, “Multi-level Speaker Authentication: An Overview and Implementation,” in 2020 IEEE 17th India Council International Conference, INDICON 2020, 2020. doi: 10.1109/INDICON49873.2020.9342423. [CrossRef]
3. S. Abhishek Anand, J. Liu, C. Wang, M. Shirvanian, N. Saxena, and Y. Chen, “EchoVib: Exploring Voice Authentication via Unique Non-Linear Vibrations of Short Replayed Speech,” in ASIA CCS 2021 - Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security, 2021. doi: 10.1145/3433210.3437518. [CrossRef]
4. B. Chettri, “Voice Biometric System Security: Design and Analysis of Countermeasures for Replay Attacks,” 2020. [CrossRef]
5. N. Kobayashi and T. Morooka, “Application of High-accuracy Silent Speech BCI to Biometrics using Deep Learning,” in 9th IEEE International Winter Conference on Brain-Computer Interface, BCI 2021, 2021. doi: 10.1109/BCI51272.2021.9385338. [CrossRef]
6. S. Kinkiri, W. J. C. Melis, and S. Keates, “Machine learning for voice recognition,” Second Medw. Eng. Conf. Syst. Effic. Sustain. Model., 2017.
7. L. Chowdhury, M. Kamal, N. Hasan, and N. Mohammed, “Curricular SincNet: Towards Robust Deep Speaker Recognition by Emphasizing Hard Samples in Latent Space,” in BIOSIG 2021 - Proceedings of the 20th International Conference of the Biometrics Special Interest Group, 2021. doi: 10.1109/BIOSIG52210.2021.9548296. [CrossRef]
8. R. Jahangir et al., “Text-Independent Speaker Identification through Feature Fusion and Deep Neural Network,” IEEE Access, 2020, doi: 10.1109/ACCESS.2020.2973541. [CrossRef]
9. S. Duraibi, W. Alhamdani, and F. T. Sheldon, “Voice Feature Learning using Convolutional Neural Networks Designed to Avoid Replay Attacks,” in 2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020, 2020. doi: 10.1109/SSCI47803.2020.9308489. [CrossRef]
10. D. R. KS, R. MD, and S. G, “Comparative performance analysis for speech digit recognition based on MFCC and vector quantization,” Glob. Transitions Proc., 2021, doi: 10.1016/j.gltp.2021.08.013. [CrossRef]
11. O. Mamyrbayev, A. Akhmediyarova, A. Kydyrbekova, N. O. Mekebayev, and B. Zhumazhanov, “BIOMETRIC HUMAN AUTHENTICATION SYSTEM THROUGH SPEECH USING DEEP NEURAL NETWORKS (DNN),” Bull., 2020, doi: 10.32014/2020.2518-1467.137. [CrossRef]
12. M. Dua, C. Jain, and S. Kumar, “LSTM and CNN based ensemble approach for spoof detection task in automatic speaker verification systems,” J. Ambient Intell. Humaniz. Comput., 2022, doi: 10.1007/s12652-021-02960-0. [CrossRef]
13. S. Bunrit, T. Inkian, N. Kerdprasop, and K. Kerdprasop, “Text-independent speaker identification using deep learning model of convolution neural network,” Int. J. Mach. Learn. Comput., 2019, doi: 10.18178/ijmlc.2019.9.2.778. [CrossRef]
14. K. Aizat, O. Mohamed, M. Orken, A. Ainur, and B. Zhumazhanov, “Identification and authentication of user voice using DNN features and i-vector,” Cogent Eng., 2020, doi: 10.1080/23311916.2020.1751557. [CrossRef]
15. Q. Wang, P. Guo, and L. Xie, “Inaudible adversarial perturbations for targeted attack in speaker recognition,” in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2020. doi: 10.21437/Interspeech.2020-1955. [CrossRef]
16. S. Nasr, M. Quwaider, and R. Qureshi, “Text-independent Speaker Recognition using Deep Neural Networks,” in 2021 International Conference on Information Technology, ICIT 2021 - Proceedings, 2021. doi: 10.1109/ICIT52682.2021.9491705. [CrossRef]
17. G. HimaBindu, G. Lakshmeeswari, G. Lalitha, and P. P. S. Subhashini, “Recognition using DNN with bacterial foraging optimization using MFCC coefficients,” J. Eur. des Syst. Autom., 2021, doi: 10.18280/JESA.540210. [CrossRef]
18. Y. Kang, W. Kim, S. Lim, H. Kim, and H. Seo, “DeepDetection: Privacy-Enhanced Deep Voice Detection and User Authentication for Preventing Voice Phishing,” Appl. Sci., vol. 12, no. 21, p. 11109, 2022, doi: 10.3390/app122111109. [CrossRef]
19. C. Z. Yang, J. Ma, S. Wang, and A. W. C. Liew, “Preventing DeepFake Attacks on Speaker Authentication by Dynamic Lip Movement Analysis,” IEEE Trans. Inf. Forensics Secur., 2021, doi: 10.1109/TIFS.2020.3045937. [CrossRef]
20. H. Park and T. Kim, “User Authentication Method via Speaker Recognition and Speech Synthesis Detection,” Secur. Commun. Networks, 2022, doi: 10.1155/2022/5755785. [CrossRef]
21. K. Khadar Nawas, M. Kumar Barik, and A. Nayeemulla Khan, “Speaker Recognition using Random Forest,” ITM Web Conf., 2021, doi: 10.1051/itmconf/20213701022. [CrossRef]
22. A. Mittal and M. Dua, “Automatic speaker verification systems and spoof detection techniques: review and analysis,” Int. J. Speech Technol., 2022, doi: 10.1007/s10772-021-09876-2. [CrossRef]
23. S. Debnath and P. Roy, “Multi-modal authentication system based on audio-visual data,” in IEEE Region 10 Annual International Conference, Proceedings/TENCON, 2019. doi: 10.1109/TENCON.2019.8929592. [CrossRef]
24. V. Gujral, J. Joshi, P. Medikonda, and N. Grover, “Advanced Speech Processing for Speaker Authentication in Communication Systems,” in International Symposium on Advanced Networks and Telecommunication Systems, ANTS, 2018. doi: 10.1109/ANTS.2018.8710076. [CrossRef]

25. W. Jiang, Z. Wang, J. S. Jin, X. Han, and C. Li, "Speech Emotion Recognition with Heterogeneous," pp. 1–15, 2019, doi: 10.3390/s19122730. [CrossRef]
26. R. Jage and S. Upadhya, "CELP and MELP speech coding techniques," Proc. 2016 IEEE Int. Conf. Wirel. Commun. Signal Process. Networking, WiSPNET 2016, pp. 1398–1402, 2016, doi: 10.1109/WiSPNET.2016.7566366. [CrossRef]
27. Z. Wu et al., "ASVspoof: The Automatic Speaker Verification Spoofing and Countermeasures Challenge," IEEE J. Sel. Top. Signal Process., vol. PP, p. 1, 2017, doi: 10.1109/JSTSP.2017.2671435. [CrossRef]
28. R. T. Al-Hassani, D. C. Atilla, and Ç. Aydin, "Development of High Accuracy Classifier for the Speaker Recognition System," Appl. Bionics Biomech., 2021, doi: 10.1155/2021/5559616. [CrossRef]
29. Z. Hao, J. Peng, X. Dang, H. Yan, and R. Wang, "mmSafe: A Voice Security Verification System Based on Millimeter-Wave Radar," Sensors, vol. 22, no. 23, 2022, doi: 10.3390/s22239309. [CrossRef]
30. A. Tahseen Ali, H. S. Abdullah, and M. N. Fadhil, "WITHDRAWN: Voice recognition system using machine learning techniques," Mater. Today Proc., 2021, doi: https://doi.org/10.1016/j.matpr.2021.04.075. [CrossRef]
31. A. B. Abdusalomov, F. Safarov, M. Rakhimov, B. Turaev, and T. K. Whangbo, "Improved Feature Parameter Extraction from Speech Signals Using Machine Learning Algorithm," Sensors, vol. 22, no. 21, p. 8122, 2022, doi: 10.3390/s22218122. [CrossRef]

AUTHORS PROFILE

Dr. Aniruddha Deka is an expert in the field of computer science with 17 years of experience in academic. He is currently working as Associate professor in the Department of Computer Science and Engineering, Assam down town University Guwahati. He obtained B.E. degree in Computer Science and Engineering (2006) from North Eastern Hill University, Shillong, Meghalaya and M.Tech. in Information Technology (2013) from Gauhati University, Guwahati, Assam. He completed his PhD in Computer Science from Bodoland University, Assam. He worked in several research projects as an Assistant Project Engineer at IIT Guwahati from the year 2007 to 2012 in the area of signal processing, computer network, IVR mainly focusing on development of speech-based application in Assamese and Bodo language. He has vast experience in curriculum development and university administration. He served Assam Royal Global University, Guwahati for a period of 10 years as Assistant professor and HOD handling various courses like B.Tech, M.Tech, BCA, MCA, BSc.IT, and MSc.IT. He guided more than 30 students in UG and PG level and published more than 25 research papers in reputed international, national journals, conferences. He acts as a reviewer in various international journals. His area of expertise includes Speech Processing, Computer Network, Operating System, and DBMS.

Dr. Debashis Dev Misra is an expert in the field of Computer Networks and Wireless Communication. He has over 15 years of experience in academia and industry. He has contributed to the development of cutting-edge wireless routing protocols and technologies in the field of wireless communication. He is currently Associate Professor in Department of Computer Science and Engineering, Assam down town University, Guwahati, Assam, India. He obtained B.E. degree in Computer Science and Engineering (2004) from Dr. Ambedkar Institute of Technology, Visvesvaraya Technological University (VTU), Karnataka and M.Tech degree in Computer Science and Engineering (2011) from Ragiv Gandhi University, Itanagar, AP. He completed Ph.D. in Computer Science and Engineering from Assam Science and Technology University, Guwahati, Assam. He has also worked for a period of more than 4 years as a senor software developer in the IT industry with a reputed MNC based in Bangalore. Dr. Misra's research in the field of Computer Networks and Wireless Communication has led to notable progress in wireless routing algorithms and bio inspired algorithms in optimization problems. He has published research papers in prestigious journals and international conferences.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)/ journal and/or the editor(s). The Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP) and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Retrieval Number: 100.1/ijrte.C78620912323
DOI: 10.35940/ijrte.C7862.0912323
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)