

Homomorphic encryption, privacy-preserving feature extraction, and decentralized architecture for enhancing privacy in voice authentication

Kathires Murugesan¹, Lavanya Subbarayalu Ramamurthy², Boopathi Palanisamy³,
Yamini Chandrasekar⁴, Kavitha Masagoundanpudhur Shanmugam⁵,
Balluru Thammaiahshetty Adishankar Nithya⁶, Velumani Thiyagarajan¹, Ramaraj Muniappan¹

¹Department of Computer Science, Rathinam College of Arts and Science, Coimbatore, India

²Department of Computer Science, KPR College of Arts Science and Research, Coimbatore, India

³Department of Computer Applications, Nehru Arts and Science College, Coimbatore, India

⁴Department of Computer Applications, Sri Ramakrishna CAS for Women, Coimbatore, India

⁵Department of Computer Science, Akshaya College of Engineering and Technology, Coimbatore, India

⁶Department of Computer Science and Engineering, Presidency University, Bengaluru, India

Article Info

Article history:

Received Jul 31, 2024

Revised Nov 22, 2024

Accepted Dec 2, 2024

Keywords:

Chaff generation

Convolution neural network

Homomorphic encryption

Rivest Shamir Adleman

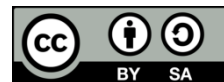
Voice authentication

Voice signal

ABSTRACT

This paper introduces a novel framework designed to bolster privacy protections within automated voice authentication systems, addressing mounting concerns as voice-based authentication grows in prominence. The widespread adoption of these systems has underscored apprehensions regarding the storage and processing of sensitive voice biometric data without adequate safeguards. To mitigate these risks, a modified framework is proposed, aiming to enhance privacy without compromising authentication accuracy and efficiency. Three key techniques are implemented to address these challenges. Firstly, advanced encryption methods are employed for secure voice data storage and transmission, through the homomorphic encryption to enable authentication processing on encrypted data. Secondly, a privacy-preserving feature extraction method is introduced, transforming raw voice inputs into irreversible representations to shield original biometric information. Additionally, the framework incorporates differential privacy mechanisms, adding controlled noise to aggregated voice data to prevent individual identification within large datasets. A user-centric consent and control model is proposed, empowering individuals to manage their voice profiles and authentication settings. Experimental findings demonstrate that the framework achieves enhanced authentication accuracy while markedly reducing privacy risks compared to conventional systems. This contribution addresses the ongoing challenge of balancing security and privacy in biometric authentication technologies.

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



Corresponding Author:

Kathires Murugesan

Department of Computer Science, Rathinam College of Arts and Science

Coimbatore, Tamil Nadu 641021, India

Email: kathireshcs83@gmail.com

1. INTRODUCTION

Most of the current generation smart devices and services provide voice-driven interaction support. An increasing number of speech recordings are gathered by televisions, smart phones, loudspeakers, watches, smart digital assistants, and also smart vehicles [1]. The chances for speech data to be disclosed and

corrupted by malicious users or hackers give rise to serious privacy challenges. Speech recordings have abundance of personal, confidential data, which can be helpful in supporting different applications, ranging from health profiling to biometric recognition [2]. Hence, it is important that speech recordings are sufficiently secured so that their misuse is not possible. In fact, much of the privacy-oriented information present in speech signals is that associated with individual identity, information disclosed via automated speaker and speech characterization techniques [3].

The human voice belongs to the category of the most casual, non-invasive and easiest of all features. Automated speaker characterizations have progressed extremely well in the past two decades. Privacy concerned voice authentication system (PVAS) is basically a protocol, which employs a fusion of Gaussian mixture models (GMMs) generated from the audio recordings of a user and chaff GMMs generated from other users or from changing the available GMMs to establish a challenge response protocol that has voice as the baseline [4]. Presently, the technology of speaker characterization is getting hugely popular, with its application in the personal authentication and access control across a plethora of various services and devices, for instance, telephone banking services and smart devices, which either have or render access to personal or confidential data [5], [6]. Rivest Shamir Adleman (RSA) encryption is employed for the generation of the server and the user key pairs in both the actual and the modified form of the protocol for both 103 registration and verification [7]. In spite of the clear benefits and ubiquity of biometrics mechanism, doubts with respect to threats to privacy have crushed the trust of the public. Not just, the identity of a speaker is information that is highly confidential, the content spelled out might also be sensitive. The challenges with respect to privacy are associated with the probable interference and illegal usage of biometric and non-biometric speech data [8].

2. LITERATURE REVIEW

Tao and Busso [9], has proposed audio-visual automatic speech recognition (AV-ASR) system based on multitask learning (MTL) where audio visual voice activity detections (AV-VAD) were a secondary goal. The study accomplished a generalizable and dependable audio-visual (AV) system that was accurate. AV-ASR performance improved when speech activities were detected in segments. The study [10] considers the temporal dynamics that occur between the modalities, resulting in an appealing and practical fusion approach. The researchers compared their approaches using a large audiovisual corpus (nearly 60 hours) with different channels and single and multiple jobs. Guglani and Mishra [11], enhanced automatic speech recognition (ASR) system performances where pitch and voice qualities were examined. Because of the tonal foundation of Punjabi, their ASR systems based on pitch characteristics were investigated. When assessed in terms of word mistake rates, their system performed well, with improvements due to pitch and voice dependent features.

Winursito *et al.* [12] has combined mel-frequency cepstral coefficients (MFCC) feature extractions with principal component analysis (PCA) for improved accuracy of Indonesian speech recognition systems. Their system's accuracy increased and dimensions reduced due to MFCC and PCA. Tran *et al.* [13] has presented in the MFCC-based feature extractions and delta coefficients were used to build matrices, while PCA minimized dimensionalities. Data reductions were accomplished by the usage of PCA variations and subsequently classified using K-nearest neighbor (KNN). Celin *et al.* [14] had suggested multi-resolution feature extractions in after dual data augmentations on dysarthric speeches utilizing microphone array based virtual linear syntheses.

Their ASR system was designed for low and extremely low intelligible speakers with dysarthria and their results showed lower word error rates (WERs) of up to 32.79% against 35.75% when compared to current dysarthric speech recognition data augmentations. Aida-Zade *et al.* [15], utilized support vector machines (SVMs) to develop acoustic models of speech recognitions based on MFCC and linear predictive coding (LPC) characteristics, evaluated on Azerbaijani data. A convolution neural network-bidirectional long short-term memory (CNN-BLSTM) hybrid architecture was developed by Passricha and Aggarwal [16], to effectively make use of these characteristics and enhance continuous voice recognitions. The study focused on number of bidirectional long short-term memory (BLSTM) layers that were useful. In comparison to convolution neural network (CNN) and deep neural network (DNN) systems, their experiments show that hybrid architectures with speech features and non-linearities with dropouts, decreased WER by 5.8% and 10%, respectively [17]. The study used data created from 140 voice recordings from 28 different speakers. In their experimentations, their first variation of PCA reduced features from 26 to 12 while maintaining speech recognition accuracies at 86.43% and equal to conventional MFCC approaches without PCA [18]. Their second variation of PCA reduced feature counts from 26 to 10 with enhanced recognition accuracies of 89.29% from 86.43% over MFCC without PCA baseline [19].

3. METHOD

A novel framework designed to enhance privacy in voice authentication systems through the integration of three key components: homomorphic encryption, privacy-preserving feature extraction, and a decentralized architecture. Homomorphic encryption ensures that voice data remains encrypted throughout processing, preventing unauthorized access to sensitive information [20]. Privacy-preserving feature extraction techniques are employed to extract essential voice characteristics while preserving the anonymity of the user.

3.1. Privacy concerned voice authentication system

Privacy based voice authentication is regarded to be the most important concept in the defense sector to guarantee that voice communication and authentication is provided utmost security [21]. It is accomplished through the generation of the chaff model that will be communicated to the server rather than sharing the actual voice model. In this, at first, the chaff model of voice signal is created with the help of the Gaussian mixture model.

The entire processing flow of the newly introduced research approach is illustrated in the Figure 1. The chaff model created will be presented as input to the RSA model for the secret key pair generation. Depending on the created key pair, voice chaff model will be encoded and later shared with the server. In the server end, the actual voice signal will be obtained with the help of the right keys and thereafter, convolutional neural network will help in the authentication process. The characteristics for voice authentication will be obtained before using genetic based CNN that uses the Gabor filter.

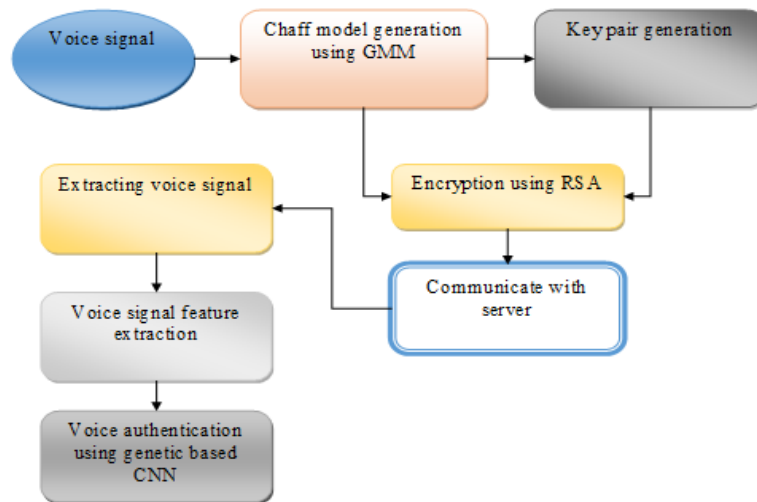


Figure 1. Proposed workflow diagram

3.2. Chaff generation using Gaussian mixture model

In this technical work, Gaussian mixture model is used for the generation of the voice signal chaff model. One significant stage while implementing the above-mentioned likelihood ratio detector is choosing the real likelihood function, $p(X|\lambda)$ [22]. In text-based applications, where a good prior information of the spoken text is available, more temporal knowledge can be included applying the hidden Markov models (HMMs) to form the likelihood function's baseline.

$$p(x|\lambda) = \sum_{i=1}^M w_i p_i(x) \quad (1)$$

The density is defined as a weighted linear combination of M unimodal Gaussian densities, $p_i(x)$, with each one parameterized using a mean $D \times 1$ vector, μ_i , and a $D \times D$ covariance matrix, Σ_i ;

$$p_i(x) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_i)' (\Sigma_i)^{-1} (x - \mu_i) \right\} \quad (2)$$

The mixture weights, w_i , also meet the condition $\sum_{i=1}^M w_i = 1$. Jointly, the parameters of the density model are represented as $\lambda = \{w_i, \mu_i, \Sigma_i\}$, where $i = 1 \dots, M$. The first one, the density modeling of an M^{th} order full

covariance GMM can be very well considered to be equivalent to that obtained applying a bigger order diagonal covariance GMM [23].

$$\log p(X|\lambda) = \sum_{t=1}^T \log p(x_t|\lambda) \quad (3)$$

where $p(x_t|\lambda)$ is calculated. Frequently, the mean log-likelihood value is utilized by dividing the value of $\log p(X|\lambda)$ by T . This is carried out so that the duration effects arising from the log-likelihood value can be normalized out.

3.3. User registration employing RSA encryption

The naming of the RSA algorithm goes after the names of its inventors in 1978, who are Ron Rivest, Adi Shamir, and Leonard Adleman [24]. It is essential to be aware of the number of steps that a computer would take for the encryption of the message so that it can be found whether a technique is quick and effective.

$$\gcd(a, b) = 1 \text{ and } J(a, b) = a^{(b-1)/2} \pmod{b} \quad (4)$$

where $J(a, b)$ refers to the Jacobi symbol, which can also be denoted as (5).

$$\left(\frac{a}{b}\right) = \left(\frac{a}{p_1}\right)^{\alpha_1} \left(\frac{a}{p_2}\right)^{\alpha_2} \dots \left(\frac{a}{p_k}\right)^{\alpha_k} \quad (5)$$

where $b = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$ refers to the primal factorization of b , and the Legendre symbol is expressed for all integers a and all odd primes p defined by (6).

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } a \equiv 0 \pmod{p} \\ +1 & \text{if } a \not\equiv 0 \pmod{p} \text{ and for some integer } x, a \equiv x^2 \pmod{p} \\ -1 & \text{if there is no such } a \equiv 0 \pmod{p} \end{cases} \quad (6)$$

$$\left(\frac{a}{1}\right) \equiv 1$$

The Jacobi symbol is only specified if a is an integer and b is a positive odd integer. In addition, $J(a, b)$ is 0 if $\gcd(a, b) \neq 1$ and ± 1 if $\gcd(a, b) = 1$. It holds true always if b is prime, else (if b is composite), it will include a probability of being false by more than 50%. If holes true 100 times for randomly selected a 's, then b is roughly too certain to be prime, with a probability of being a 1's composite in 2100.

3.4. Gabor filter-based feature extraction

A Gabor filter is basically a linear band pass filter utilized in image processing for feature detection and texture analysis [25]. The representations of Gabor filters in terms of frequency and orientation are identical to those pertaining to the human visual system, and they have been observed to be especially suitable for texture representation and differentiation [26]. In the spatial domain, a 2D Gabor filter is basically a Gaussian kernel function that is modulated using a sinusoidal plane wave.

$$g(x, y) = \exp(-(x_1^2 + \gamma^2 y_1^2 / 2\sigma^2)) \exp(i(\frac{2\pi x_1}{\lambda} + \psi)) \quad (7)$$

where $x_1 = x \cos \theta + y \sin \theta$ and $y_1 = -x \sin \theta + y \cos \theta$. The input parameter σ stands for the standard deviation of the Gaussian function, λ indicates the wavelength of harmonic function, θ refers to the orientation, γ indicates the spatial aspect ratio with a fixed value of 0.5, and ψ refers to the phase shift of harmonic function.

In the above Figure 2 is represented by the convolutional networks take their inspiration from biological processes, where the connectivity pattern between neurons is identical to the way in which the animal visual cortex is organized. It has five layers, which include, two convolution layers, C1 and C2, and two pooling layers, P1 and P2, and an output layer. The size of the filter utilized in convolution layers C1 and C2 was 5×5 while the filter size utilized in pooling layers P1 and P2 was 2×2 . 28×28 grayscale images form the input to the CNN. The output of C1 and C2 provides a feature map of 32 sets for both. The outputs from P1 and P2 contain the same number of feature map sets like C1 and C2. The CNN output was a column vector with 256×1 dimension, utilized by the SoftMax classifier for classifying the input images into one out of the two output classes.

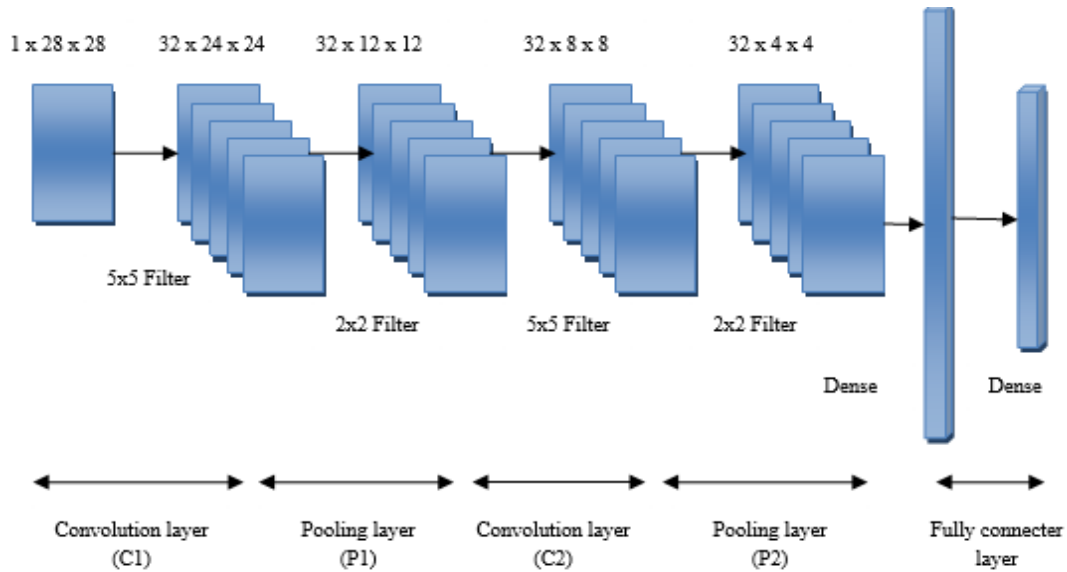


Figure 2. Enhanced CNN architecture

3.5. Hybrid genetic and CNN based voice authentication

Hybrid genetic algorithm (GA) and convolutional neural network (CNN) based voice authentication leverages the optimization capabilities of GAs to design and enhance the architecture of CNNs for robust voice recognition. Mapping the domain of artificial neural networks (ANN) to GA involves using evolutionary principles to optimize hyperparameters, such as network layers, learning rates, and activation functions [27]. This integration addresses challenges in network design by exploring a vast search space efficiently. GA aids in developing deep CNNs by automating the selection of optimal configurations, improving accuracy, and reducing training time. This hybrid approach enables adaptive and efficient voice authentication systems.

Figure 3 illustrates the enhanced genetic and CNN-based voice authentication process, highlighting the use of GA to develop a deep CNN. In this process, GA treats neural networks as individuals in a population. The performance of each network represents its fitness, guiding the selection of candidates for subsequent generations. Iterative steps in the algorithm focus on evolving the population by optimizing network parameters and architecture, aiming for improved accuracy. This synergy between GA and CNN demonstrates an effective means of automating and enhancing neural network design for robust voice authentication systems.

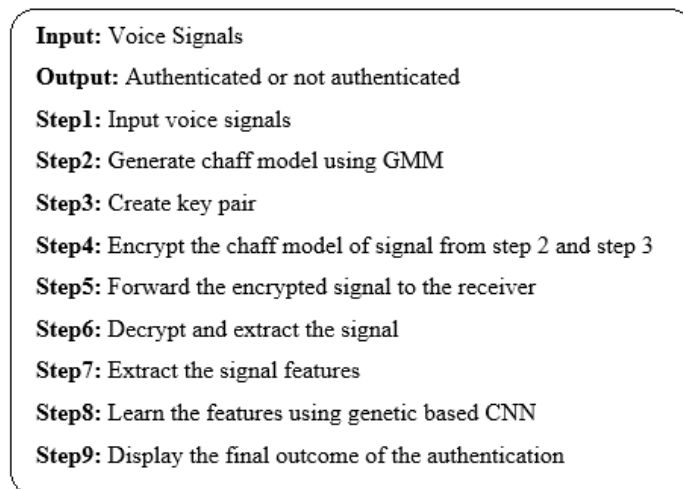


Figure 3. Enhanced genetic and CNN based voice authentication process

4. RESULTS AND DISCUSSION

The discussed two-level voice authentication system (TLVAS) assessment is carried out using the MATLAB simulation environment. The voiceprint is then compared against each other for improving the performance. The chaff model of the input signal will be produced, whose encryption will be done with the secret key pair. Then this encrypted signal will be relayed to the received where it is decrypted and then feature extraction will be carried out. The features extracted will be learned employing the genetic based CNN algorithm for the authentication process.

The presented figures illustrate various stages of the voice authentication process. Figure 4 demonstrates the input human voice processing, where the initial voice signal is captured and prepared for further analysis. This is a crucial step in ensuring that the raw audio data is suitable for subsequent processing stages. Figure 5 depicts the chaff model process, which is designed to enhance the security of the voice authentication system. The chaff model introduces decoy or “chaff” features into the data, making it resilient against unauthorized attempts and tampering. This step plays a significant role in preserving the integrity and confidentiality of the authentication process. Figure 6 outlines the feature extraction process, where the essential characteristics of the voice signal are identified and extracted. These features include key parameters such as pitch, frequency, and amplitude, which are unique to each individual and critical for accurate voice recognition. Together, these figures represent a systematic approach to achieving secure and reliable voice authentication. The input voice signal undergoes multiple layers of processing, from initial capture to feature extraction, with security enhancements provided by the chaff model. These stages collectively form the foundation of a robust and effective voice authentication system.

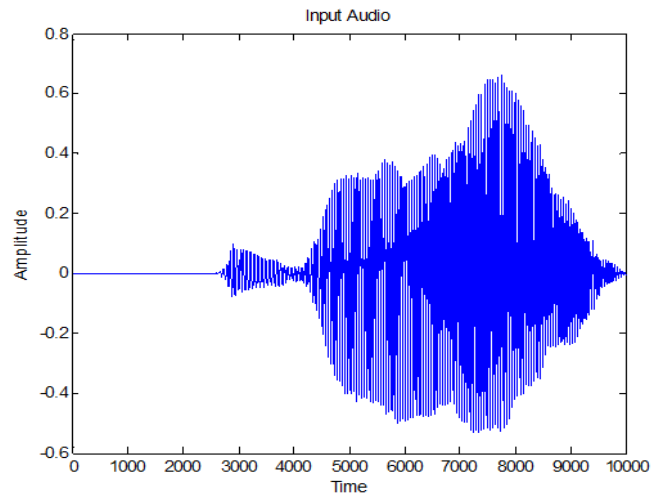


Figure 4. Input human voice processing

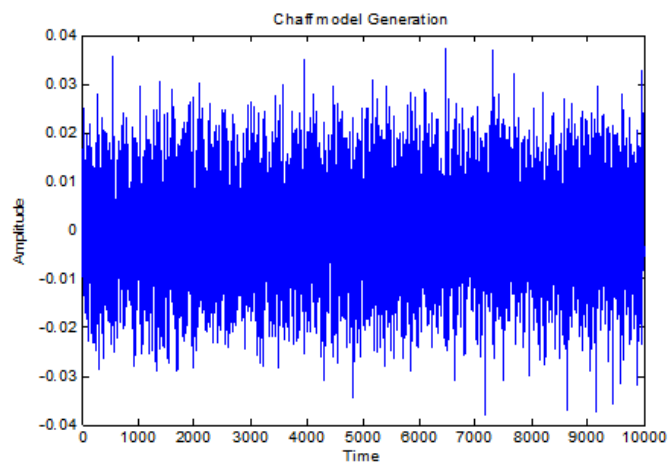


Figure 5. Chaff model process

Figure 6 shows the feature extraction process carried out on the chaff model of input signal. The learning of the extracted features will be done employing the genetic based CNN algorithm for the voice authentication results. Figure 6 outlines the feature extraction process, where the essential characteristics of the voice signal are identified and extracted. These features include key parameters such as pitch, frequency, and amplitude, which are unique to each individual and critical for accurate voice recognition. Together, these figures represent a systematic approach to achieving secure and reliable voice authentication. The input voice signal undergoes multiple layers of processing, from initial capture to feature extraction, with security enhancements provided by the chaff model. These stages collectively form the foundation of a robust and effective voice authentication system.

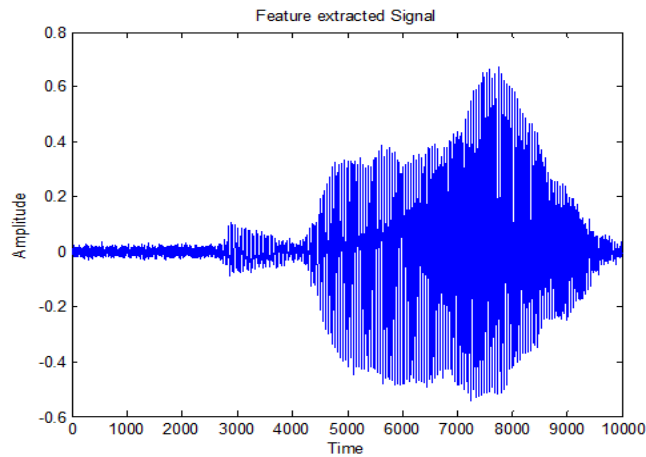


Figure 6. Extraction feature process

4.1. Sensitivity

Sensitivity is defined as the evaluation of ratio consisting of actual positive for categorizing the real voice to be the original. The sensitivity is defined as (8):

$$Sensitivity = \frac{T_p}{T_p + F_n} \quad (8)$$

where T_p refers to the actual voice identified right to be the authenticated voice. F_p indicates the intruder voice, which is wrongly authenticated to be intruder voice. F_n refers to the intruder voice is wrongly detected as the authenticated voice. The intruder voice T_n is identified correctly to be the intruder voice.

4.2. Precision

Precision can be computed by the ratio of true positives against the true positives added with false positives results of the voice signal. The precision is defined as (9).

$$Precision = \frac{T_p}{T_p + F_p} \quad (9)$$

4.3. Accuracy

Accuracy defines the overall correctness of the model. Accuracy is computed as follows: the ratio of the sum of actual classification parameters ($T_p + T_n$) to the sum of number of classification parameters ($T_p + T_n + F_p + F_n$).

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (10)$$

4.4. Signal to noise ratio

In science and engineering, signal-to-noise ratio measures the ratio between the degree of a given signal to the level of background noise. SNR is expressed as the ratio of signal power to the noise power, and its unit is generally decibels. A ratio greater than 1:1 implies more signal compared to noise.

$$SNR = \frac{P_{signal}}{P_{noise}} \tag{11}$$

4.5. Root mean square deviation

For the comparison of the encryption results with less fluctuation in the parameter, the root means square difference (RMSD) is calculated. Let $f_{\mu_1, x_1(0)}$ refer to the encryption result of the signal and let $f_{\mu_2, x_2(0)}$ be the non-encrypted signal. The RMSD between $f_{\mu_1, x_1(0)}$ and $f_{\mu_2, x_2(0)}$ is expressed as (12).

$$RMSD = \left(\frac{1}{L \times P} \sum_{i=0}^{L-1} \sum_{j=0}^{P-1} \left(f'_{\mu_1 x_1(0)}(i, j) - f'_{\mu_2 x_2(0)}(i, j) \right)^2 \right)^{1/2} \tag{12}$$

4.6. No invertibility index

The no invertibility index (NI) measures the resilience of biometric systems against inversion attacks. To safeguard confidential biometric data, some algorithms modify biometric information to make it irreversible, ensuring that attackers cannot reconstruct the original biometric features even if they gain access to the database. These transformations aim to provide robust security by rendering stored data unusable for identity theft or unauthorized access. High NI values indicate better resistance against reverse engineering of biometric traits, enhancing overall security measures.

In the Figure 7 shows the sensitivity evaluation is carried out and it is observed that the newly introduced technique yields better performance compared to the available techniques in the Table 1. The results of evaluation are given as below: for MFCC it is 87.5%, retina and fingerprint based multi-biometric system (RFBMBS) technique yields 87.9%, noise concerned accurate voice authentication system (NAVAS) achieves 91.5%, Supravallular aortic stenosis (SVAS) yields 92.4% and pediatric vasculitis activity score (PVAS) achieves 95.6%. The performance newly introduced techniques in terms of this precision assessment outperforms the available techniques. The MFCC yields 89.7%, RFBMBS yields 93.5%, NAVAS renders 93.8%. 94.6% achieves SVAS and PVAS achieves 96.2%. The evaluation of the performance of the discussed technique in terms of accuracy is much better than the available techniques. MFCC technique yields 82.00%, 85.00% yields RFBMBS, NAVAS gives 88%, SVAS gives 91% and PVAS renders 97.5%. When evaluated in terms of SVAS, the discussed technique PVAS achieves superior performance whereas it is 12.99% more than SVAS, 25.28% more than NAVAS, 82.91% more than RFBMBS and 114.65% more than MFCC. In terms of RMSD metric, PVAS demonstrates superior performance where it is 33.84% lesser compared to SVAS, 39.43% compared to NAVAS, 50.57% compared to RFBMBS and 54.25% compared to MFCC.

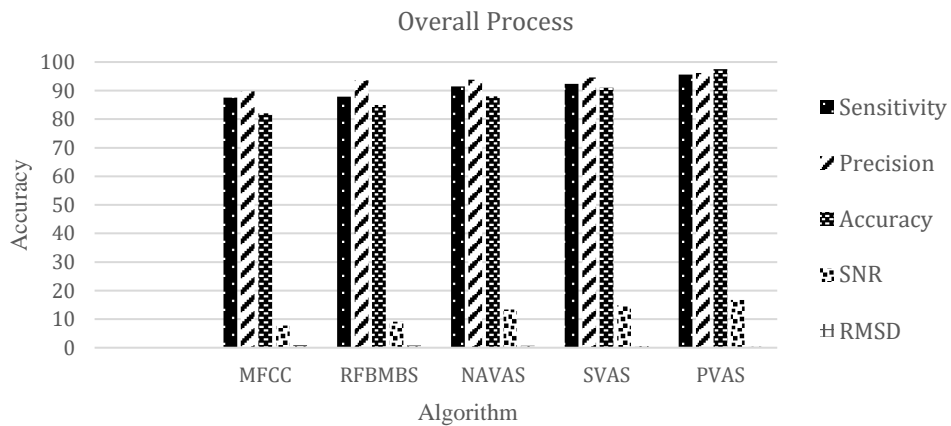


Figure 7. Overall performance

Table 1. Performance metric comparison values

Methods	Sensitivity (%)	Precision (%)	Accuracy (%)	SNR (dB)	RMSD
MFCC	87.50	89.74	82.00	7.78	0.94
RFBMBS	87.90	93.59	85.00	9.13	0.87
NAVAS	91.50	93.83	88.00	13.33	0.71
SVAS	92.40	94.60	91.00	14.78	0.65
PVAS	95.60	96.20	97.50	16.70	0.43

5. CONCLUSION

In this paper presents a pioneering framework that significantly enhances privacy protections in automated voice authentication systems, effectively addressing the increasing concerns surrounding the storage and processing of sensitive voice biometric data. The proposed methods have implementing advanced encryption techniques, privacy-preserving feature extraction, and a decentralized architecture, the proposed framework ensures robust privacy safeguards without sacrificing authentication accuracy and efficiency. Additionally, the integration of differential privacy mechanisms and a user-centric consent model further strengthens the framework's ability to protect individual privacy. Experimental results validate the framework's effectiveness, demonstrating superior authentication accuracy and a substantial reduction in privacy risks compared to traditional systems. This contribution represents a vital advancement in the quest to balance security and privacy within biometric authentication technologies.




REFERENCES

- [1] L. Schönherr, T. Eisenhofer, S. Zeiler, T. Holz, and D. Kolossa, "Imperio: robust over-the-air adversarial examples for automatic speech recognition systems," *ACM International Conference Proceeding Series*, pp. 843–855, 2020, doi: 10.1145/3427228.3427276.
- [2] A. M. Avram, V. Pais, and D. Tufiş, "Towards a Romanian end-to-end automatic speech recognition based on deepspeech2," in *Proceedings of the Romanian Academy Series A - Mathematics Physics Technical Sciences Information Science*, 2020, vol. 20, no. 4, pp. 395–402.
- [3] D. S. Park *et al.*, "SpecAugment: A simple data augmentation method for automatic speech recognition," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2019, pp. 2613–2617, doi: 10.21437/Interspeech.2019-2680.
- [4] M. Tamazin, A. Gouda, and M. Khedr, "Enhanced automatic speech recognition system based on enhancing power-normalized cepstral coefficients," *Applied Sciences (Switzerland)*, vol. 9, no. 10, 2019, doi: 10.3390/app9102166.
- [5] J. Li, R. Gadde, B. Ginsburg, and V. Lavrukhin, "Training neural speech recognition systems with synthetic speech augmentation," *arXiv preprint arXiv:1811.00707*, 2018.
- [6] H. Gupta and D. Gupta, "LPC and LPCC method of feature extraction in speech recognition system," in *Proceedings of the 2016 6th International Conference - Cloud System and Big Data Engineering, Confluence 2016*, 2016, pp. 498–502, doi: 10.1109/CONFLUENCE.2016.7508171.
- [7] C. H. H. Yang *et al.*, "Decentralizing feature extraction with quantum convolutional neural network for automatic speech recognition," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2021, pp. 6523–6527, doi: 10.1109/ICASSP39728.2021.9413453.
- [8] D. Palaz, M. Magimai-Doss, and R. Collobert, "End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition," *Speech Communication*, vol. 108, pp. 15–32, 2019, doi: 10.1016/j.specom.2019.01.004.
- [9] F. Tao and C. Busso, "End-to-end audiovisual speech recognition system with multitask learning," *IEEE Transactions on Multimedia*, vol. 23, pp. 1–11, 2021, doi: 10.1109/TMM.2020.2975922.
- [10] A. Habbal, H. Hamouda, A. M. Alnajim, S. Khan, and M. F. Alrifai, "Privacy as a Lifestyle: empowering assistive technologies for people with disabilities, challenges and future directions," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 4, 2024, doi: 10.1016/j.jksuci.2024.102039.
- [11] J. Guglani and A. N. Mishra, "Automatic speech recognition system with pitch dependent features for Punjabi language on KALDI toolkit," *Applied Acoustics*, vol. 167, 2020, doi: 10.1016/j.apacoust.2020.107386.
- [12] A. Winursito, R. Hidayat, and A. Bejo, "Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition," in *2018 International Conference on Information and Communications Technology, ICOIACT 2018*, 2018, pp. 379–383, doi: 10.1109/ICOIACT.2018.8350748.
- [13] A. T. Tran, T. D. Luong, and V. N. Huynh, "A comprehensive survey and taxonomy on privacy-preserving deep learning," *Neurocomputing*, vol. 576, 2024, doi: 10.1016/j.neucom.2024.127345.
- [14] T. A. M. Celin, T. Nagarajan, and P. Vijayalakshmi, "Data augmentation using virtual microphone array synthesis and multi-resolution feature extraction for isolated word dysarthric speech recognition," *IEEE Journal on Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 346–354, 2020, doi: 10.1109/JSTSP.2020.2972161.
- [15] K. Aida-Zade, A. Xocayev, and S. Rustamov, "Speech recognition using support vector machines," in *Application of Information and Communication Technologies, AICT 2016 - Conference Proceedings*, 2017, pp. 1–4, doi: 10.1109/ICAICT.2016.7991664.
- [16] V. Passricha and R. K. Aggarwal, "A hybrid of deep CNN and bidirectional LSTM for automatic speech recognition," *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 1261–1274, 2020, doi: 10.1515/jisys-2018-0372.
- [17] U. Shrawankar and V. Thakare, "Noise estimation and noise removal techniques for speech recognition in adverse environment," *IFIP Advances in Information and Communication Technology*, vol. 340 AICT, pp. 336–342, 2010, doi: 10.1007/978-3-642-16327-2_40.
- [18] Y. H. Taguchi and Y. Murakami, "Principal component analysis-based feature extraction approach to identify circulating microRNA biomarkers," *PLoS ONE*, vol. 8, no. 6, 2013, doi: 10.1371/journal.pone.0066714.
- [19] J. W. Hung, J. S. Lin, and P. J. Wu, "Employing robust principal component analysis for noise-robust speech feature extraction in automatic speech recognition with the structure of a deep neural network," *Applied System Innovation*, vol. 1, no. 3, pp. 1–14, 2018, doi: 10.3390/asi1030028.
- [20] Y. Zhang *et al.*, "Towards end-to-end speech recognition with deep convolutional neural networks," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2016, vol. 08-12-Sept, pp. 410–414, doi: 10.21437/Interspeech.2016-1446.
- [21] W. Han *et al.*, "Contextnet: improving convolution neural networks for automatic speech recognition with global context," *arXiv preprint arXiv:2005.03191*, 2020.
- [22] K. Chouhan, A. Singh, A. Shrivastava, S. Agrawal, B. D. Shukla, and P. S. Tomar, "Structural support vector machine for speech recognition classification with CNN approach," in *2021 9th International Conference on Cyber and IT Service Management, CITSM 2021*, 2021, pp. 1–7, doi: 10.1109/CITSM52892.2021.9588918.




- [23] A. Alsobhani, H. M. A. Alabboodi, and H. Mahdi, "Speech recognition using convolution deep neural networks," *Journal of Physics: Conference Series*, vol. 1973, no. 1, 2021, doi: 10.1088/1742-6596/1973/1/012166.
- [24] N. Khalid, A. Qayyum, M. Bilal, A. Al-Fuqaha, and J. Qadir, "Privacy-preserving artificial intelligence in healthcare: Techniques and applications," *Computers in Biology and Medicine*, vol. 158, 2023, doi: 10.1016/j.combiomed.2023.106848.
- [25] A. Iftikhar, K. N. Qureshi, M. Shiraz, and S. Albahli, "Security, trust and privacy risks, responses, and solutions for high-speed smart cities networks: a systematic literature review," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 9, p. 101788, 2023, doi: 10.1016/j.jksuci.2023.101788.
- [26] S. Singh, S. Rathore, O. Alfarrarj, A. Tolba, and B. Yoon, "A framework for privacy-preservation of IoT healthcare data using federated learning and blockchain technology," *Future Generation Computer Systems*, vol. 129, pp. 380–388, 2022, doi: 10.1016/j.future.2021.11.028.
- [27] C. Sandeepa, B. Siniarski, N. Kourtellis, S. Wang, and M. Liyanage, "A survey on privacy for B5G/6G: new privacy challenges, and research directions," *Journal of Industrial Information Integration*, vol. 30, 2022, doi: 10.1016/j.jii.2022.100405.

BIOGRAPHIES OF AUTHORS






Kathiresh Murugesan    is an assistant professor in the Department of Computer Science, Rathinam College of Arts and Science Coimbatore. He received his B.Sc. degree in mathematics from Bharathiar University. MCA degree at Bharathiar University and M.Phil. degree at Karpagam University. Ph.D. degree from Bharathiar University. He has more than 13 years of teaching experience and he published 8 research articles. His research interest includes network security, compiler design, and software engineering. He has finished one UGC sponsored minor research project. He can be contacted at: kathireshcs83@gmail.com.






Lavanya Subbarayalu Ramamurthy    is an associate professor in the Department of Computer Science at KPR College of Arts Science and Research, Coimbatore. She earned her Ph.D. in computer science from Bharathiar University. With over 19 years of academic experience, she is a leading expert in data mining and machine learning. Dr. Lavanya has published 15 papers in prestigious Scopus-indexed journals, peer-reviewed publications, online books, and has presented more than 15 papers at national and international conferences. Her work also includes patents, underscoring her significant contributions to the field. She can be contacted at email: drsrlavanya@gmail.com, lavanya.s.r@kprcas.ac.in.






Boopathi Palanisamy    is an assistant professor in the Department of Computer Applications at Nehru Arts and Science College, Coimbatore. He is pursuing a Ph.D. degree in computer science at Dr. SNS Rajalakshmi College of Arts and Science, Coimbatore with a specialization in data mining. He has more than 15+ years of experience in core computer science subjects. He has published more article papers in reputed national and international journals and also filled more than pattern papers in the same. He has been acting as a resource person in the various institutions. He can be contacted at: boopathimca1985@gmail.com.



Yamini Chandrasekar    is working as associate professor and head – BCA in Computer Science Department. She received her Ph.D. in computer applications from Anna University during 2014. Her area of research includes data mining, artificial intelligence, deep learning and machine learning. She has served as a resource person in FDP, seminar and served as a chair for national and international conference. She has organized various international conference, FDP, workshops and seminars. She has published around 30+ papers in national and international Journals which includes SCOPUS, Thomson Reuters, UGC, IEEE, etc. and also presented 25+ papers in various national and international conferences. She has published 5 books. She is the board of studies member for both PG - CS and UG-IT in Bharathiar University Coimbatore. Received sponsor to conducted FDP under AICTE ATAL FDP Scheme. She can be contacted at email: yaminics@srcw.ac.in.






Kavitha Masagoundanpudhur Shanmugam    is working as an assistant professor in the Department of Computer Science and Engineering at Akshaya College of Engineering and Technology, Coimbatore, Tamil Nadu in India. She received her bachelor's degree and Master's degree from Anna University Chennai in 2013 and 2015 respectively. She is currently pursuing her Ph.D. She has 7+ years of teaching experience. She is a lifetime member of Indian Society of Technical Education (ISTE). Her area of interest is image processing, machine learning, cloud computing and deep learning. She can be contacted at email: kavithams@acetbe.edu.in.






Balluru Thammaiahshetty Adishankar Nithya    working as an assistant professor, have more than 12 years of experience in teaching. Her area of interest is machine learning. She currently working as an assistant professor in the Department of Computer Science and Engineering with Presidency University. She can be contacted at email: nithyaba3@gmail.com.



Velumani Thiyagarajan    is working as an assistant professor in Department of Computer Science at Rathinam College of Arts and Science (Autonomous), Coimbatore 641 021 from 14-09-2020 to till date. He has got more than 12 years of teaching experience. He has obtained his B.Sc(CT), M.Sc(CS), M.Phil and MBA degrees from Periyar University Salem, Tamil Nadu, India. He has obtained his B.Ed degree from Indira Gandhi National Open University (IGNOU) at Delhi, his M.Sc (Psychology) from Madras University Madras, TamilNadu, India. He has obtained his Ph.D in computer science from Manonmaniam Sundaranar University at Tirunelveli, Tamil Nadu, India. His area of interest is image processing research directions involved knowledge discovery and data mining, pattern recognition, knowledge-based neural networks, software engineering and IoT. He has published more than 15 papers in topmost international peer-reviewed journals and conference. He can be contacted at email: velumani46@gmail.com.



Ramaraj Muniappan    is working as an assistnt professor in the Department of Computer Science at Rathinam College of Arts and Science, Coimbatore. He holds a Ph.D., degree in computer science at Bharathiar University in the year of 2020 with specialization in data mining with image process and also fuzzy logic in the image analysis. His research areas are data mining, image processing, fuzzy logic, pattern recognition and deep learning concept. He has published more research article in the reputed various national and international journals and also filed the patents in the same field. He has a reviewer of many international journals including with IEEE, ASTESJ, and JERS. He can be contacted at email: ramaraj.phdcs@gmail.com, ramaraj.phdcs123@gmail.com.