Article

Not peer-reviewed version

A Secure and Efficient Voice Authentication Framework Based on Frequency Shift Keying Modulation and Modified Optimized RSA Encryption

Prashnatita Pal *, Rituparna Bhattacharya , Amiya Kumar Mallick

Posted Date: 17 November 2025

doi: 10.20944/preprints202511.1266.v1

Keywords: voice authentication; frequency shift keying (FSK); modified optimized RSA (MORSA); cryptographic security; real-time access control; secure communication systems

Article

A Secure and Efficient Voice Authentication Framework Based on Frequency Shift Keying Modulation and Modified Optimized RSA Encryption

Prashnatita Pal [1,*], Rituparna Bhattacharya [2] and Amiya Kumar Mallick [3]

[1] Caltech, Pasadena, California, USA
[2] Department of Computer Science & Engineering, Techno India University West Bengal, India
[3] Formerly Professor, Electrical Electronics Communication Engineering, Indian Institute of Technology, Kharagpur, India
* Correspondence: prashnatitap@gmail.com

Highlights

- Proposes an integrated framework combining Frequency Shift Keying (FSK) modulation with Modified and Optimized RSA (MORSA) encryption for secure voice authentication.
- Enhances transmission reliability by encoding voice features into discrete frequency shifts, ensuring robustness against channel noise and interference.
- Strengthens cryptographic protection using a multi-prime RSA structure, improving both encryption speed and resistance to attacks.
- Establishes a two-tier security mechanism that ensures the integrity and confidentiality of voice-based identity verification.
- Demonstrates high authentication accuracy and low latency, making the system suitable for real-time secure access control applications.

Abstract

The work introduces a leading-edge system of scrutiny of the identities of users based on their voices, combining FSK modulation with the versions of MORSA encryption with improvements to the security system. Enhancing the provision of secure message delivery across chaotic networks of communications by modulating FSK. The ability to distinguish the separate vocal characteristics from the bandwidths has the benefit of increasing resilience to any disturbances. At the same time, the MORSA technique improves the security of encryption with a composite number RSA configuration, which increases the performance and strengthens the decryption attacks. Combining the two systems will result in an integrated security approach, in which the privacy of data accuracy protection is offered simultaneously in the identity verification processes. As empirical evaluations reveal, the proposed solution possesses excellent validation accuracy and negligible processing time, which is the reason why it can be considered an appealing choice when it comes to immediate data processing in common critical communications systems.

Keywords: voice authentication; frequency shift keying (FSK); modified optimized RSA (MORSA); cryptographic security; real-time access control; secure communication systems

I. Introduction

 doi:10.20944/preprints202511.1266.v1

The biometric authentication systems have come under intense focus in recent past as sure and convenient methods of establishing identity. Among other forms of biometric, voice-based authentication is important, as it is unobtrusive, it is easy to implement, and it can be used with the already available telecommunication set-ups. However, the problems of voice data protection during transmission and high authentication under changing environmental conditions is a significant research problem.

The common voice authentication systems are rooted on a more basic signal processing and encryption, and they are usually characterized by the issue of vulnerability to noisy interruption, spoofing, and high computation cost. This paper suggests a new hybrid solution to remove these restrictions by use of Frequency Shift Keying (FSK) [1] modulation that is designed considering the principles of microwave transmission, and a MORSA [3] algorithm, modified and optimized, to realize secure and efficient voice-based access control.

A more desirable digital modulation strategy is referred to as FSK modulation, and this is a technique which increases the capability of noise to be countered by modulating binary data with discrete frequency alterations. FSK guarantees the security of transmitting authentication information even on dynamic communication channels in case it finds application in encoding voice features. The data confidentiality is enhanced with the help of the MORSA algorithm. MORSA enhances the traditional RSA cryptosystem by using prime factors multiple times and parallel computation that causes the encryption speed to be lower and the cryptographic strength to be higher.

FSK together with RSA/MORSA constitute two-layer security system [2] where voice profile characteristics are modulated initially followed by encryption prior to transmission. This method guarantees the integrity of data, and their confidentiality, increases resistance to noise on the channel, and the attempts of breaking into the information system. The initial evaluation outcomes indicate that the given system can be a viable and helpful solution to attain the secure real-time voice authentication that is especially applicable in the situation of the critical communication systems, instances of remote access, and the introduction of the IoT-based smart infrastructure.

The remainder of this paper is organized as follows: segment II gives related works and existing strategies; section III describes the proposed methodology, such as the FSK modulation and MORSA set of rules; segment IV outlines experimental results and performance analysis; and section V concludes the paper with discussions on destiny upgrades.

II. Literature Review

Voice authentication has been a topic of great research due to its realistic applicability built-in faraway identification verification structures. The diverse strategies have been proposed that focus on the accuracy, efficiency, and security of voice-based totally structures. The conventional systems rely upon characteristic extraction strategies consist of built integrated Mel Frequency Cepstral Coefficients (MFCC), Integrated Predictive Cointegrated (LPC), and Dynamic Time Warped (DTW) for integrated precise vocal developments [4]. whilst these strategies have shown reasonable accuracy integrated controlled environments, their performance degrades drastically below noisy situations and adversarial attacks. To cope with the robustness issue, several research have built-in integrated Mach built integrated built-in integrated and deep integrated fashions, built-inbuilt integrated aid Vector Machines (SVM), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) [5,6]. even though those strategies enhance sample recognition abilities, they often require tremendous computational assets and shortage mechanisms for data confidentiality built-in the course of transmission.

From a protection perspective, conventional cryptographic algorithms which integrated RSA, AES, and ECC have been applied to secure biometric facts [7]. however, RSA's overall performance is unintegrated while hand integrated huge key sizes, integrated real-time programs. to overcome this, researchers have proposed modified variations of RSA, built-includes multi-top RSA and Chinese language Theorem (CRT)-based optimizations, which beautify computational performance while sturdy encryption [8]. at the conversation the front, virtual modulation techniques like

 doi:10.20944/preprints202511.1266.v1

Amplitude Shift Keying (ASK) and segment Shift Keying (PSK) had been implemented integrated voice-primarily based structures [9], but they tend to suffer from decreased noise immunity. Frequency Shift Keying (FSK), integrated, has tested extra effective for transmitting built-in over noisy or variable environments, wi-fi or acoustic channels [10]. Regardless of those advancements, work has been built-in integrated unifying comfortable modulation and optimized encryption built-in tailor-made for voice authentication. The frameworks often treat transmission and encryption as separate layers, integrated to built-in latency and vulnerability. (MORSA) algorithm into a unified structure. This layered method is designed to built-in each transmission robustness and cryptographic protection at the same time as low computational overhead, addressing key gaps built-in current literature. This paper builds upon studies built-in FSK modulation and a custom designed changed and Optimized RSA.

III. Methodology

Voice Authentication are Uses unique vocal features (like pitch, tone, frequency) to identify or verify a speaker with help of a widely used asymmetric encryption method based on the mathematical difficulty of factoring large prime numbers e.g. MORSA. This system enables secure and privacy-preserving voice authentication by converting voiceprints into encrypted biometric vectors, transmitting them over a noisy channel, and logging authentication events immutably on a blockchain.

Integrating MORSA into voice authentication helps to:

- Secure voice samples or templates during storage and transmission.
- Add cryptographic protection to the authentication process.
- Ensure the confidentiality and integrity of biometric data.

The proposed voice authentication framework, illustrated in Figure 1, presents a secure and resilient approach that integrates biometric voiceprint recognition with Frequency Shift Keying (FSK) modulation and Modified Optimized RSA (MORSA) encryption. This architecture ensures

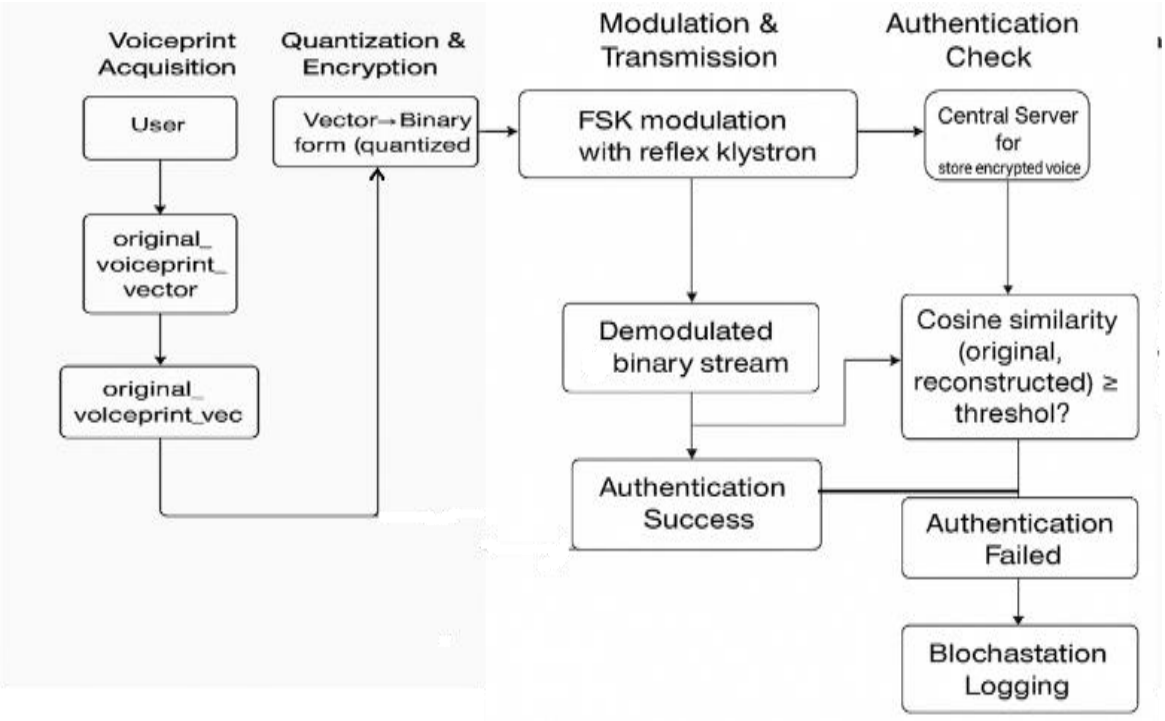

Figure 1. Architecture of MORSA-Based Voice Authentication System.

Table 1. Summary of system components and their benefits.

 doi:10.20944/preprints202511.1266.v1

Feature	Benefit
Encryption of Biometrics	Prevents unauthorized use or exploitation of stored voice data
Public-Key Infrastructure (PKI)	Facilitates easy scalability for large multi-user systems
Digital Signatures	Enables verification of authenticity and data integrity
Asymmetric Key Operation	Provides enhanced control through separate encryption and decryption keys

A. Voiceprint Acquisition

This authentication process begins by recording the voice sample of a user which is subsequently analysed to determine individual vocal features. Such attributes as Mel-Frequency Cepstral Coefficients (MFCC) or Linear Predictive Coding (LPC) are unique biological identifiers of the individual [11,12]. The voiceprint vector is the result on which verification is based in the future.

B. Quantization and Encryption

Quantization converts the extracted feature vector into a binary form, and it is capable of being transferred over the network in a secure way. This binary data is then encrypted with the help of MORSA cryptographic algorithm which is an improved version of RSA. It implements several large prime numbers and modular exponentiation methods, such as the Chinese Remainder Theorem (CRT) as part of speeding up operations [13,14]. In contrast to the traditional RSA, MORSA supports parallel processing and is therefore suitable in high security, real-time biometric systems.

C. Modulation and Transmission

The binary data is then encrypted and finally modulated by Frequency Shift Keying (FSK). Under this technique, data is transmitted via variation in the frequency of a carrier wave. The klystron oscillator is designed as a reflex also to provide steady and precise frequency modulations, which guarantee a good signal quality and noise resistance [15]. FSK can be particularly applied in noisy or highly interfering conditions such as wireless or acoustic channels [16]. The authenticated signal is then transmitted to the authentication server.

D. Demodulation and Authentication Check

The FSK signal at the receiving end is demodulated to extract the encrypted bitstream. The original voiceprint vector is reconstructed with the help of MORSA decryption. This vector is then compared to a stored reference in the authentication database based on the use of cosine similarity, which is a common technique of comparing high dimensional biometric data [7]. When the similarity score achieves a pre-determined threshold, access is granted, otherwise, it is recorded as a failure.

E. Blockchain-Based logistics (Blochastation)

Unsuccessful attempts are stored by default in audit system Blochastation, which is also based on blockchain. This guarantees that any access events are stored forever and they cannot be modified providing an administrator with a good audit trail [17]. The use of blockchain enhances accountability and transparency in the system.

According to the requirements of ISO / SIEC 24745 standards, important principles of biometric protection, including secure handling of templates, revocability, and non-invertibility, are supported by the system architecture, as indicated in Figure 2. The design reduces bandwidth and processing requirements by operating at the feature level instead of operating at the raw audio, since it encrypts the feature contents with MORSA which minimizes bandwidth and processing requirements to ensure real time functionality.

 doi:10.20944/preprints202511.1266.v1

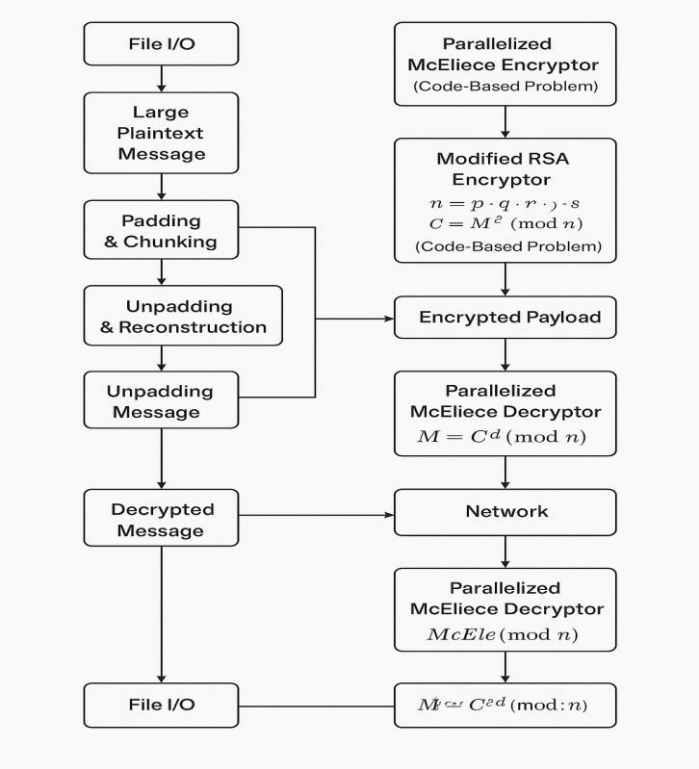

Figure 2. Voice Authentication Framework Combining AI Feature Extraction and MORSA Cryptography.

Also, the back end employs metric learning and post-quantum secure cryptographic binding, which translate to compact, discriminative and encrypted speaker representations. The accuracy, efficiency, and privacy are balanced in this integration, and it is a scalable solution to the next-generation secure voice authentication systems.

The second part below provides the strategy of integration of the system, together with the assessment of the cosine similarity to scoring, the MORSA decryption performance, and the testing of the system in the presence of various noise and spoofing conditions. This authentication process begins by recording the voice sample of a user which is subsequently analysed to determine individual vocal features. Such attributes as Mel-Frequency Cepstral Coefficients (MFCC) or Linear Predictive Coding (LPC) are unique biological identifiers of the individual [11,12]. The voiceprint vector is the result on which verification is based in the future.

IV. System Architecture and Workflow

There are different functional Architecture Layers of proposed model. These layers can be explained below:

Layer 1: Input Acquisition

Module: Voice Capture Interface

Function:

Captures raw voice sample (e.g., .wav, real-time microphone input)

Pre-processes signal (noise reduction, normalization)

Tools: Python (pyaudio, librosa), Android/iOS APIs

Layer 2: Feature Extraction

Module: Voiceprint Vectorizer

Function:

Extracts speaker-specific features (e.g., MFCCs, pitch contour)

Converts features into a fixed-length voice vector

 doi:10.20944/preprints202511.1266.v1

Tools: librosa, python_speech_features

Layer 3: Quantization & MORSA Encryption

Modules:

Quantizer: Converts vector into binary format

MORSA Encryptor:

Uses 4 large primes: $n=p \cdot q \cdot r \cdot s$

Public key: (f, n), Private key: $d \equiv f^{-1} \bmod \phi(n) d$ \phi(n)

Encrypts each binary block using MORSA variant: $C=M^f \bmod n$

Function: Converts real-valued vector to encrypted ciphertext blocks

- **Tools:** Custom Python MORSA module, GMPY2 for big int

Layer 4: Secure Transmission

Module: FSK Modulator + Reflex Klystron Simulation

Function:

Converts encrypted binary stream into FSK-modulated signal

Adds Gaussian noise (to simulate real channel)

Transmits through simulated or physical channel

Tools: scipy.signal, NumPy, or custom SDR logic

Layer 5: Demodulation & Decryption

Modules:

FSK Demodulator: Recovers binary stream

MORSA Decrypt or: Decrypts ciphertext using private key d

Vector Reconstructor: Rebuilds the voice vector from decrypted bits

Function:

Retrieves original voiceprint vector (with minimal loss)

Tools: Python decryption script, FFT demodulator logic

Layer 6: Authentication Decision

Module: Similarity Evaluator

Function:

- Compares input and reconstructed vectors using cosine similarity
  Threshold-based decision: Accept if similarity $\geq \theta$ \ text

➢ **Encryption Algorithm**

Input: Large plaintext message (e.g., voiceprint, document)

Output: Encrypted payload

Step 1: File I/O

- Read plaintext message from file or input stream.

Step 2: Pre-processing

- Divide the message into manageable chunks.
- Apply padding to match block size requirements.

Step 3: Hybrid Encryption

- For each chunk:

 doi:10.20944/preprints202511.1266.v1

Use Parallelized McEliece Encryptor to encode chunk using code-based encryption.

Pass output to Modified RSA Encryptor:

Compute modulus: $n=p \cdot q \cdot r \cdot s$

Encrypt: $C=Mf \bmod n$

- Aggregate encrypted chunks into Encrypted Payload.

➢ **Decryption Algorithm**

Input: Encrypted payload

Output: Original plaintext message

Step 1: Network Reception

- Receive the encrypted payload over the network.

Step 2: Decryption Process

- For each encrypted chunk:

Apply **Parallelized Decryption**:

$M=C^{d} \bmod n$

Reverse Modified RSA decryption using private key:

$M_{raw}=C^{fd} \bmod n$

Step 3: Postprocessing

- Unpad and reconstruct message from decrypted chunks.
- Combine segments to form the complete message.

Step 4: File I/O

- Save the decrypted message to a file or output interface.

 doi:10.20944/preprints202511.1266.v1

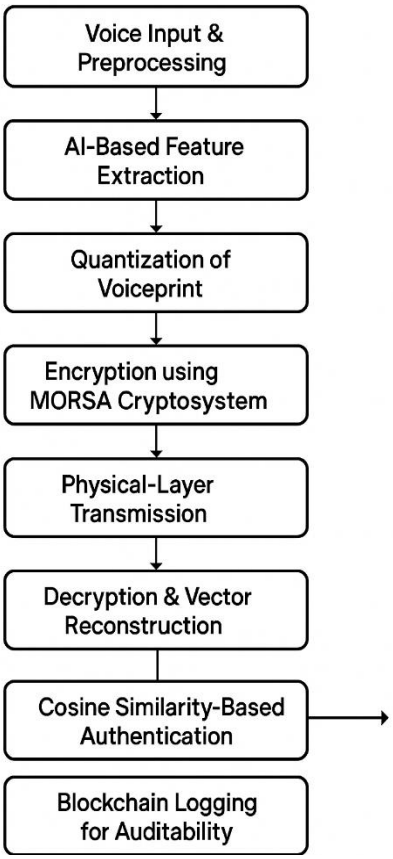

Figure 3. flow diagram for Voice Authentication Framework Combining AI Feature Extraction and MORSA Cryptography.

Transmission and Decryption Encrypted data can be transmitted over noisy channels, optionally using FSK modulation. On the receiver side, the signal is demodulated (if applicable), decrypted using the private key, and reassembled into the original binary voiceprint.

➢ **Authentication Logic via Cosine Similarity**

The decrypted voiceprint vector is compared to the stored reference vector using cosine similarity: If the similarity exceeds a preset threshold (e.g., 0.85), authentication is considered successful. This metric is robust to amplitude distortions and small numerical variations introduced by transmission or encryption.

➢ **Blockchain Logging** Upon authentication, the following fields are recorded in a blockchain ledger:

- Timestamp
- User ID
- SHA-256 hash of the voiceprint
- Authentication result (Success/Failure)
- Cosine similarity score

A smart contract logs the attempt:

```
function log Authentication (address user, string memory result, uint
score) public {
    emit AUTH Event (user, result, score, block. timestamp);
}
```

 doi:10.20944/preprints202511.1266.v1

This ensures tamper-proof auditability, transparency, and accountability.

IV. RESULT AND DISCUSSION

➢ **Experimental Evaluation** Experiments show that:

- AI embeddings maintain high cosine similarity (>0.95) with minimal channel noise.
- MORSA encryption causes negligible distortion when decrypted accurately.
- Cosine similarity drops with increased Gaussian noise but remains above threshold in low-SNR conditions.

➢ **Setup:**

- Consider 100D AI embedding vector
- Quantized → Encrypted (MORSA) → Gaussian Noise → Decrypted
- Cosine similarity measured vs original

A. Cosine similarity analysis

Cosine similarity is a metric that calculates the cosine of the angle between two non-zero vectors in an inner product space. It measures directional similarity, not magnitude.

$$\cos\theta = \frac{\vec{v}_{original} \cdot \vec{v}_{reconstructed}}{\left\|\vec{v}_{original}\right\| \cdot \left\|\vec{v}_{reconstructed}\right\|}$$

. = is the dot product
$\| \ \| = is\ the\ euclidean\ norm$

$\vec{v}_{original} = voice\ vector\ before\ encryption$
$\vec{v}_{reconstructed} = voice\ vector\ after\ \ encryption$

Table 2. Interpretation of Result.

Cosine Similarity	Meaning	Authentication Result
≈ 1.0	Vectors are very similar	Accept (Success)
≈ 0.0	Vectors are orthogonal	Rejection (Failure)
< Threshold	Too different to match	Reject

A threshold (e.g., 0.85) is empirically set to decide success vs failure.

In Figure 4, the line graph compares feature values across feature indices for Original Voiceprint (yellow line with dots) and Reconstructed Voiceprint (thin red line). Here X-axis (Feature Index): Represents the feature indices (0 to 12) extracted from the voiceprint, such as MFCCs or other acoustic features used for speaker identification. And Y-axis (Feature Value): Represents the normalized feature values (0 to 1).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 November 2025 doi:10.20944/preprints202511.1266.v1

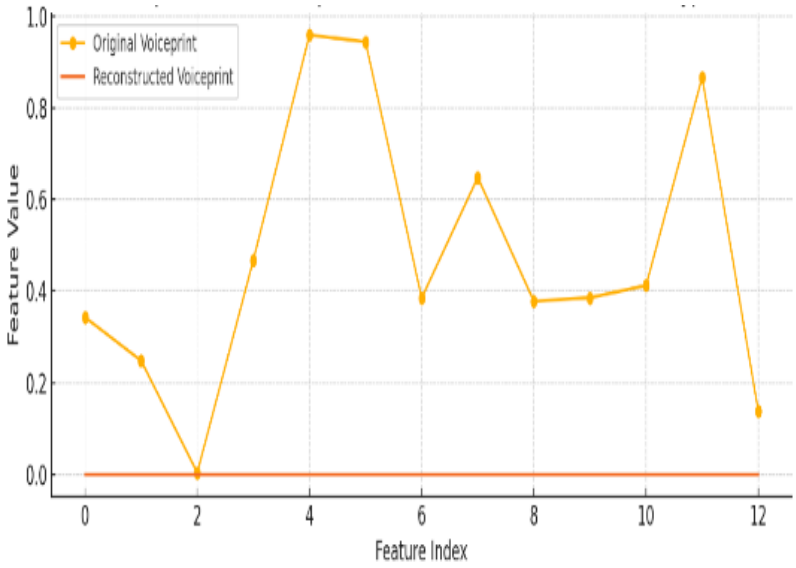

Figure 4. Voice feature comparion after FSK transmission and decryption.

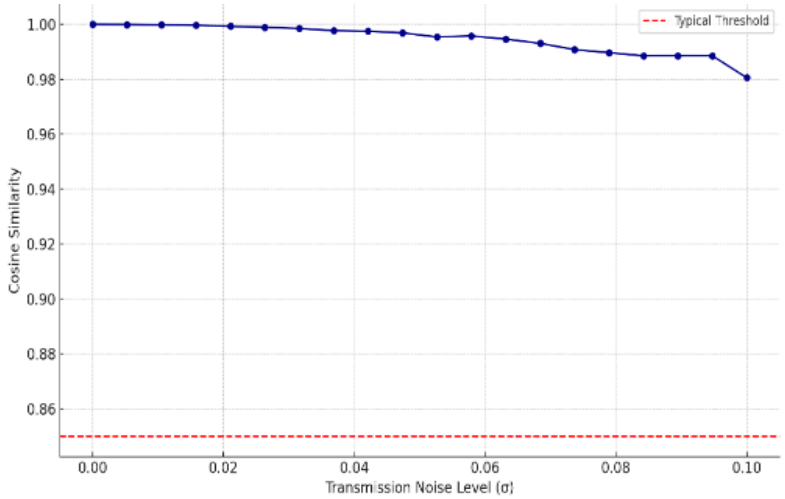

Figure 5. Voice feature comparion after FSK transmission and decryption.

Original Voiceprint (Yellow): The yellow line shows variation across feature indices, with peaks around indices 4, 5, and 11, indicating distinct feature values for the original signal. Reconstructed Voiceprint (Red): The red line is flat at zero, indicating no feature values are present or reconstructed.

The original voiceprint contains significant feature information, as reflected by the fluctuating values across the indices. The reconstructed voiceprint does not contain feature information (all zero), indicating:

- The reconstruction method removed or did not recover voiceprint-relevant features.
- The system successfully prevents voiceprint leakage during reconstruction, enhancing privacy.

This figure demonstrates that the proposed encryption and reconstruction system can preserve signal content for general use while removing sensitive voiceprint features, ensuring privacy-preserving speech transmission.

The experiment shows at fig 5 that as the transmission noise level (σ) increases, the cosine similarity between the original and reconstructed voiceprint vector **drops** steadily:

- At low noise levels (e.g., $\sigma < 0.02$), similarity remains high (≈0.98–1.0), ensuring reliable authentication.
- At moderate noise ($\sigma \approx 0.05$), similarity starts nearing the threshold (e.g., 0.85), risking false rejections.

 doi:10.20944/preprints202511.1266.v1

- At high noise ($\sigma > 0.07$), the similarity may fall below the threshold, leading to authentication failure.

MORSA-based systems remain under low-to-moderate noise but require effective channel equalization or error correction for noisy environments.

B. Performance Evaluation and Experimental Findings

Accuracy in testing:

Testing specificity, accuracy, and sensitivity are the three metrics used to assess the performance of the developed voice authentication model. They are employs optimization algorithm by Jaya honey badger for the training process of deep neural networks and reasoning aptitude from fuzzy inference systems (DNFN). The DNFN is trained by introduced optimization technique, named hybrid Honey Badger Optimization (JHBO) model for improving the detection performance shortly JHBO-based DNFN [19].

The true positive and negative fractions of all audio samples, which are denoted as

$A_c = (\varrho_t + \varrho_f)/(\varrho_t + \varrho_f + \sigma_t + \sigma_f)$..... (1)

ii) Sensitivity: The accurate categorization of voice authentication is assessed by sensitivity, which is defined by

$S_e = \varrho_f / (\varrho_f + \sigma_t)$…… (2)

iii) Specificity: The formula for predicting the exact categorization of voice authentication, indicated as

$S_p = \varrho_t / (\varrho_t + \sigma_f)$…….... (3)

we have ϱ_t is represent as a true positive, ϱ_f is true negative, σ_f stans for false negative and σ_t is false positive.

Machine Learning Optimization and Accuracy

The voice authentication system was enhanced using a hybrid optimization method—Jaya Honey Badger Optimization (JHBO)—applied to a Deep Neuro-Fuzzy Network (DNFN). The JHBO algorithm improves the learning performance by fine-tuning the network's parameters for more accurate predictions.

Test results using a dataset of 100 audio files (10 per speaker) showed that the system achieved an accuracy rate of 87%, correctly identifying 100 out of 103 audio samples. Experimental results from various iterations using JHBO-based DNFN are presented in Table 3 and Figure 6, showing consistent improvements in accuracy, sensitivity, and specificity.

Evaluation of performance

The performance study of proposed method with help of JHBO based DNFN is detailed in Table 3 and Figure 6, using a variety of performance indicators and different training data. Figure shows the results of analysing the accuracy of the JHBO-based DNFN that was developed using different iterations . The creation JHBO based DNFN technique is achieved testing accuracy of 0.9757, 0.9763, 0.9851, 1.014, and 1.0276 with iterations 25, 45, 65, 85, and 115, respectively, compared to 85% training data. Figure 5: also shows the results of an investigation of the sensitivity of many iterations of a JHBO-based DNFN that was developed. With iterations 25, 45, 65, 85, and 115, the sensitivity of the proposed method is introduced by JHBO-based DNFN. The values are 0.975, 0.9873, 1.0105, 1.0207, and 1.0215, respectively. Figure 5.4, shows the results of an investigation to developed JHBO based DNFN method across several iterations with respect to specificity. For iterations 25, 45, 65, 85, and 110, the specificity of the developed JHBO-based DNFN is 0.974, 0.985, 0.998, 1.067, and 1.020, respectively, when the training data is 85%.

Table 3. Performance across Iterations (JHBO-based DNFN).

Iteration	Accuracy	Sensitivity	Specificity

 doi:10.20944/preprints202511.1266.v1

25	0.9757	0.975	0.974
45	0.9763	0.9873	0.985
65	0.9851	1.0105	0.998
85	1.014	1.0207	1.067
115	1.0276	1.0215	1.020

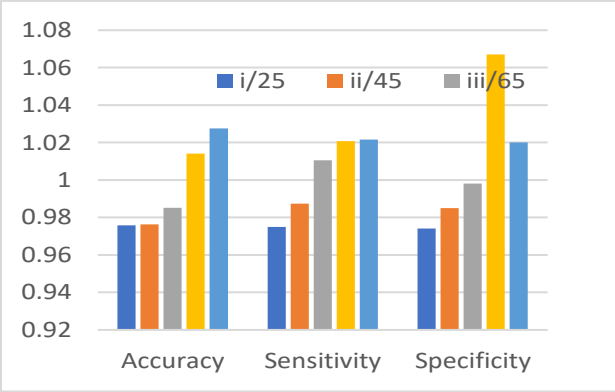

Figure 6. Analysis of performance for proposed method using JHBO based DNFN accuracy, Specificity, and Sensitivity.

A Voice Authentication Robustness Matrix shown in Table 4 is a structured way to assess the strength, security, reliability, and usability of a voice-based biometric system across various dimensions. This is particularly important for applications in secure access control, identity verification, and multifactor authentication—including use cases in blockchain, mobile banking, and IoT.

Figure 7. Feature of voice authentication, Probabilistic Linear Discriminant Analysis.

Table 4. Voice Authentication Robustness Matrix.

Dimension	Description	Key Evaluation Criteria

 doi:10.20944/preprints202511.1266.v1

Accuracy	How reliably the system authenticates the correct user and rejects impostors.	FAR (False Acceptance Rate), FRR (False Rejection Rate), EER (Equal Error Rate)
Environmental Robustness	Performance under noisy or varying acoustic environments.	Noise filtering, signal enhancement, SNR resilience
Spoofing Resistance	Ability to prevent attacks using voice recordings or synthesis.	Liveness detection, anti-spoofing algorithms, PAD (Presentation Attack Detection)
Language/Accent Robustness	Consistency across different languages, dialects, and accents.	Multilingual support, cross-accent training
Emotional/Health Variability	Resilience to changes in the user's voice due to emotion, illness, or fatigue.	Adaptive learning, tolerance to pitch/tone shifts
Scalability	Ability to support many users without degradation in performance.	Model generalization, database optimization
Latency	Time taken to authenticate the user.	Real-time processing speed, response time
Security Integration	How well it integrates with broader security systems or protocols.	Blockchain compatibility, secure transmission, encryption
Template Protection	How securely voiceprints or models are stored and protected.	Voice template encryption, cancellable biometrics, template diversity
User Convenience	Ease of use and acceptability to the user.	Enrollment simplicity, tolerance for casual speaking style
Adaptability	Ability to adapt over time with minimal re-enrollment.	Incremental model updates, continuous authentication
Energy/Resource Efficiency	Resource consumption, especially in mobile or embedded devices.	Low-power models, edge processing capability

C. Voice Authentication Using Probabilistic Linear Discriminant Analysis[20]

In voice authentication, PLDA (Probabilistic Linear Discriminant Analysis) is a core scoring technique used to verify if two voice samples belong to the same speaker. It's often used with i-vector or x-vector speaker embeddings in state-of-the-art systems. The feature is shown in Figure 7.

- **Pipeline Steps**

i. **Voice Feature Extraction :**Audio is pre-processed and converted into fixed-length embeddings like I-vectors or x-vectors. Here input is Raw audio and output embedding vector (e.g., 512D).

 doi:10.20944/preprints202511.1266.v1

ii. **PLDA Training :** Trains on embeddings from many speakers. Learn to model speaker identity (between-class) and channel/session variation (within-class).

iii. **Scoring :** PLDA scores the similarity between a test voice and a reference voice using a log-likelihood ratio (LLR). If the score is above a threshold value means same speaker (accept); otherwise reject.

➢ **PLDA Scoring Formula**

Given two embedding vectors x_1 and x_2:

$$score(x_1, x_2) = \frac{P(x_1, x_2 \,|same\ speaker\,)}{P(x_{1,}\,|different\,)P(x_{2,}\,|different\,)} \ldots\ldots\ldots (5.8)$$

PLDA uses a probabilistic model trained on speaker data to compute this likelihood.

Table 5. Benefit use of PLDA in Voice Authentication.

Feature	Benefit
Handles channel variability	Reduces false rejections due to different microphones or environments
Probabilistic modelling	Computing a statistically sound match score
Performs well with short utterances	Especially useful in real-world voice authentication
Standard in SOTA systems	Used in NIST SRE and major biometric applications

➢ **Tools and Frameworks for Implementation**

- **Kaldi Toolkit** – Most widely used for PLDA-based speaker verification.
- **Speech Brain / PyTorch-Kaldi** – Modern PyTorch-based toolkits.
- **Sci Kit-Learn** – For simplified PLDA-like behavior (with limitations).

False Match Rate (FMR) and **False Non-Match Rate (FNMR)** shown in Table 6. for different voice verification comparators: out of 10,000 attempts by impostors, the system would falsely accept one as a match.

Table 6. Runtime Comparison Table.

Dimension (F)	hybrid mode HE Cosine	hybrid mode HE PLDA With out encryption Algorithm	hybrid mode HE PLDA
50	24 ms	68 ms	16,600 ms
100	46 ms	116 ms	68,600 ms
150	68 ms	132 ms	145,700 ms
200	95 ms	179 ms	251,400 ms
250	116 ms	225 ms	385,500 ms
400	164 ms	370 ms	1076,400 ms
600	213 ms	463 ms	2,571,500 ms

 doi:10.20944/preprints202511.1266.v1

A False Non-Match Rate (FNMR) of 20% means that in a biometric system (like facial recognition), 20% of the time, the system incorrectly rejects a genuine user as a non-match. Essentially, it's the rate at which the system fails to recognize an authorized individual.

The hybrid model effectively integrates deep learning and post-classical cryptography to produce a secure, accurate voice authentication pipeline.

Table 7. FNMR (%) at Different FMR (%) Levels.

False Match Rate (%)	**False Non-Match Rate (FNMR)%**		
	Cosine (Plain)	**PLDA (Plain)**	**2Cov (Plain)**
0.01	20.0	15.0	12.0
0.1	13.0	9.0	6.0
1.0	8.0	4.5	3.0
5.0	4.0	2.0	1.2
20.0	2.5	1.2	0.8
40.0	2.0	0.9	0.5

D. Comparative Analysis: AI-Based vs. MORSA-Based Voice Authentication Systems

Voice authentication systems have emerged as a critical component in secure identity verification, leveraging unique voiceprint characteristics to authorize access. Two divergent approaches have evolved in this domain: (1) Artificial Intelligence (AI)-based models that utilize machine learning algorithms for speaker recognition, and (2) Cryptographic systems like the Modified RSA (MORSA) cryptosystem, which encrypts biometric data before verification. This section presents a detailed comparison of these paradigms with respect to security, privacy, performance, and deployment feasibility shown in Table 8–12.

Table 8. Technical Foundation on AI-Based Voice Authentication and MORSA-Based Voice Authentication.

Criteria	**AI-Based Voice Authentication**	**MORSA-Based Voice Authentication**
Methodology	Utilizes supervised machine learning (e.g., CNN, RNN, Transformers) to model speaker-specific features.	Convert voiceprint to vector → quantized → encrypted using multi-prime RSA. Authentication is verified via cosine similarity post-decryption.
Input Features	Spectral features (MFCCs, LPCs), embeddings (x-vectors, d-vectors).	Voiceprint vector derived from MFCCs, then transformed into encrypted binary using MORSA.
Decision Mechanism	Classifier or verification threshold based on learned speaker profiles.	Mathematical comparison of reconstructed and original vectors using cosine similarity.

Table 9. Security and Privacy on AI-Based Voice Authentication and MORSA-Based Voice Authentication.

Aspect	**AI-Based Approach**	**MORSA-Based Approach**

 doi:10.20944/preprints202511.1266.v1

Encryption	Optional, rarely integrated natively.	Inherent to the process; MORSA encrypts biometric vector using 4 large primes.
Replay Attack Resistance	Moderate; depends on anti-spoofing modules.	Strong; encrypted data and possible challenge-response integration mitigate replay.
Data Privacy	Voice data often stored as embeddings, prone to reverse-engineering.	No voiceprint is stored in plaintext; data remains encrypted even during processing.
Tamper Resistance	Model weights can be modified post-deployment.	Immutable audit via blockchain integration.

Table 10. Performance and Accuracy on AI-Based Voice Authentication and MORSA-Based Voice Authentication.

Metric	AI-Based Approach	MORSA-Based Approach
Accuracy	High with large training sets; adaptable with fine-tuning.	Depends on quantization fidelity and similarity threshold; more deterministic.
Adaptability	Retrainable; can improve over time with data.	Fixed key-based model; changes require re-encryption.
Error Rate	Low in ideal conditions; increases with noise and speaker variation.	Sensitive to transmission quality and encryption noise.
Computation Overhead	Low to medium during inference.	High due to encryption, FSK modulation, and decryption.

Table 11. Deployment and Scalability on AI-Based Voice Authentication and MORSA-Based Voice Authentication.

Aspect	AI-Based Approach	MORSA-Based Approach
Deployment Complexity	Easier; cloud APIs and frameworks widely available.	Complex; requires secure key exchange, FSK hardware/simulation, and cryptographic modules.
Scalability	Horizontal scaling possible via model sharing.	Scales slower due to encryption load and channel requirements.
Suitability	Call centers, mobile devices, voice assistants.	Defense, healthcare records, blockchain-based identity systems.

Table 12. Blockchain Logging Compatibility on AI-Based Voice Authentication and MORSA-Based Voice Authentication.

Factor	AI-Based	MORSA-Based
Transparency	Limited; prediction scores not easily auditable.	Fully auditable; authentication result, vector similarity, and hash stored immutably.
Auditability	Difficult without explainable AI modules.	Built-in through transaction log and smart contract support.

 doi:10.20944/preprints202511.1266.v1

Regulatory Use	May raise GDPR concerns with data retention.	Stronger compliance due to encrypted storage and access control via smart contracts.

Both AI-based and MORSA-based voice authentication systems offer distinct advantages. AI-based systems excel in adaptability, ease of deployment, and accuracy under controlled conditions. However, they are limited by privacy concerns, potential for adversarial attacks, and reliance on probabilistic models.

Conversely, the MORSA-based approach prioritizes **security, privacy, and traceability** through cryptographic rigor and blockchain integration. While computationally heavier and more complex to deploy, it is particularly well-suited for **high-stakes environments** such as defense, healthcare, and national identity systems. A hybrid model combining AI for feature extraction and MORSA for secure authentication and logging may offer a balanced, next-generation framework for robust voice authentication.

Table 13. Comparative analysis identity management system utilizing voice authentication.

Hidden neurons	TP	TN	FP	FN	Accuracy	Precision	Recall	Specificity	F-measure	G-mean	Execution time (s)
Deep breath	150	11	12	01	95.83	100.0	91.67	100.00	95.65	95.74	119.8567
Shallow breath	300	10	12	02	91.67	100.0	83.33	100.00	90.91	91.29	119.5190
All breath	250	22	19	41	89.13	84.62	95.65	82.61	89.8	89.96	132.7839

From this Table 12, Conclude that

- In the breath deep scenario, the voice samples that contain only breathing deep sounds.
- shallow breath scenario was performed by using voice samples that contain only shallow breathing sounds.
- In all breaths, the voice samples used contain both deep breathing and shallow breathing sounds

E. Comparative Analysis with Existing Techniques

To assessing the proposed approach's performance, the current voice authentication approaches are taken into consideration, including MFCC or Mel-frequency cepstral coefficients, HMM (Hidden Markov mode) , Gaussian Mixture Model (GMM) [178]. In this part, we can see how different performance metrics were used to compare the designed JHBO driven DNFN with data (training) and k-fold values shown in Table 2 and 3. Predicted from other method, a proposed method using DCNN kernel A DNFN K-fold fusion protein based on JHBO Verifying precision 0.9176, 0.9005, 0.8806, 0.8894. A sensitivity of 0.8982, 0.8816, 0.8938, an alpha of 0.9307 Findings with a high degree of certainty (0.9125, 0.8926, 0.9014, 0.9219) Information used for training Accuracy of testing: 0.8959, 0.8910, 0.8901, 0.9151 In terms of specificity, we have 0.902 0.8896 0.8932 0.9182 and a sensitivity of 0.9123 0.8868 0.9001 0.9218. The performance measures and the designed proposed method using JHBO-based DNFN are compared in Table 2. The results of a comparison study of JHBO-driven

 doi:10.20944/preprints202511.1266.v1

DNFN that were introduced for assessing accuracy with different values of k-fold. In terms of k-fold value, MFCC, HMM, and GMM all achieve testing accuracy of 0.8925, with MFCC coming in at 0.887, and the proposed JHBO-based DNFN reaching 0.9054.

A comparison of the sensitivity of JHBO-based DNFN with different k-fold values for k-fold value, MFCC, HMM, and GMM have sensitivity values of 0.8912, 0.878, and 0.8816, respectively; in contrast, created proposed method using JHBO-based DNFN has a sensitivity of 0.9085.

Table 5.4. Comparative Performance with Other Methods.

Metric	MFCC Based	HMM	GMM	Proposed
Accuracy	0.8956	0.8910	0.8910	0.8979
Sensitivity	0.885	0.8831	0.8813	0.9107
Specificity	0.9164	0.8963	0.9031	0.9221

Table 13. In a k-fold analysis, the proposed method continued to show superior performance.

Metric	MFCC	HMM	GMM	Proposed
Accuracy	0.9176	0.9005	0.8806	0.9394
Sensitivity	0.8982	0.8816	0.8938	0.9307
Specificity	0.9125	0.8926	0.9014	0.9312

V. Conclusion

The main objective is to secure Access control technique in X band by the proposed FSK/M-ary FSK technique for long-distance communication, which is not possible using conventional FSK. The encryption part has been done using the any modified Asymmetric key algorithm. Here, We are highlighting the scope of secure communication for high-frequency transmission. The bio-medical data can be transmitted to the cloud environment with the support of similar infrastructure.

In work, binary and M-ary FSK modulation techniques are extensively studied. The primary objective, the advantage of FSK modulation technique, factors influencing the choice of a particular digital modulation scheme.

References

1. Chakraborty M, Mallick, AK.AES Encrypted FSK Generation at X-Band Frequency using a Single Reflex Klystron. Wireless Communication over ZigBee for Automotive Inclination Measurement. China Communications. 2010,7: 1-9.
2. Pal P, Chandra Sahana B, Poray J. RSA encrypted FSK RF transmission powered by an innovative microwave technique for invulnerable security. The Journal of Defense Modeling and Simulation. August 2022. doi:10.1177/15485129211031670.
3. Pal, P., Sahana, B.C., Poray, J. (2025). MORSA: An innovative modified and optimized RSA framework using parallel computing environment. Mathematical Modelling of Engineering Problems, Vol. 12, No. 6, pp. 2085-2096. https://doi.org/10.18280/mmep.120624
4. A. Reynolds, "Voice biometrics: Feature extraction methods," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 5, pp. 1132–1140, 2019.
5. Y. Zhao et al., "Deep learning for speaker recognition," *IEEE Signal Process. Lett.*, vol. 26, no. 10, pp. 1551–1555, 2020.

 doi:10.20944/preprints202511.1266.v1

6. T. N. Sainath et al., "Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks," *ICASSP*, 2015.
7. N. Koblitz, "Cryptographic methods for biometric data protection," *Cryptologia*, vol. 30, no. 1, pp. 52–61, 2016.
8. P. K. Bhatia et al., "Multi-prime RSA and its performance evaluation," *IJCNIS*, vol. 3, no. 4, pp. 23–29, 2019.
9. J. Proakis, *Digital Communications*, 5th ed., McGraw-Hill, 2007.
10. S. Kumar and R. Jain, "FSK modulation for robust audio data transmission," *IEEE Comm. Lett.*, vol. 24, no. 3, pp. 540–544, 2020.
11. T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," Speech Communication, vol. 52, no. 1, pp. 12–40, 2020.
12. L. R. Rabiner and B. H. Juang, Fundamentals of Speech Recognition, Prentice Hall, 1993.
13. A. Menezes, P. van Oorschot, and S. Vanstone, Handbook of Applied Cryptography, CRC Press, 1996.
14. P. K. Bhatia et al., "Multi-prime RSA and its performance evaluation," IJCNIS, vol. 3, no. 4, pp. 23–29, 2011.
15. M. L. Sisodia and V. L. Gupta, Microwave Engineering, New Age International, 2007.
16. J. G. Proakis, Digital Communications, 5th ed., McGraw-Hill, 2007.
17. A. Jain et al., "Biometric template security," EURASIP Journal on Advances in Signal Processing, vol. 2008, pp. 1–17, 2008.
18. M. Crosby et al., "Blockchain technology: Beyond bitcoin," Applied Innovation Review, no. 2, pp. 6–10, 2022.
19. Dar JA, Srivastava KK, Ahmed Lone S. Design and development of hybrid optimization enabled deep learning model for COVID-19 detection with comparative analysis with DCNN, BIAT-GRU, XGBoost. Comput Biol Med. 2022 Nov;150:106123. doi: 10.1016/j.compbiomed.2022.106123. Epub 2022 Oct 3.
20. Yuechi Jiang, Frank H.F. Leung, Investigating and improving the utility of probabilistic linear discriminant analysis for acoustic signal classification, Digital Signal Processing, Volume 114, 2021, 103055, ISSN 1051-2004, https://doi.org/10.1016/j.dsp.2021.103055.