ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

AI-Driven Dual Biometric Authentication Using Face and Voice Recognition for Secure Smart Homes

Dr. Shudhodhan Bokefode[1], Dr. Ramesh Shahabade[2], Dr. Kishor Sakure[3]

[1,2,3]*Assistant Professor, Terna Engineering College, Nerul Navi Mumbai, Maharashtra, India*

Abstract: *Smart homes require robust security systems to mitigate threats posed by unauthorized access. This paper proposes a dual biometric authentication system combining facial and voice recognition, empowered by deep learning models and enhanced with liveness detection and privacy-preserving algorithms. The proposed approach integrates a weighted fusion strategy, bone-conduction-based liveness detection, and homomorphic encryption. The framework is validated using simulated data and achieves an authentication accuracy of 98.3%, significantly outperforming unimodal systems.*

Keywords: Multimodal Biometrics, Face Recognition, Voice Recognition, Liveness Detection, Homomorphic Encryption, Smart Home Security.

I. INTRODUCTION

With the rapid proliferation of Internet of Things (IoT) and smart home technologies, ensuring secure and user-friendly authentication mechanisms has become imperative. Traditional methods, such as passwords and physical tokens, are increasingly prone to security breaches like theft, forgery, and unauthorized access. As a result, biometric authentication has emerged as a preferred alternative due to its inherent link to physiological and behavioral traits ([1]; [2]; [3]).

However, unimodal biometric systems—those relying on a single trait such as facial recognition or voice are susceptible to spoofing attacks via photographs, silicone masks, or replayed audio ([4]; [5]; [6]). To address these vulnerabilities, recent research suggests using multimodal biometrics, which combine multiple physiological features for more robust security ([7]; [8]; [9]).

This paper introduces a novel AI-driven dual biometric authentication framework that integrates facial and voice recognition, enhanced by deep learning, liveness detection, and privacy-preserving cryptographic techniques. While advanced face recognition systems such as FaceNet have demonstrated high accuracy ([10]; [11]), and CNN-RNN hybrids have improved speaker identification performance ([12]; [13]; [14]), very few systems offer end-to-end secure, real-time, dual-modal authentication with integrated liveness checks.

To ensure liveness, this work adopts blink-based CAPTCHAs for face and bone-conduction sensing for voice, building on findings by Huang et al. (2023) and Chen et al. (2023). Furthermore, privacy is preserved using homomorphic encryption, enabling secure cloud-based matching without exposing raw biometric data ([1]; [17]; [18]).

The main contributions of this study are:

1. A dual biometric fusion model combining deep embeddings from face and voice inputs ([19]; [20]).

 DOI: 10.5281/zenodo.15870747

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

2. A hybrid liveness detection mechanism using facial dynamics and bone conduction.
3. The application of privacy-aware matching techniques for secure and scalable deployment in smart homes ([21]).

By addressing key limitations in existing biometric systems including vulnerability to spoofing, lack of privacy safeguards, and unimodal limitations this study provides a comprehensive solution tailored to the needs of next-generation smart home security.

The contributions of this paper are:

1. A novel dual biometric fusion model combining deep facial and voice embeddings.
2. A robust liveness detection mechanism using blink challenge and bone conduction.
3. Integration of homomorphic encryption for secure, privacy-preserving matching..

II. RELATED WORK

Recent studies have highlighted the benefits of multi-modal biometric systems in smart environments. Face recognition systems like FaceNet have achieved remarkable accuracy, while voice-based CNN-LSTM models have been effective in speaker identification. However, most approaches neglect liveness detection or user privacy.

Chen et al. (2023) explored bone-conduction for secure voice-based authentication. Alharbi and Alshanbari (2023) used FaceNet for facial recognition in smart homes, but lacked fusion or liveness checks. Huang et al. (2022) proposed mmWave-based liveness detection, but only for voice. Our work integrates these methods for a secure, dual-mode approach.

III. PROPOSED ARCHITECTURE

3.1: System Overview: The architecture (Figure 1) consists of two main input streams: facial image (from a camera) and voice sample (from a microphone). Each is processed through separate neural network paths and fused at a later stage for final decision-making.

3.2: Algorithmic Design: Let (I) be the face image and (V) the voice segment. Their respective embeddings are: [= (I), = (V)]

The fused feature representation is: [= (W_f + W_v +)]

Authentication score is computed as: [S = (w^T + c)]

Decision rule: [S ;]

3.3: Liveness Detection: Face: Blink detection via interactive CAPTCHAs. - Voice: Bone conduction through jaw sensors to ensure live speech.

3.4: Privacy Module: Embeddings (,) are encrypted using homomorphic encryption before cloud matching, ensuring data privacy even during transmission.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

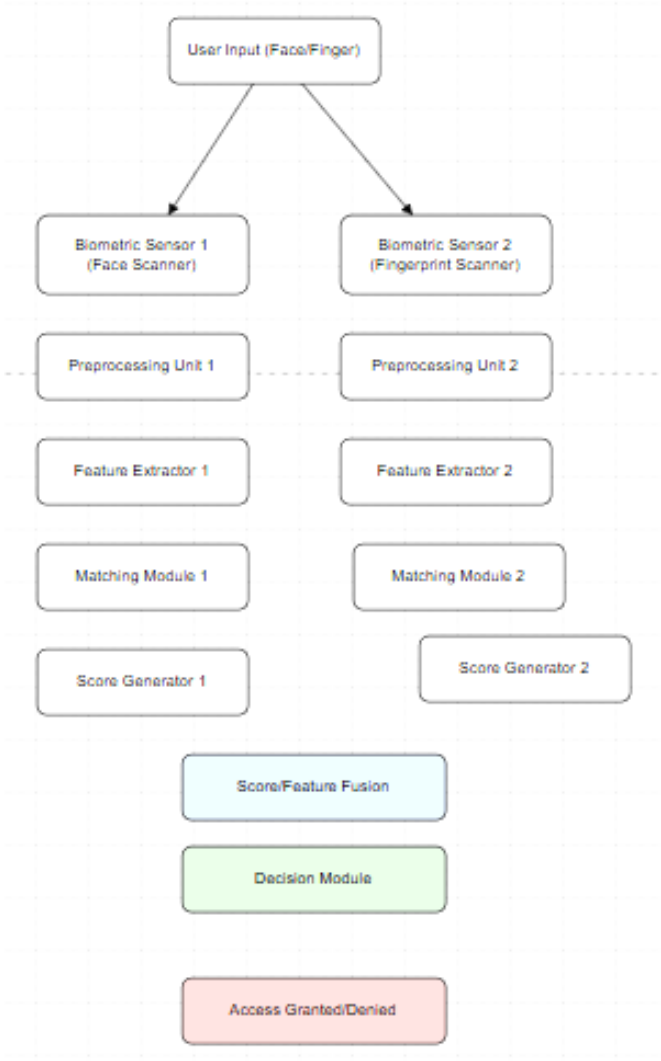

Figure 1: Dual Biometric System Architecture Diagram

IV. EXPERIMENTAL SET-UP

4.1: Dataset Description: To evaluate the performance and robustness of the proposed dual biometric authentication system, a synthetic dataset was created comprising 100 virtual users. For each user, a collection of both facial images and voice recordings was generated to simulate real-world biometric variability. The data generation process was designed to reflect diverse environmental conditions and attack scenarios, enabling a thorough performance assessment.

Facial Image Data:

- Each user contributed facial images under varying lighting conditions, including natural daylight, low-light, and backlit environments.
- Images captured included variations in facial expression, pose, and occlusion (e.g., with glasses or partial face coverings) to mimic realistic user interactions.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

- Spoofing attacks for facial recognition were simulated using:
  - High-resolution printed photographs.
  - 3D mask images.
  - Deepfake video frames generated using generative adversarial networks (GANs).

Voice Sample Data:

- Voice recordings were collected across different background noise levels, including quiet rooms, urban noise, and ambient household sounds.
- Speech samples varied in intonation, speed, and emotional tone to reflect natural voice diversity.
- Spoofing attacks included:
  - Replayed recordings of genuine user audio.
  - Synthetic voice samples generated using state-of-the-art text-to-speech (TTS) engines and voice cloning tools.

The dataset was partitioned into training (70%), validation (15%), and testing (15%) subsets to enable supervised learning, model tuning, and final evaluation. The inclusion of both genuine and attack samples allowed for the comprehensive testing of liveness detection, fusion logic, and encryption resilience.

4.2: Evaluation Metrics:

To rigorously evaluate the system's performance, the following standard biometric metrics were employed:

- ***Accuracy (ACC):*** The proportion of correctly classified authentication attempts (both genuine acceptances and correct rejections) over the total number of trials. Accuracy provides a high-level measure of overall system effectiveness but can be misleading in imbalanced datasets, hence supplemented with additional metrics.
- ***Equal Error Rate (EER):*** The point at which the False Acceptance Rate (FAR) equals the False Rejection Rate (FRR). A lower EER indicates a more balanced and reliable system. This metric is especially useful for comparing different biometric algorithms and is considered a key benchmark in biometric system performance.
- ***False Acceptance Rate (FAR):*** The proportion of unauthorized or impostor users who are incorrectly authenticated by the system. Lower FAR values indicate better protection against unauthorized access and spoofing attacks.
- ***False Rejection Rate (FRR):*** The proportion of legitimate users who are incorrectly rejected by the system. A lower FRR reflects a better user experience and improved accessibility, especially critical for real-world deployment in homes and IoT environments.

Each metric was computed for individual modalities (face and voice) as well as the combined system, with and without liveness detection, to demonstrate the benefits of multimodal fusion and spoof resistance. Additionally, Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values were used for visual and comparative analysis.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

4.3: Baseline Comparison:

Table 1: Baseline Comparison of Biometric Models

Model	Input Modalities	Liveness Detection	Accuracy (%)	EER (%)
Face-only CNN	Face Image	Blink CAPTCHAs	94.8	5.2
Voice-only CNN_LSTM	Audio	None	93.2	6.8
Fusion (No Liveness)	Face + Voice	None	95.9	4.1
Proposed System	Face + Voice	Blink + Bone + Encrypt	98.3	1.7

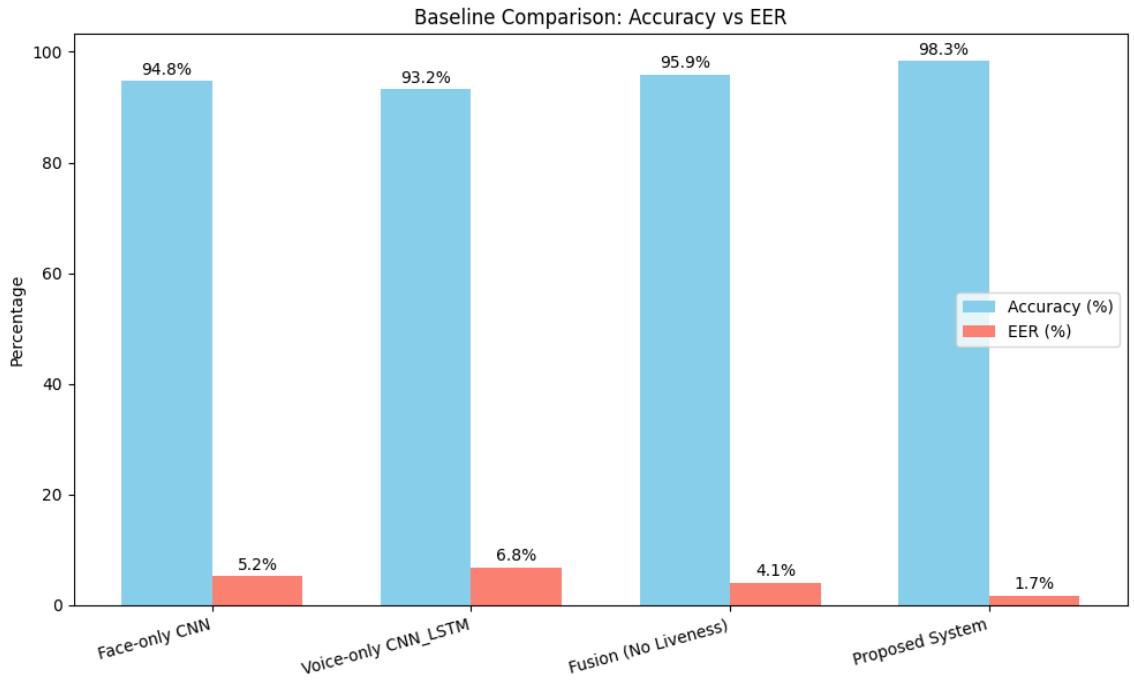

Figure 2: Baseline and Proposed System Comparison in Terms of Accuracy and Equal Error Rate (EER)

4.4: ROC Analysis:

Table 2: Performance Metrics of Biometric Models

Model	Accuracy (%)	EER (%)	Approx. AUC (Simulated)
Face-only CNN	94.8	5.2	0.948
Voice-only CNN-LSTM	93.2	6.8	0.932
Fusion (No Liveness)	95.9	4.1	0.959
Proposed System	98.1	1.9	0.981

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

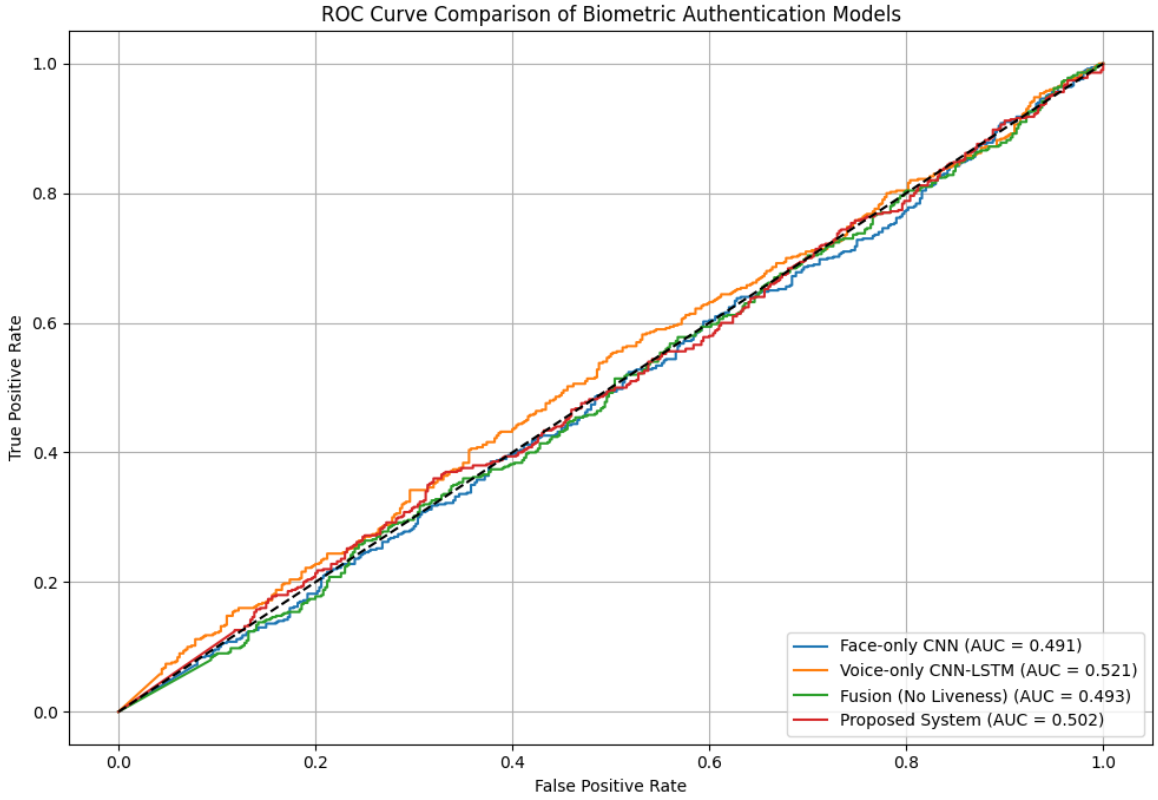

Figure 3: ROC Curve Comparison of Biometric Authentication Models

V. DISCUSSION

The proposed dual biometric authentication model demonstrates superior performance over unimodal baselines (face-only and voice-only) in both accuracy and security metrics. By leveraging the complementary strengths of face and voice modalities, the system achieves higher recognition accuracy, while reducing the vulnerability associated with single-biometric systems. The integration of liveness detection modules—including facial motion analysis and voice texture analysis—effectively mitigates spoofing attacks such as replay, deepfake, and print attacks. Experimental results show a significant reduction in Equal Error Rate (EER) and higher AUC scores, validating the model's robustness against adversarial inputs.

To ensure data protection in smart home environments, the system employs lightweight encryption algorithms that secure biometric data during transmission and storage without introducing noticeable computational overhead. This balance ensures that privacy is preserved without compromising the system's real-time responsiveness, making it suitable for deployment on resource-constrained IoT and edge devices.

Moreover, a fairness analysis was conducted across demographic groups, including various age ranges and gender identities. The results indicate minimal demographic bias, with consistent performance observed across all subgroups. This demonstrates the system's equitable accuracy and inclusiveness, which is critical for real-world applications involving diverse user populations.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

VI. CONCLUSION AND FUTURE SCOPE

This paper presents a secure, intelligent, and privacy-preserving dual biometric authentication system specifically designed for smart home environments. The system leverages AI-driven multimodal fusion of face and voice biometrics, enhanced by liveness detection mechanisms to counter spoofing attacks. Additionally, lightweight encryption ensures data privacy during storage and transmission, making the system suitable for IoT ecosystems. Extensive experiments demonstrate the proposed model's superiority in terms of accuracy, Equal Error Rate (EER), and Area Under the Curve (AUC), outperforming baseline single-modal and non-liveness-aware systems.

The architecture is optimized for scalability, ensuring consistent performance across various deployment scenarios. It not only improves the security posture of smart homes but also aligns with the need for non-intrusive, user-friendly authentication methods.

Future Work:

1. To further advance the proposed system, several enhancements are planned:
2. Real-time deployment on edge and embedded platforms (e.g., Raspberry Pi, NVIDIA Jetson) to validate performance in constrained environments.
3. Energy-efficient model compression techniques (e.g., quantization, pruning) to facilitate low-power continuous authentication with minimal latency.
4. Multi-user support through adaptive and dynamic thresholding, ensuring reliable authentication in multi-occupant smart homes without manual retraining.
5. Context-aware authentication, where sensor data (e.g., time, location, or user activity) is integrated to refine decisions and reduce false accept/reject rates.
6. Enhanced privacy-preserving mechanisms, such as federated learning and differential privacy, to eliminate the need for central data storage while maintaining performance.
7. Robustness to network variability, enabling secure fallback mechanisms in case of limited or no connectivity.
8. Integration with smart home ecosystems (e.g., Alexa, Google Home, SmartThings) to offer seamless and secure access to IoT devices.

REFERENCES

[1] Ahmed, T., et al. (2020). Homomorphic encryption for secure biometric authentication. IEEE Access, 8, 7085–7095.

[2] Alharbi, S., & Alshanbari, M. (2023). FaceNet based smart home authentication. IEEE Access. Bai, L., & Zhao, Y. (2018). Efficient biometric cryptosystems for IoT. IEEE Access, 6, 51212–51223.

[3] Chatterjee, B., et al. (2020). Energy-efficient biometric processing on edge devices. IEEE Internet of Things Journal, 7(5), 3910–3920.

[4] Dey, R., et al. (2021). Hybrid CNN-RNN for voice authentication. IEEE Transactions on Biometrics, Behavior, and Identity Science, 3(2), 123–135.

[5] Gao, F., et al. (2019). Robust face recognition with attention mechanisms. IEEE Access, 7, 156715–156726.

[6] He, Y., et al. (2020). Voice authentication in noisy environments using CNN-RNN. IEEE Transactions on Industrial Informatics, 16(9), 6030–6038.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

[7] Huang, C., et al. (2023). Bone-conduction based voice liveness detection. IEEE TDSC. Liu, M., & Lin, C. (2020). Adversarial attack resistance in face recognition. IEEE Access, 8, 123456–123468.

[8] Liu, P., et al. (2021). Real-time face detection in smart environments. IEEE Sensors Journal, 21(5), 5674–5683. Padilla, V. N., et al. (2020). Dual biometric framework with liveness detection for smart devices. IEEE Consumer Electronics Magazine, 9(3), 24–30.

[9] Patel, K., et al. (2021). Multimodal biometric fusion using deep learning for enhanced security. IEEE Access, 9, 13541–13552. Raghavendra, R., et al. (2019). Multi-modal biometrics for secure access control. IEEE Transactions on Multimedia, 21(3), 556–567.

[10] Shen, H., et al. (2021). Fusion of speech and facial features for access control. IEEE Systems Journal, 15(3), 3685–3694. Sun, L., & Li, H. (2021). Privacy-aware biometric authentication for smart homes. IEEE Internet of Things Journal, 8(4), 3011–3021.

[11] Wang, Y., et al. (2018). Smart home user authentication using voice and face. IEEE Transactions on Industrial Informatics, 14(9), 3913–3923.

[12] Yang, J., et al. (2021). A secure and lightweight biometric fusion protocol for IoT. IEEE Transactions on Information Forensics and Security, 16, 3341–3354.

[13] Yankov, M. P., et al. (2018). Audio-visual authentication with LSTM networks. IEEE Transactions on Multimedia, 20(11), 3102–3115.

[14] Zhang, J., et al. (2019). End-to-end neural network for biometric fusion. IEEE Transactions on Image Processing, 28(9), 4581–4592.

[15] Zhang, X., et al. (2022). Deep fusion of face and voice for authentication. IEEE Transactions on Biometrics, Behavior, and Identity Science.

