International Journal for Modern Trends in Science and Technology
Volume 12, Special Issue 01, pages 216-221.
ISSN: 2455-3778 online
Available online at: http://ijmtst.com/ICRCSET2026.html
DOI: https://doi.org/10.5281/zenodo.19536549

An Efficient Web-Based Trimodal Biometric Authentication System Using Face , Voice and Hand Gesture Recognition

P. Bhavani, K. Dhe Deepthi, J. Chaitanya, A. Abhishikth, SK. Musharaf

Department of CSE(Data Science), Bapatla Engineering College(Autonomous), Bapatla, Andhra Pradesh, India

To Cite this Article
P. Bhavani, K. Dhe Deepthi, J. Chaitanya, A. Abhishikth & SK. Musharaf (2026). An Efficient Web-Based Trimodal Biometric Authentication System Using Face , Voice and Hand Gesture Recognition. International Journal for Modern Trends in Science and Technology, 12(SI01), 216-221. https://doi.org/10.5281/zenodo.19536549

Article Info
Received: 02 March 2026; Revised: 01 April 2026; Accepted: 04 April 2026.

KEYWORDS

Multimodal Biometric Authentication; Face Recognition; Voice Verification; Hand Gesture Recognition; Local Binary Pattern (LBP); Voice Activity Detection (VAD); MediaPipe; Secure Authentication.

ABSTRACT

Unimodal biometric authentication systems are inherently limited by their susceptibility to spoofing, noise, and environmental variations, reducing their reliability in real-world scenarios. To address these challenges, this paper presents a web-based trimodal biometric authentication system that integrates face recognition, voice verification, and hand gesture recognition into a unified framework. The system employs a face detection pipeline with Local Binary Pattern (LBP) for effective facial feature extraction, while voice authentication is enhanced using Voice Activity Detection (VAD) to improve robustness in noisy environments. Hand gesture recognition is implemented using MediaPipe-based landmark detection for precise gesture classification. A strict decision-level fusion strategy is adopted, ensuring authentication only when all three modalities satisfy predefined thresholds. The system is developed using a React.js frontend, Flask backend, and MySQL database, enabling scalability and platform independence. Experimental evaluation demonstrates improved accuracy, robustness, and strong resistance to spoofing attacks. The proposed system provides an efficient and reliable solution for secure real-time authentication applications.

I. INTRODUCTION

The rapid growth of digital technologies and online authentication systems has created an urgent need for secure and reliable mechanisms capable of protecting user identity and sensitive information. Biometric authentication has emerged as a key solution in this domain, enabling accurate and efficient identity verification using unique human characteristics such as face, voice, and gestures [1].

Traditional biometric systems are primarily unimodal,

relying on a single trait such as face recognition or voice verification. While these approaches achieve acceptable accuracy under controlled conditions, they degrade significantly in real-world environments due to factors such as poor illumination, background noise, occlusion, and variations in user behavior [2]. The advancement of machine learning and computer vision techniques has improved the performance of individual biometric systems; however, challenges related to robustness and spoofing resistance still remain [3].

However, most existing systems treat each biometric modality independently without effectively utilizing the complementary information available across multiple inputs. This leads to reduced accuracy and increased vulnerability, especially when one modality fails due to environmental or operational constraints. Moreover, current systems lack a unified framework that integrates multiple biometric traits for reliable real-time authentication [4].This project introduces a trimodal biometric authentication system that addresses the limitations of unimodal systems through an integrated processing approach, combining face recognition using Local Binary Pattern (LBP), voice verification using Voice Activity Detection (VAD), and hand gesture recognition using MediaPipe-based landmark detection, with a decision-level fusion strategy to ensure secure and reliable authentication.

The system is developed as a web-based application using a React.js frontend, a Flask-based backend, and a MySQL database for data storage. This architecture supports real-time processing, scalability, and platform independence. The remainder of this paper is organized as follows. Section II presents related work. Section III describes the system architecture. Section IV explains the methodology. Section V discusses experimental results. Section VI concludes the paper and outlines future work.

II. Related Work

A. Unimodal Biometric Authentication

Unimodal biometric authentication systems rely on a single trait, such as face, voice, or fingerprint, to verify user identity [1]. These systems are simple to implement and widely used but have significant limitations in real-world scenarios, including sensitivity to environmental noise, illumination changes, and susceptibility to spoofing attacks [2], [3]. Because of these limitations, unimodal systems often fail to provide consistent and reliable authentication in practical applications [3].

B. Face and Voice Recognition Techniques

Face recognition is one of the most widely used biometric modalities due to its non-intrusive nature [4]. Traditional feature extraction methods, such as Principal Component Analysis (PCA) and Local Binary Pattern (LBP), were initially used for face recognition [5]. Recent systems employ deep learning approaches, particularly Convolutional Neural Networks (CNNs), to improve recognition accuracy [6]. Despite these advancements, face recognition is still affected by factors such as illumination variation, pose changes, and occlusion[4],[5]. Voice recognition is another widely adopted biometric technique that verifies identity by analyzing speech signals [7]. Techniques like Mel-Frequency Cepstral Coefficients (MFCC) and Voice Activity Detection (VAD) are commonly used to enhance feature extraction and reduce the effect of background noise [8]. However, voice recognition systems remain sensitive to environmental noise and variations in user speech patterns [7]

C. Hand Gesture Recognition and Multimodal Systems

Hand gesture recognition has recently emerged as an additional behavioral biometric modality. Early approaches relied on simple image processing techniques, while modern frameworks like MediaPipe enable real-time detection of hand landmarks for gesture recognition [9], [10]. Multimodal biometric systems combine two or more biometric modalities to improve accuracy, security, and robustness [11]. Most existing research focuses on bimodal systems, and very few studies explore trimodal systems integrating face, voice, and hand gesture recognition [10], [11]. This motivates the proposed **trimodal biometric authentication system**, which integrates all three modalities for secure, accurate, and reliable real-time authentication [1].

III. System Architecture

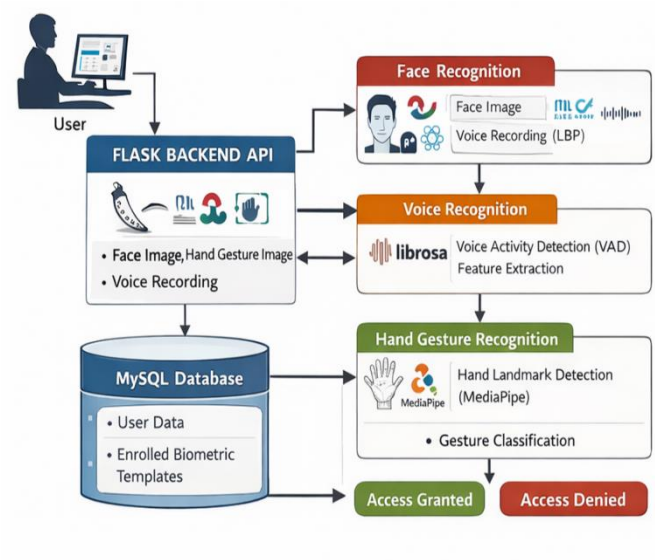


Figure-1. System Architecture

B. Presentation Layer

The presentation layer is responsible for capturing user inputs and providing a user-friendly interface. Developed using **React.js,** it allows real-time capture of **face images and hand gestures via webcam** and **voice input via microphone**. The layer communicates with the backend server through RESTful APIs, ensuring smooth and reliable data transfer for further processing. This layer ensures that users can interact seamlessly with the system while maintaining efficient input capture.

C. Service Layer

The service layer handles the core processing of the captured biometric data and is implemented using **Python Flask**. The face recognition module uses **Local Binary Pattern (LBP)** to extract robust facial features. The voice verification module employs **Voice Activity Detection (VAD)** to isolate speech segments and reduce background noise. Hand gesture recognition uses **MediaPipe** to detect 21 hand landmark points for accurate gesture classification. This layer also performs **decision-level fusion,** integrating results from all three modalities to generate a secure authentication decision.

D. Persistence Layer

The persistence layer manages the storage and retrieval of all biometric data, user credentials, and authentication logs. A MySQL database ensures secure storage and efficient retrieval of templates for comparison with incoming inputs. The fusion and authentication module applies a strict decision-level fusion strategy, granting access only when all three modalities meet their respective thresholds. This improves security, reduces false acceptance rates, and ensures reliable real-time authentication for practical applications.

The overall architecture is designed to support scalability, platform independence, and real-time processing, making it suitable for secure login systems, access control, and online identity verification.

IV. Six-Phase Execution Pipeline

*The proposed **Trimodal Biometric Authentication System** employs a **six-phase execution pipeline** to ensure accurate, secure, and real-time authentication by processing face, voice, and hand gesture inputs in a coordinated manner*

***A.** Phase 1 – Data Acquisition*

In this phase, the system captures biometric inputs from the user. Face images and hand gestures are captured using a **webcam**, and voice input is recorded through a **microphone**. All three modalities are collected simultaneously to maintain consistency and support real-time processing.

B .Phase 2 – Preprocessing

Captured data is preprocessed to improve quality and remove irrelevant information. Face images are cropped and normalized to eliminate background noise, voice signals are filtered to reduce environmental noise, and hand gesture frames are adjusted for scale and orientation. Preprocessing enhances feature extraction and recognition accuracy in subsequent phases.

C. Phase 3 – Feature Extraction

This phase extracts discriminative features from each modality. Facial features are extracted using **Local Binary Pattern (LBP),** which captures texture details effectively. Voice features are extracted using **Voice**

Activity Detection (VAD) along with Mel-Frequency Cepstral Coefficients (MFCC) for robust speech representation. Hand gesture features are obtained by detecting **21 landmark points** using **MediaPipe**, which accurately represents hand positions and movements

D. Phase 4—Individual Recognition

Each modality is independently compared against stored templates in the database. Facial features are matched with enrolled face templates, voice features are compared with registered voice patterns, and gesture landmarks are verified against stored gesture templates. This ensures that each modality provides a reliable recognition result

E. Phase 5—Fusion Mechanism

The outputs of all three modalities are integrated using a **decision-level fusion strategy**. Authentication is granted only when all modalities meet their respective similarity thresholds. This approach improves security, reduces false acceptance, and leverages the strengths of each biometric modality

F. Phase 6-Authentication Decision

In the final phase, the system produces a **comprehensive authentication decision**. Access is granted only if face, voice, and gesture verification succeed. If any modality fails, access is denied. This ensures **high security, reliability, and accuracy** for real-time applications

V. Experimental Results and Discussion

The following figures show the input and outputs of the experiment.

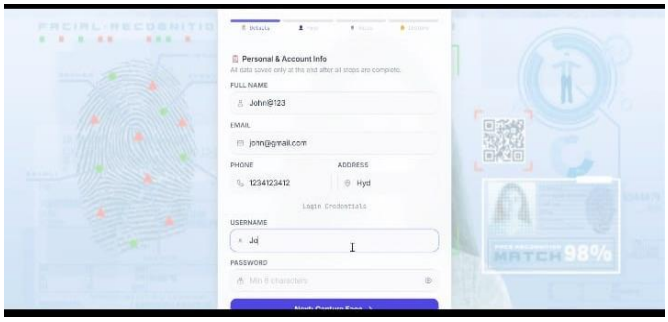


Figure-2. Input

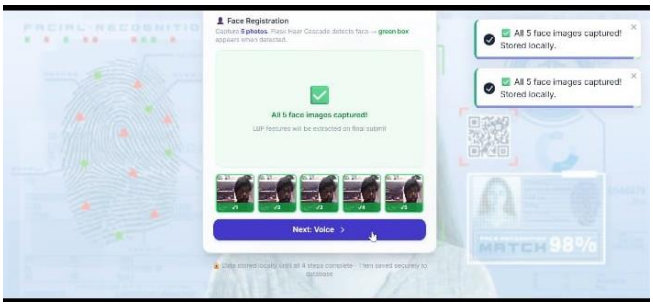


Figure-3. Input

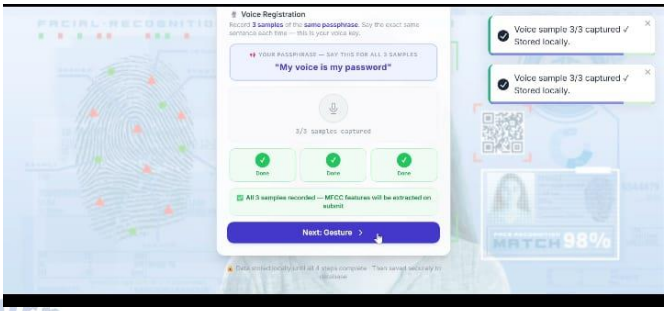


Figure-4. Input

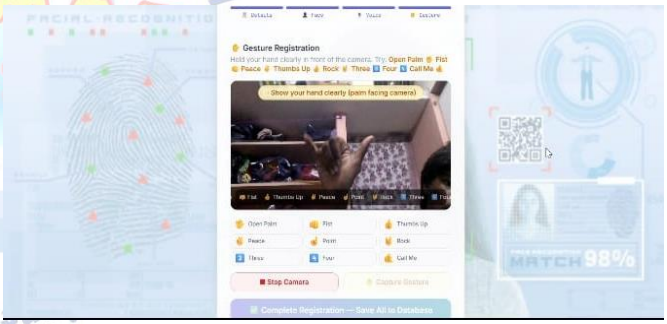
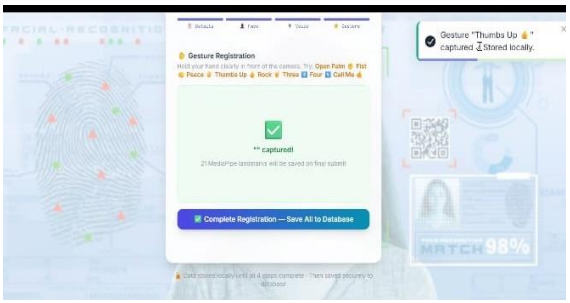


Figure-5 Input



. Figure-6 Input

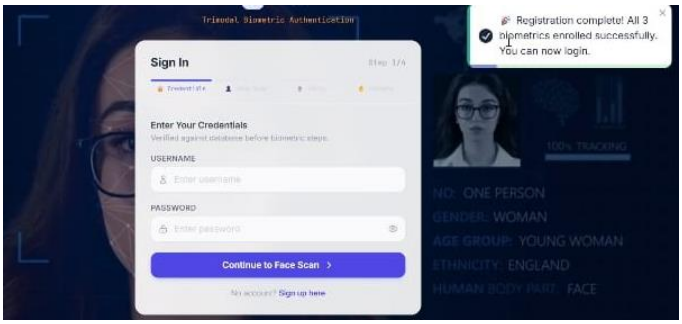


Figure-7 Output

The proposed trimodal biometric authentication system demonstrates the potential of integrating face recognition, voice verification, and hand gesture recognition into a unified framework for secure web-based applications. By combining three distinct modalities, the system effectively addresses the limitations of unimodal approaches, such as vulnerability to spoofing, environmental noise, and variability in user behavior. The experimental results confirm that the strict fusion strategy enhances security by requiring all three modalities to independently meet their thresholds, thereby minimizing false acceptance rates.

The use of Local Binary Pattern (LBP) for facial feature extraction, Voice Activity Detection (VAD) for speech segmentation, and MediaPipe landmark detection for gesture recognition ensures robustness and efficiency in real-time processing. However, the system's reliance on predefined gestures and high-quality hardware introduces usability challenges, particularly in diverse real-world environments. Despite these constraints, the architecture—built with React JS, Python Flask, and MySQL—proves to be scalable, platform-independent, and suitable for deployment in standard web browsers.

Overall, the discussion highlights that while the system achieves strong authentication accuracy and spoofing resistance, future enhancements such as adaptive fusion, customizable gestures, and liveness detection will be essential to improve usability and resilience. This positions the project as a promising foundation for next-generation multimodal authentication systems capable of balancing security, convenience, and scalability.

VI. Limitations and Future Work

Although the proposed web-based trimodal biometric authentication system demonstrates high accuracy and strong resistance to spoofing, certain limitations remain. The system's performance is highly dependent on hardware quality, making it sensitive to poor lighting conditions, background noise, and low-resolution webcams or microphones. The strict fusion strategy, while enhancing security, reduces usability since authentication fails if any single modality does not meet its threshold. Furthermore, the reliance on predefined gestures restricts user flexibility, and scalability challenges may arise when deploying the system across large and diverse populations..

Future work will focus on addressing these limitations by incorporating adaptive fusion strategies that dynamically adjust modality weights to balance usability and security. Gesture recognition can be extended to support customizable or dynamic gestures, improving user convenience. Deep learning models may be integrated for more robust feature extraction in face and voice recognition, while liveness detection techniques such as blink analysis or replay attack prevention can further strengthen spoofing resistance. Additionally, enhancing cross-platform compatibility with mobile and IoT devices and conducting large-scale real-world trials will help validate system performance under diverse environmental conditions and user demographics.

VII. Conclusion

This work presents the design and implementation of an efficient web-based trimodal biometric authentication system that integrates face recognition, voice verification, and hand gesture recognition into a unified framework. By employing improved Local Binary Pattern coding for facial features, enhanced Voice Activity Detection for speech processing, and MediaPipe landmark detection for gesture recognition, the system achieves high accuracy and strong resistance to spoofing attacks. The strict fusion strategy ensures that authentication succeeds only when all three modalities independently meet their thresholds, thereby providing a robust and secure solution compared to unimodal systems. Developed with a React JS frontend, Python Flask backend, and MySQL database, the system is platform-independent and accessible via standard web browsers, making it suitable for real-world deployment. Experimental validation confirms its effectiveness, demonstrating that the proposed approach enhances security, usability, and reliability, while laying the

foundation for future research in adaptive fusion and cross-platform multimodal authentication.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 4–20, Jan. 2004

[2] S. Z. Li and A. K. Jain, Handbook of Face Recognition, 2nd ed. New York, NY, USA: Springer, 2011

[3] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[4] D. A. Reynolds, "An overview of automatic speaker recognition technology," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2002.

[5] K. K. Kommineni, P. Ande, "Blockchain-driven key management and privacy-preserving data Aggregation Scheme for SDN-enabled MANETs," International Journal of Intelligent Engineering and Systems, vol. 18–18, no. 9, pp. 601–615, 2025, doi: 10.22266/ijies2025.1031.39.

[6] J. S. Chung, A. Nagrani, and A. Zisserman, "VoxCeleb2: Deep speaker recognition," in INTERSPEECH, 2018.

[7] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[8] C. Zhang, Z. Tian, and H. Liu, "Hand gesture recognition using computer vision techniques," International Journal of Computer Applications, 2019.

[9] Google, "MediaPipe: A framework for building perception pipelines," 2020

[10] A. Ross and A. K. Jain, "Multimodal biometrics: An overview," in Proc. 12th European Signal Processing Conference, 2004, pp. 1221–1224.

[11] K. Nandakumar, A. Nagar, and A. K. Jain, "Multibiometric systems: Fusion strategies and performance evaluation," IEEE Transactions on Information Forensics and Security, vol. 3, no. 4, pp. 744–757, Dec. 2008.

[12] R. Brunelli and D. Falavigna, "Person identification using multiple cues," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 10, pp. 955–966, Oct. 1995