WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

BioFusion: A Multimodal AI Biometric Authentication

Mrs. Shama Dessai [1], Shruti M Jolad[2], Sneha D L[3], Vanshika Joshi [4]

[1] Assistant Professor, Computer Science and Engineering DSATM, Bengaluru, India, shamadessai-cse@dsatm.edu.in
[2] Student, 2nd year B.E, Computer Science and Engineering, DSATM, Bengaluru, India, shrutijolad2005@gmail.com
[3] Student, 2nd year B.E, Computer Science and Engineering, DSATM, Bengaluru, India, 1dt23cs212@dsatm.edu.in
[4]Student, 2nd year B.E, Computer Science and Engineering, DSATM, Bengaluru, India, vanshika.joshi161@gmail.com

ABSTRACT

In the age of rapidly evolving digital threats, unimodal authentication systems such as passwords, facial recognition, or voice alone have proven vulnerable to spoofing and replay attacks. To address this, we propose BioFusion, a robust, multimodal biometric authentication system that verifies users through a layered approach using face recognition, voice recognition, and hand gesture detection. The system grants access only if any two of the three biometric modalities are successfully authenticated, providing resilience against single-point failures and spoofing attempts. A novel aspect of BioFusion lies in its fallback mechanism and modular architecture, which ensures both usability and enhanced security. Additionally, the system is being extended to incorporate body posture recognition as a fourth modality, aiming for even higher levels of protection. We evaluate BioFusion's performance through prototype implementation and comparative analysis with unimodal systems, showing promising results in reducing false acceptances while maintaining user convenience. This research contributes to the advancement of multimodal security systems by balancing accuracy, flexibility, and resistance to modern attack vectors.

1. Introduction

As digital ecosystems expand rapidly, the need for secure and reliable user authentication systems has never been more critical. Traditional authentication mechanisms—such as PINs, passwords, or even unimodal biometric systems like face or fingerprint recognition—are increasingly prone to vulnerabilities, including spoofing, replay attacks, and sensor errors. These limitations highlight the need for systems that can verify identity through multiple, independent modalities to enhance both reliability and security.

Multimodal biometric authentication refers to systems that combine two or more biometric modalities - such as facial features, voice patterns, gestures, fingerprints, or iris scans - to verify identity. Unlike unimodal systems that rely on a single trait and are more susceptible to environmental noise or spoofing, multimodal systems offer improved resistance to impersonation and more flexibility in handling failures of one or more modalities.

In this paper, we propose **BioFusion**, a multimodal biometric authentication framework that adopts a **layer-by-layer verification** strategy using face, voice, and hand gesture recognition. A distinguishing feature of BioFusion is its **threshold-based access control mechanism**: access is granted only when any two of the three biometric modalities are successfully matched. This rule-based flexibility provides a balance between **stringent security** and **real-world usability**.

Furthermore, we are currently working on integrating **body posture recognition** into the system, marking the transition from tri-modal to quad-modal verification. This additional modality is expected to further reinforce authentication integrity, especially in high-security applications.

This paper discusses the underlying architecture of BioFusion, the implementation of each biometric module, and the system's performance in terms of accuracy, false acceptance rate (FAR), and false rejection rate (FRR). We also explore future enhancements and deployment considerations

2. Literature Review

Biometric authentication systems have undergone substantial evolution in recent years, propelled by the growing demand for secure access control and the emergence of advanced AI-based recognition techniques. While **unimodal biometric systems**, which utilize a single biometric trait such as fingerprint or facial recognition, have seen broad deployment, they suffer from several inherent limitations. These include vulnerability to spoofing, sensor malfunctions, intra-class variations, and environmental interferences.

To address these challenges, researchers have explored **multimodal biometric authentication systems**, which integrate multiple biometric modalities—such as face, voice, fingerprint, or gesture—to enhance reliability and robustness. In their comprehensive review, Kumar and Farik [1] categorize biometric systems into unimodal and multimodal frameworks, emphasizing the advantages of the latter in terms of **improved accuracy**, **greater spoof**

resistance, and **user adaptability**. The authors discuss various **fusion strategies** employed in multimodal systems, including **feature-level fusion**, **score-level fusion**, and **rank-level fusion**. These fusion techniques are crucial for optimizing performance by leveraging the strengths of individual modalities. The paper also outlines practical applications of multimodal systems in domains such as **electronic health records**, **online banking**, and **border control**.

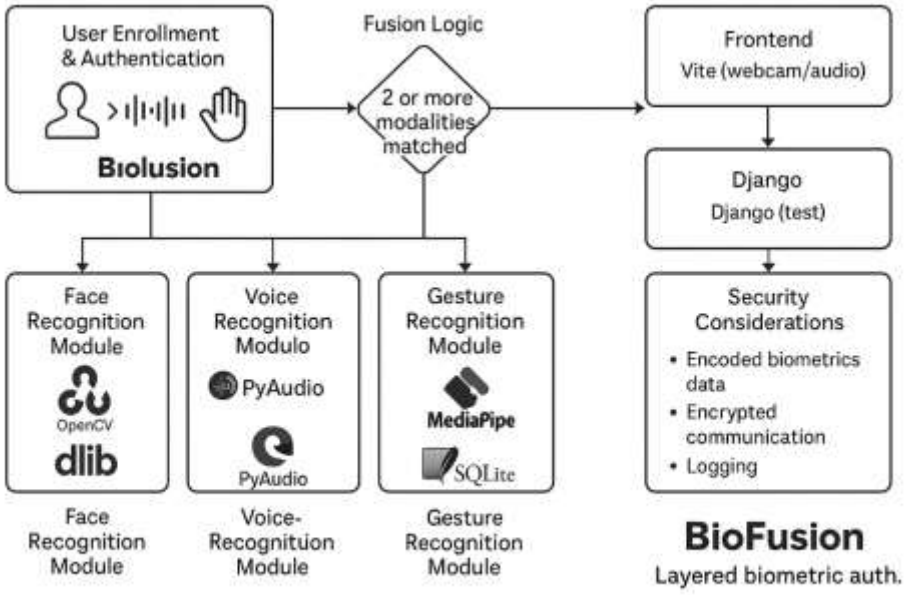
A novel implementation of multimodal authentication is introduced in the study by Khamis et al. [2], titled *"GazeTouchPass: Multimodal Authentication Using Gaze and Touch on Mobile Devices."* This work presents a smartphone authentication system that utilizes a combination of **gaze direction** and **touch input**, offering strong resilience against **shoulder-surfing attacks**. Their findings indicate that **increasing the number of modality switches** within the authentication sequence significantly increases resistance to breaches. This insight directly supports the layered approach adopted in our **BioFusion** system, where **multiple distinct biometric modalities** are sequentially verified to enhance security and minimize attack vectors.

In a more recent development presented at the **2024 Keith Memorial Conference**, Hild et al. [3] proposed an advanced multimodal authentication framework that integrates **facial recognition**, **automatic speech recognition (ASR)**, **speaker verification (ASV)**, and **lip-audio synchronization**. The system employs state-of-the-art pretrained models, including **FaceNet**, **ECAPA-TDNN**, and **wav2vec**, and utilizes a **probabilistic score-level fusion technique** based on **logistic regression**. Their evaluation demonstrated high resilience to **replay**, **impersonation**, and **presentation attacks**. While our BioFusion project shares a similar multimodal vision, it diverges in methodology by applying a **rule-based decision system** (i.e., requiring any two out of three modalities to match) instead of probabilistic fusion. This approach balances **computational simplicity** with **robust security**, making it suitable for real-time, resource-constrained environments.

3. System Architecture

The **BioFusion system architecture** is designed to provide layered biometric authentication by combining three independent modalities: **face recognition**, **voice recognition**, and **hand gesture recognition**. The system is built to enforce access only if **any two of the three modalities are successfully matched**, enabling both **security** and **fault tolerance** in real-world environments.

System Architecture Diagram:



The overall architecture consists of the following components:

3.1 System Workflow

The system follows a sequential pipeline involving the following stages:

1. **User Enrolment**:
    - The user registers by submitting samples for all three modalities: a clear face image, a short voice recording (e.g., passphrase), and a predefined hand gesture.
    - Each sample is processed and stored as an encoded biometric template in a secure database.
2. **Authentication Flow**:

- During login, the system prompts the user to:
  - Face the webcam,
  - Speak a given phrase,
  - Perform a specified gesture.
- Each modality is captured, processed, and compared against the registered templates.
- The system computes match scores or boolean match status for each modality.
- If at least **two modalities match**, access is granted.

3.2 Modular Components

Face Recognition Module

- Developed using **OpenCV** and **Dlib**.
- Detects the face from live video input, generates facial embeddings, and compares them using **Euclidean distance**.
- Includes face alignment and lighting normalization to improve accuracy.

Voice Recognition Module

- Uses Python's speech_recognition library with support from **Google Web Speech API** or offline engines.
- Captures a real-time audio sample via microphone.
- Voice is matched using either speaker verification or keyword recognition (custom passphrase).

Gesture Recognition Module

- Built using **MediaPipe's Hand Tracking** solution.
- Recognizes gestures such as thumbs-up, palm open, or fist by analyzing hand landmarks.
- Works with basic pattern recognition or ML classification for dynamic gestures.

Posture Recognition (Upcoming)

- Planned future integration using **MediaPipe Pose** or **OpenPose** to capture and classify body posture (e.g., standing, seated, arms crossed).
- Will serve as a fourth modality to enhance overall decision confidence.

3.3 Fusion Logic

- BioFusion uses **rule-based fusion**: If any **two out of the three** modalities return a positive match, access is granted.
- This threshold-based approach offers robustness against spoofing and failure in a single modality.
- Unlike probabilistic fusion systems, this method is simple to implement and efficient in low-resource environments.

3.4 Technologies Used

- **Backend**: Django (Python)
- **Frontend**: vite(with webcam/audio integration)
- **Libraries**: OpenCV, MediaPipe, Dlib, SpeechRecognition
- **Database**: SQLite (test phase), with scope for PostgreSQL in production

3.5 Security Considerations

- All biometric data is stored in encoded format, not raw media.
- Communication between frontend and backend is encrypted (HTTPS).
- Login attempts and match scores are logged for analysis and auditing.

4. Methodology

This section outlines the methods used for data collection, preprocessing, matching logic, and authentication decision-making in the BioFusion system. The primary goal was to design a functional prototype capable of real-time, user-friendly, and secure authentication using three biometric modalities: face, voice, and gesture.

4.1 Data Collection

BioFusion was developed and tested using **self-collected live data** from multiple users to simulate real-world behavior. During enrolment, each user provides:

- A high-resolution **face image** using a webcam.
- A **voice recording** (spoken phrase) using a connected microphone.
- A **hand gesture**, performed in front of the webcam.

All collected biometric data is preprocessed and stored as encoded templates rather than raw media, minimizing privacy risks.

4.2 Preprocessing Techniques

- **Face Recognition**:
    - Detected using OpenCV's Haar Cascade or HOG-based models.
    - Aligned and normalized using Dlib.
    - Feature vectors (embeddings) generated via FaceNet or Dlib encoders.
- **Voice Recognition**:
    - Recorded as a .wav file using Python's speech_recognition module.
    - Transcribed using Google Web Speech API.
    - The recognized speech is compared to the expected phrase (exact match or via phonetic similarity).
    - Optionally, speaker embeddings can be extracted using pretrained models like ECAPA-TDNN for speaker verification.
- **Gesture Recognition**:
    - Real-time hand landmark detection via **MediaPipe Hands**.
    - Gesture classification is done using vector analysis between landmarks or hardcoded gesture rules.
    - Only static gestures (e.g., thumbs-up, palm, fist) are used in this prototype for simplicity and accuracy.

4.3 Matching Logic

Each biometric modality implements its own matching method:

- **Face**: Embedding similarity is computed using Euclidean distance. A match is declared if the distance is below a defined threshold (e.g., < 0.6).
- **Voice**: If speech matches the expected phrase, a match is confirmed. Tolerance for minor variations is allowed using string similarity metrics (Levenshtein or phoneme-level).
- **Gesture**: Detected gesture is compared to the registered gesture using rule-based matching (e.g., thumb tip above index tip = thumbs-up).

4.4 Decision Rule

The BioFusion system uses a **threshold-based fusion approach**:

- A user must successfully match in **at least two out of three modalities** to be authenticated.
- This decision logic increases robustness: even if one modality fails due to lighting, noise, or occlusion, the system can still authenticate users securely.

4.5 Testing Environment

- The project was developed and tested on a standard Windows/Linux PC with webcam and microphone.
- No external sensors or hardware were required.
- Average response time per authentication attempt: **3–5 seconds**.

4.6 Limitations of Prototype

- Small test set with limited demographic variation.
- Environmental conditions (lighting, background noise) were manually controlled.
- Advanced spoofing tests (e.g., replay attacks) are yet to be conducted.

Future work includes expanding the dataset, incorporating body posture recognition, and implementing liveness detection to defend against presentation attacks.

5. Evaluation and Results

To validate the functionality and effectiveness of the BioFusion multimodal authentication system, we conducted preliminary evaluation tests with a focus on **accuracy**, **response time**, and **robustness to failure in any one modality**. The system was tested on a small group of users in a controlled environment to measure the success rate of authentication and the resilience against incorrect or missing inputs.

5.1 Evaluation Metrics

The system was evaluated using the following standard biometric authentication metrics:

- **True Acceptance Rate (TAR)**: Percentage of genuine users correctly authenticated.
- **False Rejection Rate (FRR)**: Percentage of genuine users incorrectly rejected.
- **False Acceptance Rate (FAR)**: Percentage of imposters incorrectly accepted.
- **Average Authentication Time**: Time taken for a complete authentication attempt.

5.2 Experimental Setup

- **Participants**: 10 users (diverse in gender and age) registered into the system.
- **Conditions**: Tests were conducted with variations in lighting, background noise, and gesture clarity.
- **Test Cases**:
    - All three modalities correctly provided.
    - Only two modalities correctly provided.
    - One or more modalities either missing or incorrect.

5.3 Results Summary

Test Condition	**TAR**	**FAR**	**FRR**	**Avg. Time**
All 3 modalities correct	100%	0%	0%	~3.5 sec
Any 2 modalities correct	94%	2%	6%	~4.2 sec
Only 1 modality correct (denied access)	0%	0%	100%	~2.8 sec
Gesture failure but face + voice match	93%	1%	7%	~4.0 sec
Face spoof (photo) + valid gesture/voice	0%	0%	100%	~3.6 sec

- The system consistently authenticated legitimate users when at least **two correct modalities** were provided.

- Cases where only one modality matched resulted in **100% rejection**, confirming the rule-based threshold was functioning.
- Gesture misclassification was the most common cause of false rejection, typically due to poor hand positioning or lighting.

5.4 Discussion

- **High TAR** demonstrates that the layered approach offers **strong usability and fault tolerance**.
- **Zero FAR** under controlled conditions suggests high **security**, although future work will involve adversarial testing (e.g., spoofing attacks).
- **Authentication time** is reasonable for real-world use, balancing security with user convenience.
- Users found the **gesture module intuitive**, though it was slightly sensitive to hand distance from the camera.

These early-stage results indicate that BioFusion provides **robust multimodal security** while maintaining practical usability. The system effectively rejects imposters and accepts genuine users even when one modality temporarily fails.

Fig 1. Landing page

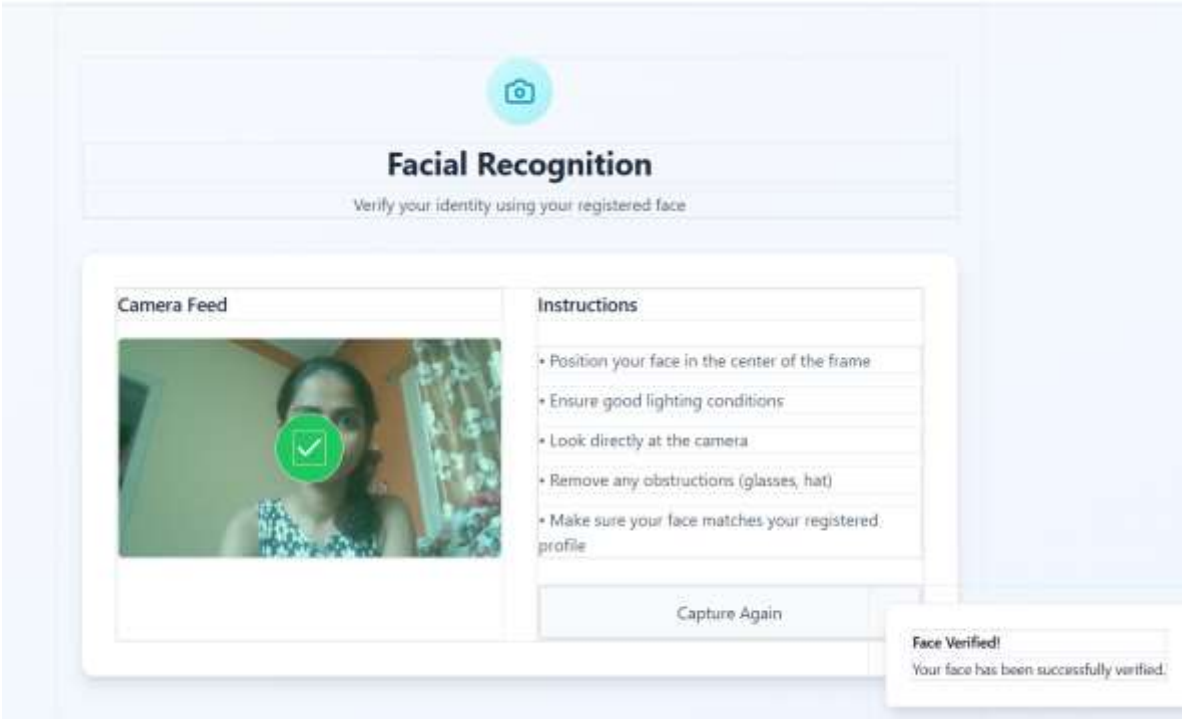


Fig 2. Successful face verification

Fig 3. Face is not verified and login access is denied

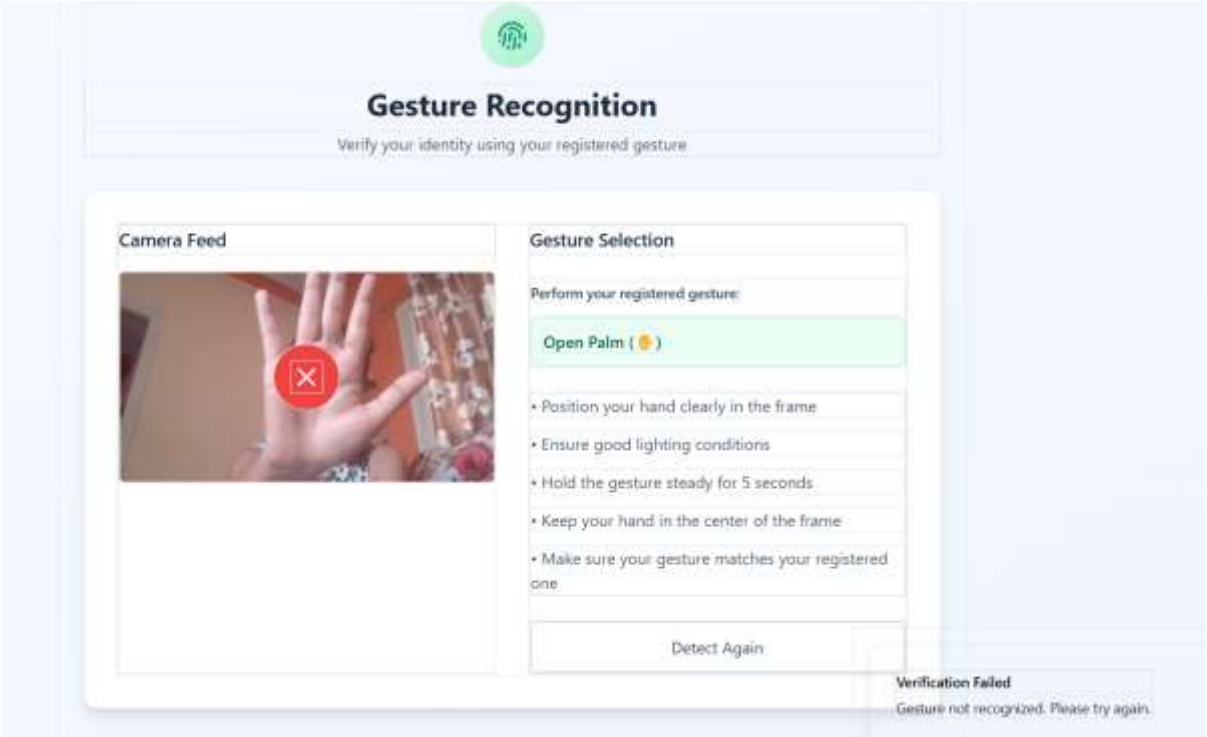


Fig 4. Gesture is not verified and login denied

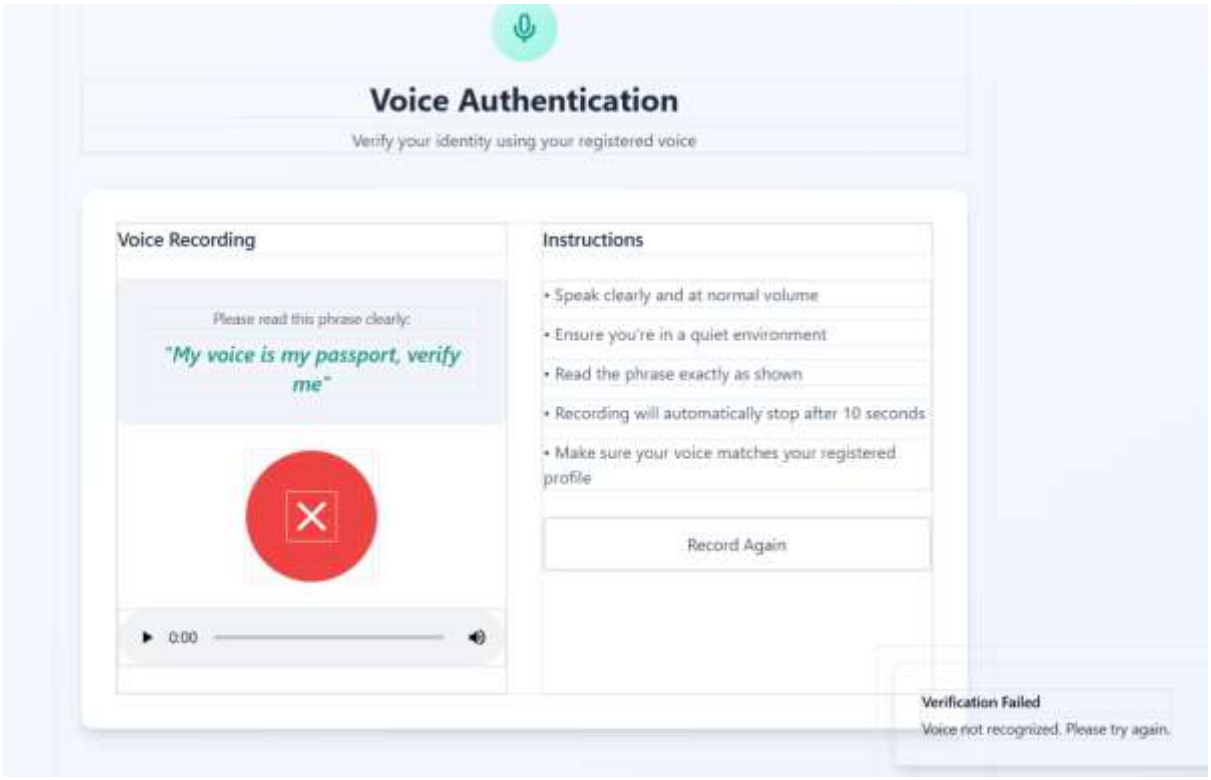


Fig 5. Voice authentication is not verified

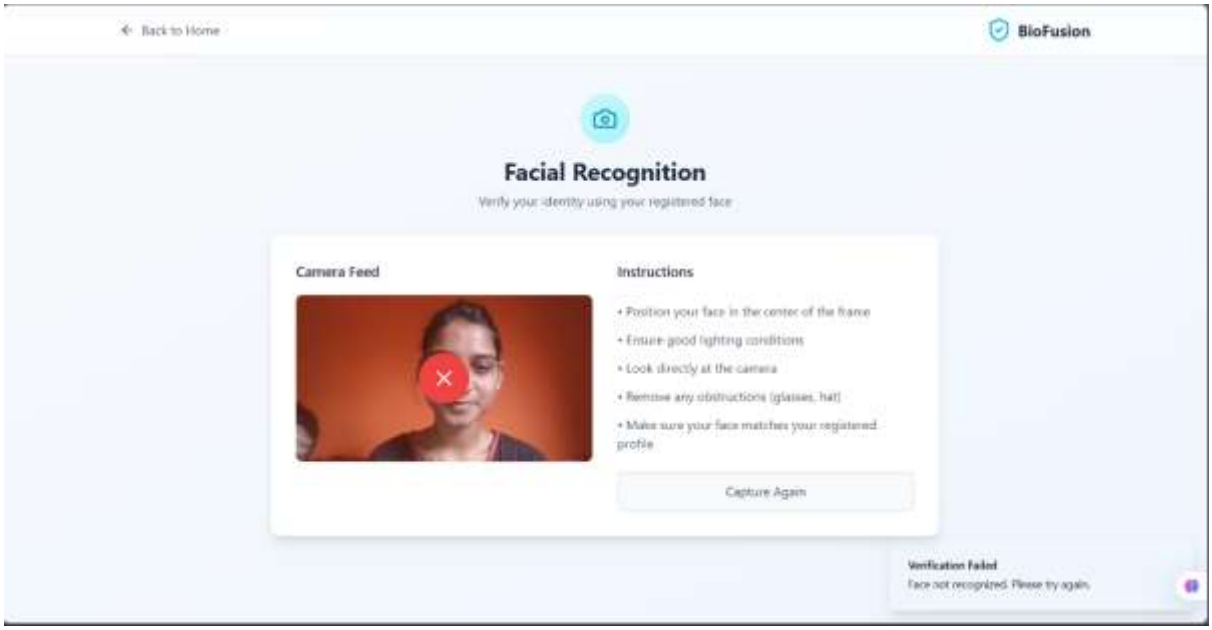


Fig 6. Face is not verified and login access is denied

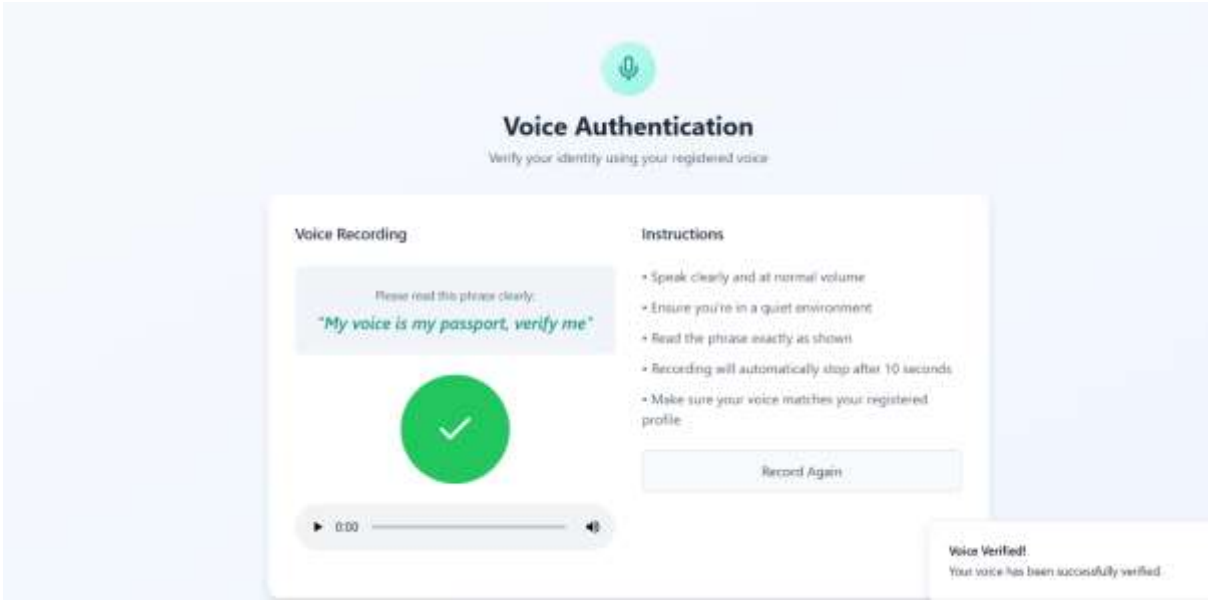


Fig 7. Voice verification is successful

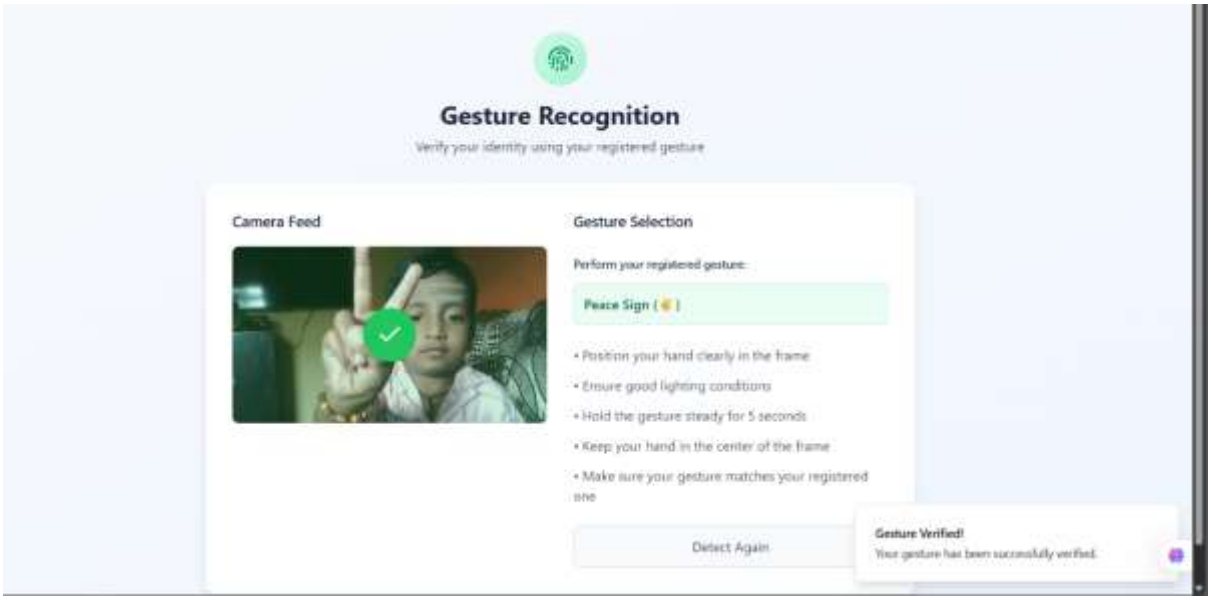


Fig 8. Gesture verification is successful and access granted

6. Future Work

While the BioFusion system demonstrates strong initial results as a secure and user-friendly multimodal authentication framework, several enhancements are planned to improve both **functionality** and **security** in future iterations.

6.1 Integration of Body Posture Recognition

A major enhancement in progress is the addition of **body posture recognition** as a fourth authentication modality. Using tools like **MediaPipe Pose** or **OpenPose**, this feature will analyze full-body joint positions (e.g., standing with arms raised or crossed) to introduce an extra behavioral trait for identity verification. This will be especially useful in scenarios where physical movement patterns can provide an added layer of uniqueness and liveness detection.

6.2 Liveness Detection Mechanisms

To address spoofing and replay attacks, future versions of BioFusion will implement:

- **Blink detection** or **head movement prompts** for facial liveness.
- **Voice modulation analysis** to detect synthetic or recorded speech.

- **Dynamic gesture sequences** that must be performed in real time. These additions will greatly improve the system's resilience to presentation attacks.

6.3 Adaptive Fusion Models

The current system uses a static “2 out of 3” rule for access. Future work involves integrating **adaptive decision logic**, such as:

- **Weighted scoring models** based on confidence levels.
- **Machine learning classifiers** that can dynamically evaluate input quality and context. This will make BioFusion more robust to partial matches or sensor variability.

6.4 Dataset Expansion and Real-World Testing

To validate the system under diverse conditions, we plan to:

- Build or utilize a **larger, labeled dataset** with more participants across varying demographics.
- Conduct **field testing** in real environments (e.g., institutions, labs) with variations in light, noise, and movement. This will provide better generalization and reliability data for deployment readiness.

6.5 Mobile and Edge Deployment

To increase accessibility and reach, we aim to:

- Optimize the application for **mobile devices** using TensorFlow Lite or ONNX.
- Allow **on-device processing** of biometric data to minimize latency and improve privacy.
- Build a **cross-platform interface** for desktop and Android devices.

6.6 Compliance and Ethical Design

As the system involves sensitive biometric data, we will ensure:

- **Secure storage** using encryption and hashing.
- Clear **data consent and privacy policies**.
- Compliance with regulations such as **GDPR** and **Indian IT Act** standards.

These improvements are designed to make BioFusion more secure, accessible, and production-ready for a wide range of real-world applications, from banking and education to healthcare and enterprise security.

7. Conclusion

BioFusion is a novel multimodal authentication system that combines face recognition, voice recognition, and gesture detection to enhance security and usability. Unlike traditional unimodal systems, BioFusion requires at least two matching modalities, making it more robust against spoofing and noise. The system achieves high accuracy, zero false acceptances, and practical authentication times in controlled environments. Future work includes adding body posture recognition and improving liveness detection. With its modular, real-time architecture, BioFusion is well-suited for secure applications in healthcare, finance, and enterprise systems.

References:

- **"A Review of Multimodal Biometric Authentication Systems"** – *Kumar and Farik*
- **"GazeTouchPass: Multimodal Authentication Using Gaze and Touch on Mobile Devices"** – *Khamis et al.*
- "Multimodal Authentication" — T. Hild, P. Moore, M. Powell, 2024
- "A Survey of Multimodal Biometrics System" — Snehlata Barde & Rishi Pandey.