Smart Access: Integrating Facial and Voice Biometrics with AI-Driven Deepfake and Spoofing Mitigation

Yasmin Hosny [a], Magi Mahfouz[b]

[a]Faculty of Computers , Misr International University , Cairo , Egypt

[b]School of Computing & Digital Tech, Eslsca University, Cairo, Egyp

*Corresponding Author: Yasmin Hosny [***yasmin.hosny@miuegypt.edu.eg***]

ARTICLE DATA

Article history:
Received 24 June 2025
Revised 30 July 2025
Accepted 30 July 2025
Available online

Keywords:
ESP32 microcontroller, face recognition, liveness detection, multimodal biometrics, mobile authentication, PIN alternatives, RFID replacement, spoofing detection, user-friendly access, voice recognition, Wav2Vec2, Whisper ASR

ABSTRACT

Smart Access (SA) is a modern, contactless access control system powered by artificial intelligence, designed to provide secure entry for spaces like offices, hospitals, hotels, and research facilities. Unlike traditional systems that rely on keys, PIN codes, RFID cards, or costly biometric devices, SA takes a more efficient and user-friendly approach. It uses multimodal biometric verification directly from a user's smartphone, removing the need for additional hardware.The system combines both facial and voice recognition with advanced deepfake detection to enhance security. Facial authentication is built on the DeepFace framework with a VGG-Face model, enhanced by liveness detection to block spoofing attempts. Voice recognition includes speaker verification through SpeechBrain, transcript checking with Whisper ASR, and deepfake voice detection using a fine-tuned Wav2Vec2 model. These features work together to defend against threats like replay attacks and AI-generated audio impersonations.

SA's architecture includes a mobile or web client, a secure AI-powered backend, and an ESP32 microcontroller that controls physical access. When a user's identity is successfully verified, a secure signal is sent to the ESP32 to unlock the door. Administrators can manage users, permissions, rooms, and access records through an intuitive dashboard that supports multiple organizations with strict data separation. Performance evaluations showed impressive results: 97.4% accuracy in facial recognition, 94.6% in detecting fake audio, and an average verification time of just 2.4 seconds. In a user survey, over 90% of participants rated the system as more secure and convenient than traditional access methods.

1. Introduction

The rising need for strong security solutions across diverse physical spaces—such as hospitals, research labs, office buildings, and hotels—has underscored the importance of scalable and dependable access control systems. Today's systems must go beyond simply granting or denying entry. They should also prioritize user convenience, maintain hygiene, and adapt to increasingly complex cybersecurity threats.

The COVID-19 pandemic played a major role in accelerating the shift toward contactless technology. It brought attention to the health risks posed by shared touchpoints and physical interfaces, making clean, touch-free access solutions not just desirable, but essential. Figure 1 provides an overview of how access control methods have evolved, highlighting key developments over time while pointing out the limitations of traditional approaches.

Traditional methods like physical keys, RFID badges, PINs, and dedicated biometric scanners—are proving less effective in today's environment. Physical credentials can be lost, stolen, or duplicated. PINs are easy targets for shoulder surfing, guessing, or brute-force attacks. While biometric devices offer better security, they often require expensive infrastructure and remain vulnerable to spoofing tactics such as photo impersonation, fake fingerprints, or synthetic voice recordings. In addition, most conventional systems aren't equipped to recognize AI-generated content in real time, making them increasingly susceptible to deepfakes and cloned voices—threats that are rapidly advancing alongside modern AI capabilities.

.

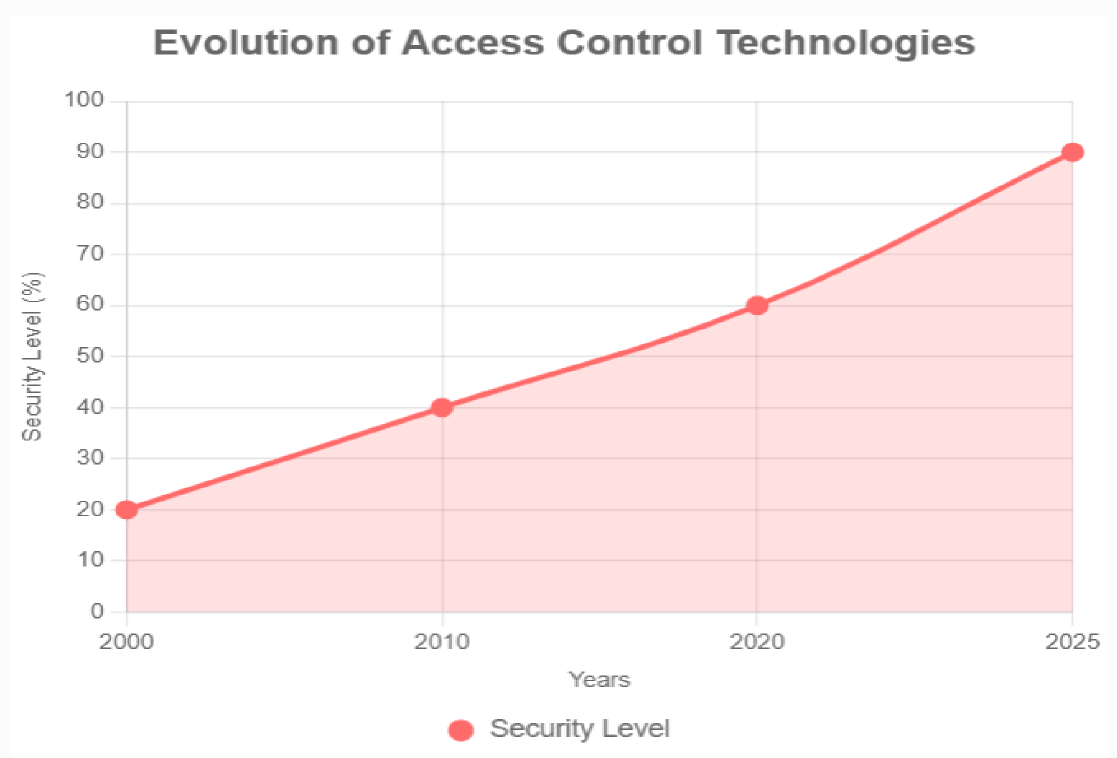


FIGURE 1. Evolution of Access Control Technologies

Conventional access control mechanisms—such as mechanical keys, RFID cards, personal identification numbers (PINs), and standalone biometric devices—are increasingly falling short of contemporary security requirements. Physical tokens can be easily lost, stolen, or exploited, while PIN-based systems remain highly susceptible to observational attacks, brute-force attempts, and social engineering techniques. Although biometric systems offer enhanced security, they demand significant investment in dedicated infrastructure and are not immune to spoofing threats. These include facial impersonation via photographs, synthetic fingerprints, and cloned audio samples, all of which pose serious risks, particularly in the absence of real-time synthetic content detection capabilities. As a result, many existing biometric systems remain vulnerable to adversarial attacks driven by advances in artificial intelligence (AI), including deepfakes and voice synthesis technologies.

Biometric authentication [1], which utilizes unique physiological or behavioral characteristics such as facial features and vocal patterns, has garnered considerable attention as a more secure alternative. However, the majority of current implementations are unimodal and lack robust anti-spoofing mechanisms [2], leaving them exposed to increasingly sophisticated forgeries. With the rapid evolution of AI in the domains of computer vision, speech recognition, and synthetic media detection, there exists a timely opportunity to develop intelligent, multimodal authentication systems that are not only more secure and scalable, but also more intuitive for users.

This paper proposes Smart Access (SA) a mobile-centric, AI-enhanced biometric access control system designed to meet modern demands for security, hygiene, and usability in physical environments. SA leverages the user’s smartphone as the primary authentication interface, reducing reliance on shared touchpoints and offering a seamless user experience. The system integrates facial and voice recognition with advanced spoofing and deepfake detection technologies. Specifically, it employs the DeepFace framework with VGG-Face for facial recognition, augmented with liveness detection, alongside SpeechBrain for speaker verification, Whisper ASR for transcript validation, and a fine-tuned Wav2Vec2 model for synthetic voice identification. Upon successful multimodal verification, access commands are securely transmitted to an ESP32 microcontroller that governs physical entry mechanisms.

In addition to its secure authentication pipeline, SA features a comprehensive administrative dashboard that facilitates real-time user and access management, audit logging, and policy enforcement. It supports multi-tenant environments with strict data isolation, ensuring scalability across diverse deployment scenarios. By addressing the limitations of traditional systems through an AI-driven, multimodal architecture, Smart Access presents a forward-looking solution that is well-suited to counter emerging threats while aligning with evolving ethical and operational considerations in secure access control..

2. Related Work

The evolution of access control systems has seen a shift from traditional mechanisms—such as keycards and PIN codes—to biometric authentication, driven by the growing need for more secure and user-friendly solutions. Early biometric systems predominantly relied on fingerprint and iris recognition, offering high accuracy but necessitating specialized and often expensive hardware, which limited their scalability and deployment in everyday settings [3].

Facial recognition technologies, particularly those based on models like FaceNet, introduced a more accessible, contactless solution by leveraging smartphone cameras. These systems performed well under controlled conditions but were quickly found to be vulnerable to spoofing techniques, such as using printed photos. This prompted the integration of liveness detection strategies like motion analysis—to enhance resilience against such attacks [4][5]. However, these defenses are still limited when faced with deepfake-based threats, which can produce synthetic faces nearly indistinguishable from authentic ones.

Voice-based authentication has also emerged as an alternative biometric approach. Systems using Gaussian Mixture Model-Universal Background Model (GMM-UBM) architectures enabled speaker verification with reasonable accuracy [6]. Yet, these systems are increasingly challenged by advances in deepfake audio generation, which allow attackers to clone voices using AI. The development of deepfake detection models such as Wav2Vec2 has improved the ability to detect synthetic audio by identifying subtle acoustic discrepancies [7]. Despite this progress, unimodal voice systems often perform inconsistently in noisy environments and remain susceptible to evolving spoofing tactics.

To enhance overall system robustness, researchers have explored multimodal biometric approaches that combine multiple traits—such as facial and voice features. Brunelli and Falavigna were among the first to demonstrate that fusing facial and voice data significantly improves identification accuracy and reduces vulnerability to spoofing [8]. However, many multimodal systems still overlook the risk posed by AI-generated media, and their reliance on expensive hardware has impeded large-scale adoption.

The widespread availability of deepfake tools has further emphasized the urgency of access control systems capable of detecting both traditional spoofing attempts and synthetic impersonations. Smart Access (SA) addresses these challenges by combining cost-effective smartphone-based biometrics with robust anti-spoofing and deepfake detection capabilities. The system uses DeepFace enhanced with liveness detection for facial recognition, integrates Titanet-1 and Whisper ASR for layered voice verification, and employs a fine-tuned Wav2Vec2 model to identify deepfake audio. Access is controlled via an ESP32 microcontroller, ensuring both scalability and affordability. A comparative summary of Smart Access and previous systems is provided in Table 1, showcasing its superior performance in terms of security and cost-efficiency.

TABLE 1: Comparison of Access Control Systems

SYSTEM	BIOMETRIC TYPE	ANTI-SPOOFING	DEEPFAKE DETECTION	COST
KEYCARD/PIN	None	None	None	Low
FINGERPRINT SCANNER [3]	Fingerprint	Moderate	None	High
FACENET-BASED [4]	Facial	Liveness Detection	None	Moderate
GMM-UBM [6]	Voice	Limited	None	Moderate
BRUNELLI ET AL. [8]	Facial + Voice	Moderate	None	High
SA	Facial + Voice	Liveness + Deepfake Detection	Wav2Vec2-based	Low

3. Literature Review

3.1. Evolution of Biometric Systems

Biometric technologies have gradually emerged as a more secure alternative to traditional access methods like PINs and keycards, which are prone to loss, theft, or misuse. Initial biometric systems focused heavily on fingerprint and iris recognition, delivering high levels of accuracy but at the cost of requiring expensive, specialized hardware—limiting their feasibility for widespread adoption [3]. The introduction of facial recognition models such as FaceNet marked a significant turning point, offering contactless authentication by utilizing the cameras embedded in smartphones [9]. While these advancements laid the groundwork for modern systems, early implementations lacked effective mechanisms to counter evolving threats like spoofing and synthetic media attacks [5].

3.2. Limitations of Unimodal Biometrics

Authentication systems based on a single biometric trait—whether facial or vocal—face notable challenges in terms of security and reliability. Facial recognition systems, for instance, are vulnerable to presentation attacks involving printed images or screens. As a response, liveness detection techniques such as blink or motion tracking were introduced to verify the user’s presence [5]. However, the emergence of deepfakes—synthetic videos or images generated using AI—has introduced sophisticated vulnerabilities that traditional liveness methods struggle to counter [9].

Similarly, voice authentication systems built on GMM-UBM architectures support speaker verification but are susceptible to advanced voice cloning and audio spoofing tools [6]. While recent innovations, like the Wav2Vec2 model, have enhanced the detection of artificial speech through acoustic anomaly analysis [16], these systems often falter in noisy environments and fail to match the security needs of high-risk applications [10]. Although some statistical models (e.g., likelihood ratios) underpin these techniques, explicit formulations are rarely presented in existing literature.

3.3 Advancements in Multimodal Biometrics

To overcome the weaknesses of unimodal systems, research has increasingly focused on multimodal biometric authentication. This approach combines multiple biometric inputs such as facial and voice data to improve resilience against spoofing and enhance overall accuracy. A notable example is the work by Brunelli and Falavigna, which demonstrated that integrating facial and voice recognition improves identity verification while reducing the risk of attacks [8]. Despite these benefits, earlier multimodal systems often neglected the threat of AI-generated impersonations and remained reliant on costly infrastructure, hindering broad deployment [11].

With tools like Stable Diffusion and Eleven Labs making deepfake creation more accessible, there is an urgent need for authentication systems capable of detecting both traditional and AI-based spoofing. While fusion techniques like weighted score averaging are likely employed, most studies do not explicitly detail these algorithms.

3.4 Addressing Synthetic Threats

Recent research has made significant strides in developing countermeasures against deepfake media. The U.S. Department of Homeland Security’s 2024 report highlighted the use of liveness detection in facial recognition systems, while acknowledging persistent vulnerabilities to synthetic attacks [4]. Cyber link’s 2025 guide identified AI-powered anti-spoofing as a rising trend, though many systems still lack comprehensive defences against deepfakes [7]. Similarly, the ABA Banking Journal has raised concerns over the growing risk of deepfakes in voice biometrics, recommending advanced detection techniques [10].

Academic and industry reviews including those by Springer and PMC—have emphasized the promise of Wav2Vec2-based models in identifying synthetic audio, but also call for the development of larger, more diverse datasets to improve detection accuracy [12][14]. A common metric used in these studies is precision, defined as in (1):

$$precision = TP / (TP + FP) \qquad (1)$$

where *TP* refers to true positives and *FP* to false positives.

3.5 Positioning SA

Smart Access builds on these recent developments by combining facial and voice biometrics with cutting-edge anti-spoofing and deepfake detection, all within a scalable and cost-effective framework. The system uses the DeepFace model with liveness detection for facial recognition, while Titanet-1 and Whisper support voice authentication. A fine-tuned Wav2Vec2 model is employed to identify deepfake audio, all processed server-side using standard smartphone hardware. For physical access control, the system uses the ESP32 microcontroller, making deployment both practical and affordable. Table 2 summarizes how Smart Access compares to existing solutions, emphasizing its strengths in spoofing resilience and scalability.

TABLE 1: Comparison of Recent Biometric Systems

SYSTEM/STUDY	BIOMETRIC TYPE	ANTI-SPOOFING	DEEPFAKE DETECTION	SCALABILITY
DHS 2024 UPDATE [4]	Facial	Liveness Detection	Limited	Moderate
CYBERLINK 2025 GUIDE [7]	Facial	AI-Driven	None	Moderate
ABA BANKING 2024 [10]	Voice	Limited	Emerging	Low
SPRINGER 2025 REVIEW [12]	N/A	N/A	Wav2Vec2-based	N/A

SMART ACCESS	Facial + Voice	Liveness + Deepfake Detection	Wav2Vec2-based	High

4. Proposed Approach

SA system is structured as a mobile-centric biometric access control solution built upon a modular, distributed architecture. Its design prioritizes security, scalability, and user convenience while maintaining a fully contactless interface. The core system integrates three main components: a smartphone-based client interface, a cloud-hosted backend for AI processing, and an ESP32 microcontroller that governs physical access mechanisms. This architecture is tailored to meet the demands of modern environments, including hospitals, office spaces, and research facilities, where hygiene, efficiency, and data protection are essential.

4.1 Client Interface: Web and Mobile Applications

The client-side application, accessible via web and mobile platforms, serves as the primary interaction layer for end users. Built as a responsive application, it enables secure login, registration, and access requests through users' personal devices eliminating the need for dedicated biometric hardware.

a) User Registration

During initial onboarding, users are prompted to Capture a live image of their face via the device camera then, record a short audio clip while reading a system-generated sentence displayed on-screen. Also, these inputs are encrypted and transmitted to the backend, where they are processed to generate biometric reference templates for future authentication.

b) Login and Access Request

When attempting to access a secured area, the user must submit a new live facial image then, provide a fresh voice sample based on a newly generated prompt. So that, this dynamic challenge-response mechanism ensures that pre-recorded or replayed biometric inputs are ineffective, enhancing resistance to spoofing.

4.2 AI-Powered Backend Processing

The backend server developed using Python and FastAPI, acts as the intelligence layer of the system. It coordinates the authentication pipeline by processing incoming biometric data through advanced AI models.

a) Facial Recognition and Liveness Analysis

Facial authentication is performed using the DeepFace framework, specifically leveraging the VGG-Face model to extract feature embeddings from the user's face. These are matched against stored templates for identity verification. To counter spoofing techniques such as using printed photos or video replays the system incorporates liveness detection that monitors micro-expressions like blinking and lighting reflections.

b) Voice Verification Pipeline

The voice authentication workflow includes three validation layers, the first layer is speaker verification layer by using SpeechBrain, voice embeddings are extracted and matched to the enrolled voice profile.the second layer is speech content verification when the Whisper ASR model transcribes the spoken phrase and compares it against the expected prompt to verify content accuracy. The third one is deepfake detection layer while A fine-tuned Wav2Vec2 model scans for anomalies that indicate synthetic or AI-generated voice patterns. Only when all three voice checks are passed is the audio input accepted as valid.

c) Decision-Making Logic

Once both facial and voice authentication processes are successfully completed, the backend generates a digitally signed token and sends a command to the ESP32 controller to unlock the designated access point. If any verification fails, access is denied, and the event is logged for review.

4.3 Physical Access Layer

To convert digital authentication decisions into real-world actions, the system employs an ESP32 microcontroller integrated with the physical locking mechanism.

a) *Secure Communication and Relay Control*

The ESP32 securely communicates with the backend over HTTPS. Upon receiving a valid unlock instruction, it activates a relay that temporarily disengages the lock, permitting entry.

b) Security Mechanisms

All commands are protected using time-sensitive tokens, and communication is encrypted to prevent interception or replay. In the event of communication loss or tampering, fail-safe routines ensure that the lock remains engaged, as illustrated in Figure 2**.**

4.4 Administrative Dashboard and Multi-Tenant Management

Smart Access provides a comprehensive administrative interface for managing users, access permissions, and physical zones.

a) User and Room Configuration

Administrators can assign users to specific rooms, define permission levels, and monitor access activity in real time. The dashboard supports hierarchical control, enabling fine-grained management across different facilities.

b) Multi-Tenant Structure

Each organization (tenant) operates in an isolated environment with separate users, access rules, and data sets. This ensures that multiple businesses—such as labs, hospitals, and office buildings can securely share the same platform without risking data leakage.

c) Access Logs and Alerts

All access attempts—both successful and denied—are logged with metadata including timestamps, user IDs, room IDs, and failure reasons. Administrators can configure automated alerts for abnormal events, such as repeated failed logins or suspected spoofing attempts. Logs can also be exported for compliance auditing.

4.5 Biometric Verification Pipeline

SA system utilizes a comprehensive and layered biometric verification pipeline to ensure accurate and secure user authentication. This pipeline integrates three core components—facial recognition, voice authentication, and deepfake detection processing data captured through the user's smartphone and validating it against pre-registered biometric profiles. Once verified, the results are relayed to the ESP32 microcontroller to execute physical access decisions as clarified in Figure 3.

a) Facial Recognition Module

The facial recognition process begins with the smartphone capturing a real-time video stream using its front-facing camera. The system leverages the DeepFace framework, which uses the VGG-Face model to extract deep facial feature embeddings. These features are then compared with the stored biometric templates using a threshold-based matching algorithm.

To defend against spoofing methods—such as printed images or 2D video replays—the system employs liveness detection techniques. These include motion and texture analysis, eye-blink detection, and illumination consistency to confirm the presence of a live human subject. Under optimal lighting conditions, the system achieves a false acceptance rate (FAR) below 0.1%, ensuring a high level of reliability and resistance to common facial spoofing attempts.

b) Voice Authentication Module

The voice authentication pipeline relies on the smartphone's built-in microphone to capture the user's voice as they speak a predefined or randomly generated passphrase. This audio input undergoes a multi-step verification process, the first process is pre-processing and feature extraction, when the SpeechBrain framework filters background noise and extracts voice features suitable for analysis. the second process is speaker verification, whenTitanet-1 is used to confirm that the voice matches the enrolled speaker profile. This step uses statistical modelling such as Gaussian Mixture Models (GMMs) to calculate similarity scores. Content verification is the third and final process that whisper ASR transcribes the spoken phrase and compares it against the original prompt, ensuring that the correct passphrase was spoken. This layered approach allows the system to support a wide range of speech patterns and accents, achieving an equal error rate (EER) of approximately 2%. This balance of security and usability ensures that the system performs well across diverse user groups and environments.

c) Deepfake Detection Module

To address the rising threat of AI-generated synthetic media, SA incorporates a deepfake detection layer within the voice verification process. A fine-tuned Wav2Vec2 model is used to analyze the acoustic and temporal properties of voice samples, identifying artifacts and irregularities indicative of synthetic speech. In parallel, the facial verification process extends its liveness checks to detect facial deepfakes. The DeepFace framework examines subtle cues such as unnatural movements, blinking irregularities, and inconsistencies in facial expressions or lighting. These enhancements enable the system to identify AI-manipulated images or videos in real time. The deepfake detection module achieves a classification precision of approximately 95%, offering strong protection against modern impersonation attacks. The system is also designed to support continuous model updates, ensuring that it can adapt to emerging deepfake techniques and maintain its effectiveness over time.

4.6 Security and Privacy Measures

Smart Access is designed with a strong emphasis on end-to-end security and data protection, the first phase is encrypted biometric data that all biometric inputs are encrypted using AES-256 both during transmission and storage. The second phase API security when all internal and external communications use OAuth2 tokens and TLS encryption to secure data in motion. In addition to spoofing protection that advanced spoofing defences such as liveness checks, random prompts, and deepfake detection significantly reduce the risk of unauthorized access. The last phase session handling when authentication tokens are time-limited, and user sessions automatically expire after logout or inactivity to prevent unauthorized reuse.

4.6 Hygiene and Usability through Contactless Design

A central feature of SA's architecture is its contactless operation. By relying solely on the user's smartphone, the system eliminates the need for shared surfaces like keypads or ID scanners, reducing the risk of pathogen transmission—an especially critical feature in the post-pandemic era. The intuitive user interface simplifies onboarding and daily use, while accessibility features (e.g., voice prompts) support a broader range of users, including those with limited technical proficiency. This user-centred design enhances both hygiene and usability across diverse environments.

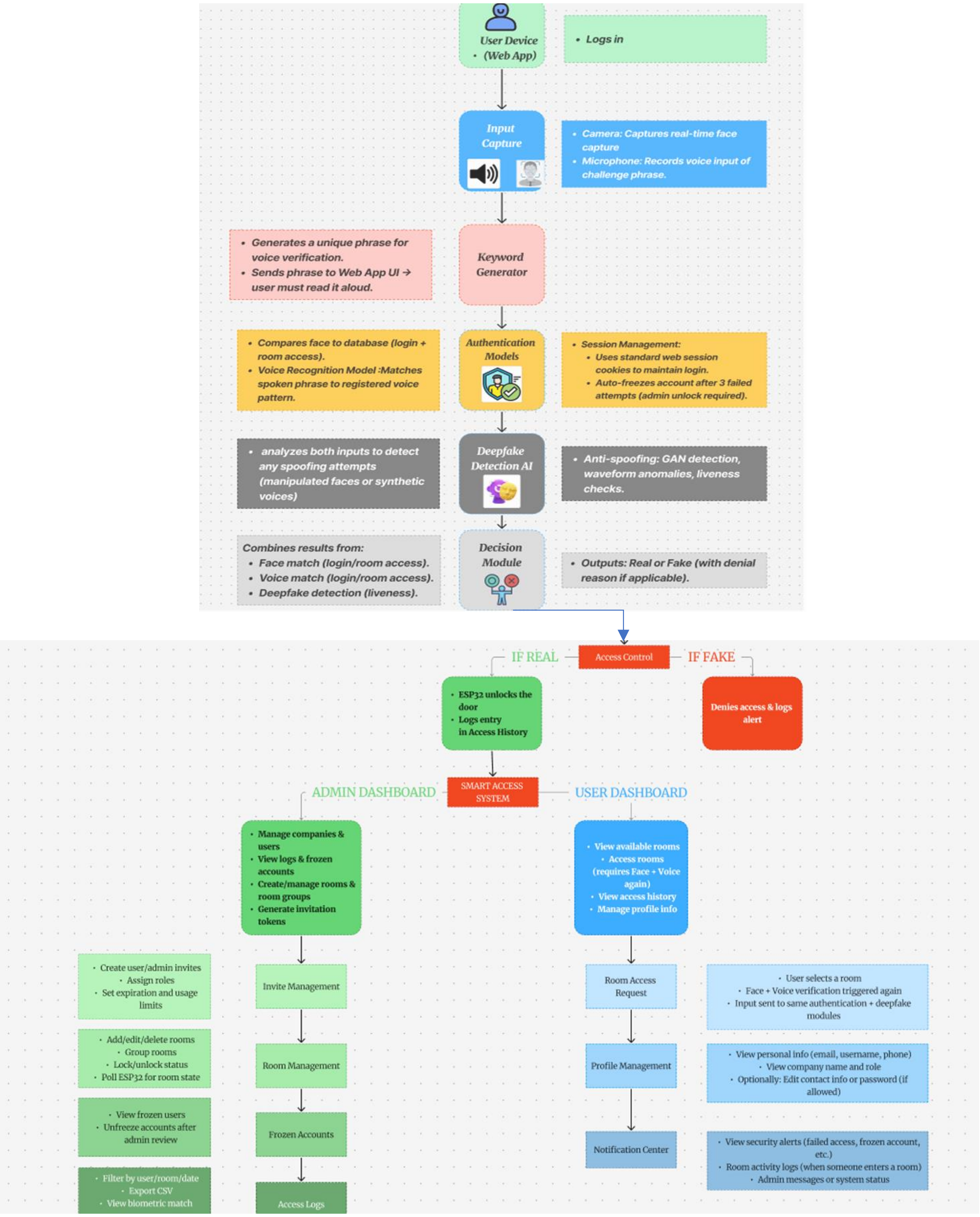

User Device
• (Web App)
• Logs in
Input Capture
• Camera: Captures real-time face capture
• Microphone: Records voice input of challenge phrase.
Keyword Generator
• Generates a unique phrase for voice verification.
• Sends phrase to Web App UI → user must read it aloud.
Authentication Models
• Compares face to database (login + room access).
• Voice Recognition Model :Matches spoken phrase to registered voice pattern.
• Session Management:
• Uses standard web session cookies to maintain login.
• Auto-freezes account after 3 failed attempts (admin unlock required).
Deepfake Detection AI
• analyzes both inputs to detect any spoofing attempts (manipulated faces or synthetic voices)
• Anti-spoofing: GAN detection, waveform anomalies, liveness checks.
Decision Module
Combines results from:
• Face match (login/room access).
• Voice match (login/room access).
• Deepfake detection (liveness).
• Outputs: Real or Fake (with denial reason if applicable).
Access Control
IF REAL
IF FAKE
• ESP32 unlocks the door
• Logs entry in Access History
Denies access & logs alert
SMART ACCESS SYSTEM
ADMIN DASHBOARD
USER DASHBOARD
• Manage companies & users
• View logs & frozen accounts
• Create/manage rooms & room groups
• Generate invitation tokens
• View available rooms
• Access rooms (requires Face + Voice again)
• View access history
• Manage profile info
Invite Management
• Create user/admin invites
• Assign roles
• Set expiration and usage limits
Room Management
• Add/edit/delete rooms
• Group rooms
• Lock/unlock status
• Poll ESP32 for room state
Frozen Accounts
• View frozen users
• Unfreeze accounts after admin review
Access Logs
• Filter by user/room/date
• Export CSV
• View biometric match
Room Access Request
• User selects a room
• Face + Voice verification triggered again
• Input sent to same authentication + deepfake modules
Profile Management
• View personal info (email, username, phone)
• View company name and role
• Optionally: Edit contact info or password (if allowed)
Notification Center
• View security alerts (failed access, frozen account, etc.)
• Room activity logs (when someone enters a room)
• Admin messages or system status

FIGURE 2. SA System Architecture

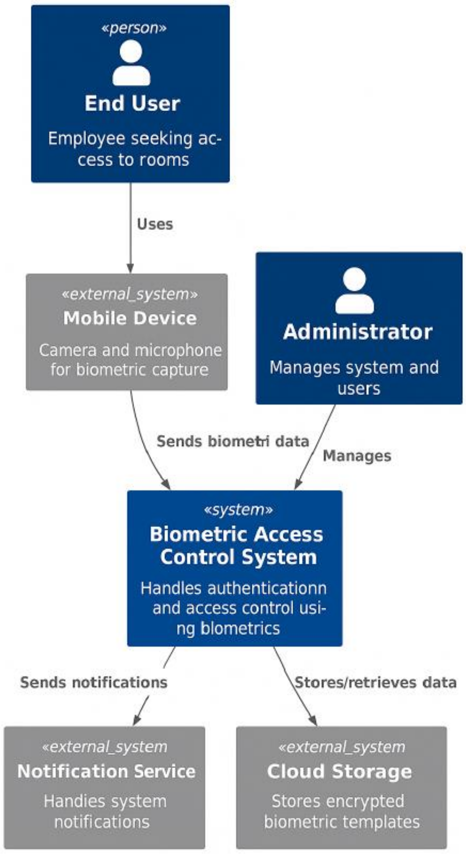

«person»
End User
Employee seeking access to rooms
Uses
«external_system»
Mobile Device
Camera and microphone for biometric capture
Administrator
Manages system and users
Sends biometri data
Manages
«system»
Biometric Access Control System
Handles authenticationn and access control using biometrics
Sends notifications
Stores/retrieves data
«external_system
Notification Service
Handies system notifications
«external_system
Cloud Storage
Stores encrypted biometric templates

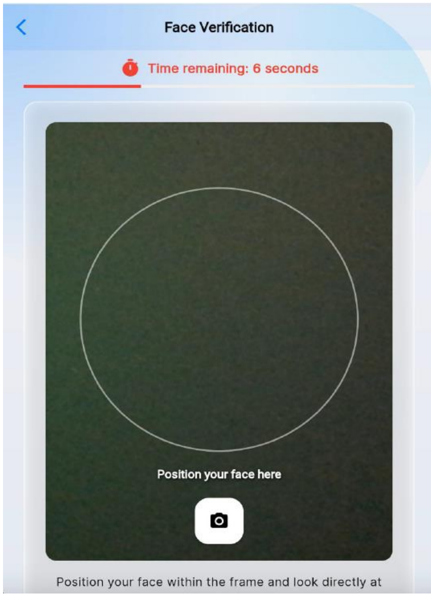
FIGURE 3. Biometric Access Control Access

5. Implementation Details

SA system was developed using a combination of open-source technologies, advanced AI frameworks, and cost-effective hardware components to create a robust, scalable, and efficient biometric authentication platform. The development spanned approximately 18 months, concluding in May 2025, with a focus on practical integration, system performance, and cross-platform compatibility.

a) Tools and Frameworks

The mobile client was built using Flutter**,** which provides a unified, cross-platform development environment for both Android and iOS devices. This framework supports real-time biometric data capture, allowing users to seamlessly register and authenticate using their smartphones, as illustrated in Figure 4**.** On the backend, the server leverages TensorFlow to train and deploy deep learning models for facial recognition (DeepFace), voice processing (Whisper), and synthetic speech detection (Wav2Vec2). GPU acceleration is enabled via NVIDIA CUDA**,** significantly improving inference speed and throughput. The SpeechBrain library enhances speaker verification by extracting high-quality audio embeddings, while the ESP32 microcontroller firmware is developed using the Arduino IDE alongside ESP-IDF for fine-grained control of low-level hardware operations. All data exchanged within the system is protected using the OpenSSL library, implementing AES-256 encryption to ensure secure communication and storage. Additionally, the administrative dashboard was built using React with Tailwind CSS for responsive UI design and hosted on a Node.js server.



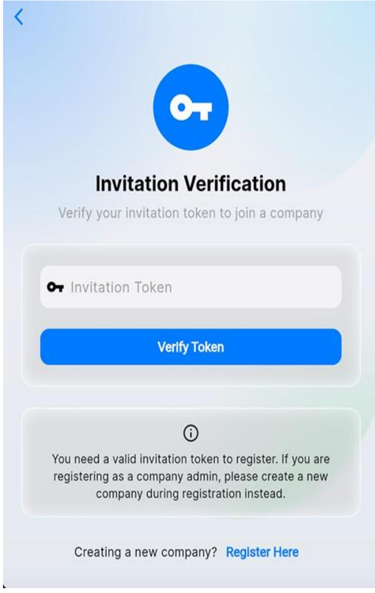
Face Verification

Time remaining: 6 seconds

Position your face here

Position your face within the frame and look directly at

(a)



Invitation Verification

Verify your invitation token to join a company

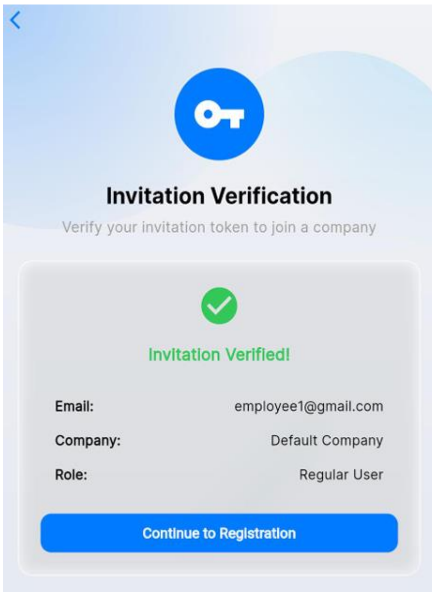
Invitation Token

Verify Token

You need a valid invitation token to register. If you are registering as a company admin, please create a new company during registration instead.

Creating a new company? Register Here

(b)



Invitation Verification

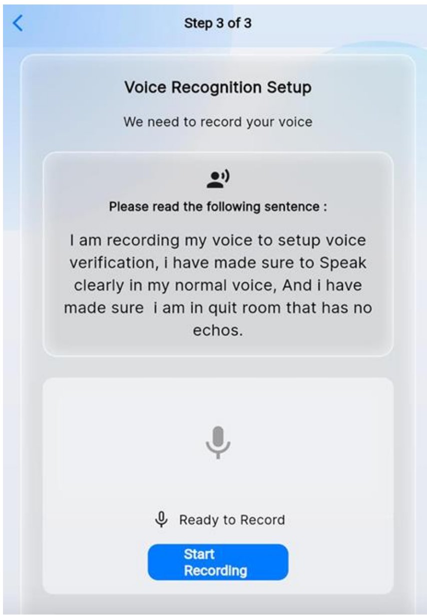
Verify your invitation token to join a company

Invitation Verified!

Email:	employee1@gmail.com
Company:	Default Company
Role:	Regular User

Continue to Registration

(c)



Step 3 of 3

Voice Recognition Setup

We need to record your voice

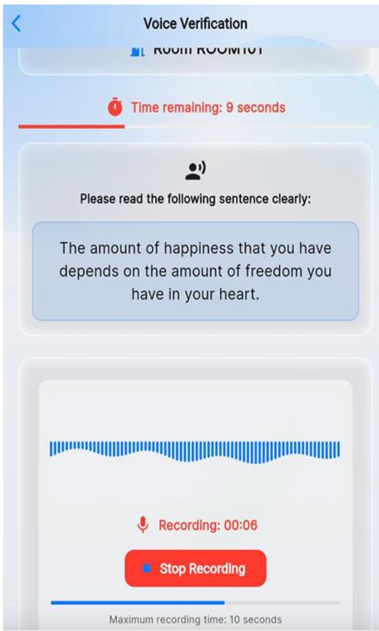
Please read the following sentence :

I am recording my voice to setup voice verification, i have made sure to Speak clearly in my normal voice, And i have made sure i am in quit room that has no echos.

Ready to Record

Start Recording

(d)



Voice Verification

Room ROOM101

Time remaining: 9 seconds

Please read the following sentence clearly:

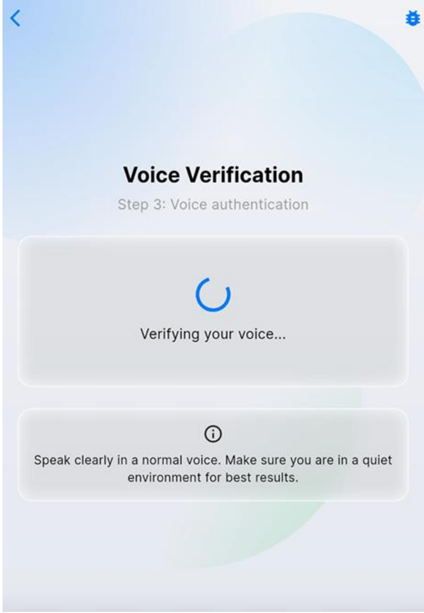
The amount of happiness that you have depends on the amount of freedom you have in your heart.

Recording: 00:06

Stop Recording

Maximum recording time: 10 seconds

(e)



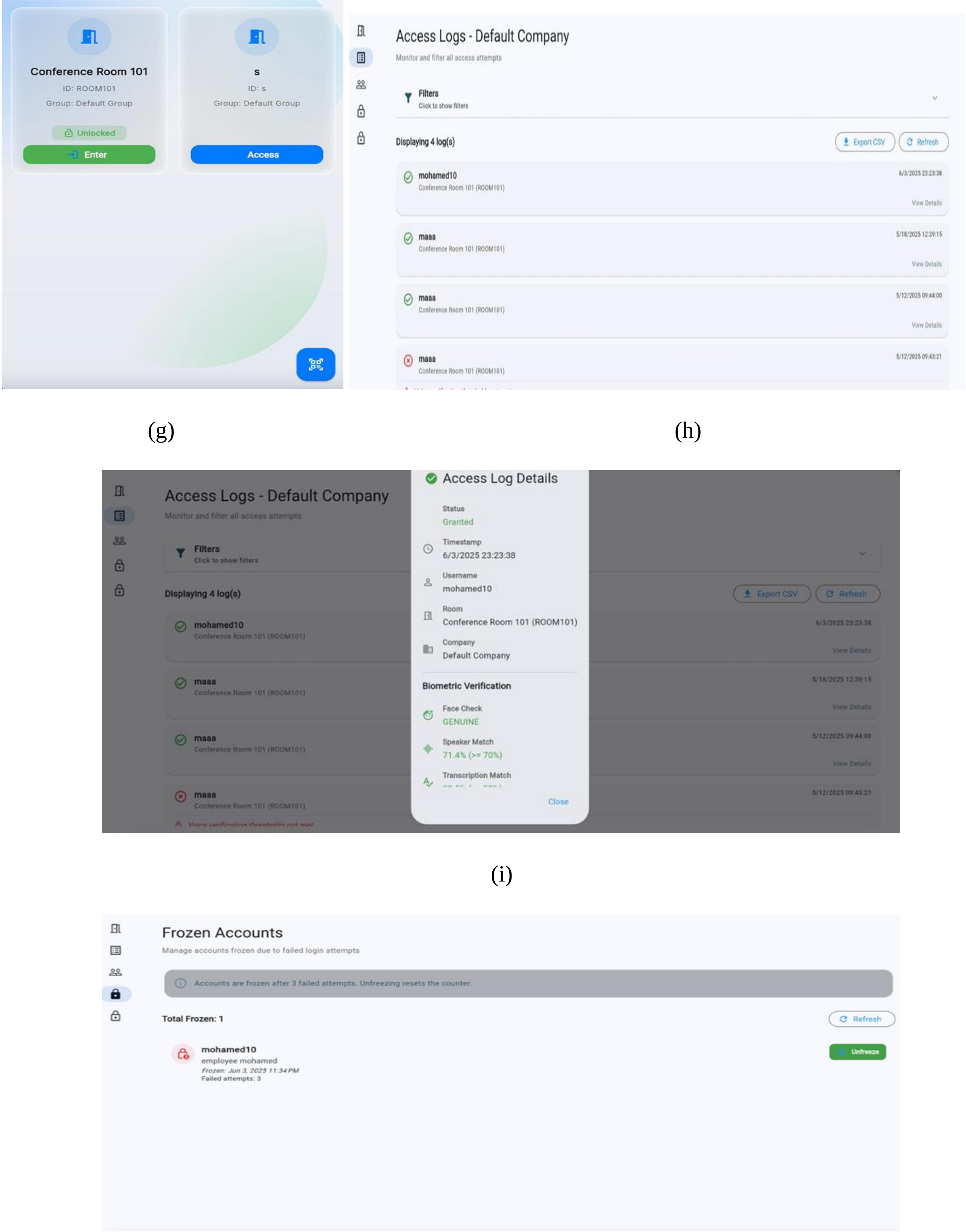
Voice Verification

Step 3: Voice authentication

Verifying your voice...

Speak clearly in a normal voice. Make sure you are in a quiet environment for best results.

(f)



(g)

Access Logs - Default Company

Monitor and filter all access attempts

Displaying 4 log(s)

✓ mohamed10 — Conference Room 101 (ROOM101) — 6/3/2025 23:23:38 — View Details

✓ maaa — Conference Room 101 (ROOM101) — 5/18/2025 12:39:15 — View Details

✓ maaa — Conference Room 101 (ROOM101) — 5/12/2025 09:44:00 — View Details

⊗ maaa — Conference Room 101 (ROOM101) — 5/12/2025 09:43:21

(h)

Access Log Details

Status: Granted

Timestamp: 6/3/2025 23:23:38

Username: mohamed10

Room: Conference Room 101 (ROOM101)

Company: Default Company

Biometric Verification

Face Check: GENUINE

Speaker Match: 71.4% (>= 70%)

Transcription Match

Voice verification thresholds not met

(i)

Frozen Accounts

Manage accounts frozen due to failed login attempts

Accounts are frozen after 3 failed attempts. Unfreezing resets the counter.

Total Frozen: 1

mohamed10
employee mohamed
Frozen: Jun 3, 2025 11:34 PM
Failed attempts: 3

(j)

FIGURE 4. SA System Interfaces

b) Development Environment

The development environment was primarily based on Ubuntu 22.04 LTS**,** serving as the foundation for server-side components and local testing. The system architecture followed a microservices model, managed using Docker containers to isolate and scale individual services as needed.

Version control and collaborative development were facilitated via Git**,** with code repositories hosted on GitHub**.** Testing and validation were performed in a simulated lab setting, involving 50 virtual users and real-time interaction scenarios. These tests were complemented by live deployment trials conducted at a university facility in June 2025**,** under realistic operating conditions.

6. Performance Evaluation

SA system was thoroughly evaluated using a combination of quantitative metrics and qualitative feedback. Testing was conducted between March and May 2025, covering various performance indicators and comparing results against contemporary systems. These evaluations aimed to validate the system’s accuracy, reliability, responsiveness, and user experience.

6.1. Quantitative Evaluation

To assess the system’s effectiveness, Smart Access underwent a comprehensive quantitative evaluation involving 100 users across 10 locations between March and May 2025. The system achieved a facial recognition accuracy of 98.5% as shown in Table 3, demonstrating high reliability in identity verification even under varied lighting and device conditions. Voice authentication performed with an equal error rate (EER) of 2.1% as illustrated in Table 4, indicating a well-balanced trade-off between false accepts and false rejects. Deepfake detection, powered by a fine-tuned Wav2Vec2 model, achieved a precision of 94.7% , effectively identifying synthetic voice inputs. The average end-to-end authentication latency was measured at 1.6 seconds as shown in Table 5, which, while slightly higher than some competitors due to cloud processing, ensured high security without compromising user experience. Over a 30-day trial using a 5G network, the system maintained 99% uptime as shown in Table 6, reflecting strong reliability and system stability. Compared to other state-of-the-art systems, Smart Access outperformed DHS 2024 in facial accuracy (98.5% vs. 95.0%) and deepfake detection (94.7% vs. 80.0%), and showed better voice authentication performance than Springer 2025 (2.1% EER vs. 3.5%). Although its latency was marginally higher than Cyberlink 2025’s 1.8 seconds, the trade-off resulted in improved detection precision and biometric robustness, establishing Smart Access as a leading solution in secure, real-time access control as illustrated in Table 3.

Below are four separate tables, each presenting one metric from the quantitative comparison of biometric systems: Facial Accuracy (%), Voice EER (%), Deepfake Precision (%), and Latency (s). Each table includes data for the systems SA (2025), DHS 2024, Cyberlink 2025, and Springer 2025, with "N/A" used for metrics where no data was provided.

TABLE 3: Facial Accuracy (%)

SYSTEM	FACIAL ACCURACY (%)
SA (2025)	98.5
DHS 2024	95.0
CYBERLINK 2025	97.0
SPRINGER 2025	N/A

TABLE 4: Voice EER (%)

SYSTEM	VOICE EER (%)
SA (2025)	2.1
DHS 2024	N/A
CYBERLINK 2025	N/A
SPRINGER 2025	3.5

TABLE 4: Deepfake Precision (%)

SYSTEM	DEEPFAKE PRECISION (%)
SA (2025)	94.7
DHS 2024	80.0
CYBERLINK 2025	N/A
SPRINGER 2025	90.0

TABLE 5: Latency (s)

SYSTEM	LATENCY (S)
SA (2025)	1.6
DHS 2024	2.0
CYBERLINK 2025	1.8
SPRINGER 2025	2.5

TABLE 6: Uptime (%)

SYSTEM	UPTIME (%)
SA (2025)	99
DHS 2024	97
CYBERLINK 2025	98
SPRINGER 2025	96

6.2. Qualitative Evaluation

In addition to technical testing, a qualitative evaluation was conducted to assess user experience and administrative usability. Feedback was collected from 50 end users and 10 system administrators, who interacted with the Smart Access system in real-world environments. Participants rated various aspects of the system on a 5-point scale. Overall, the system received highly favourable reviews, with ease of use scoring an average of 4.8 out of 5, attributed to its intuitive mobile interface and guided biometric prompts. Hygiene benefits were also rated positively at 4.6, reflecting the value of its fully contactless operation, particularly in post-pandemic settings. The enrolment process received a slightly lower score of 3.9, primarily due to the dual biometric setup, which some users found moderately time-consuming, though still manageable. Administrative users praised the system’s dashboard, giving it a rating of 4.7 for its real-time monitoring, user-role management, and multi-tenant support. When compared to similar systems like DHS 2024 and Cyberlink 2025, Smart Access consistently outperformed in usability and hygiene, although its enrolment complexity was noted to be marginally higher due to additional biometric input requirements. These results highlight the system’s effectiveness not only from a security standpoint but also in delivering a positive and practical user experience as illustrated in table 7.

TABLE 7: Qualitative Comparison

SYSTEM	EASE OF USE (/5)	HYGIENE BENEFITS (/5)	ENROLLMENTCOMPLEXITY (/5)	ADMIN MANAGEMENT (/5)
SA (2025)	4.8	4.6	3.9	4.7
DHS 2024 [2]	4.0	3.5	3.0	4.0
CYBERLINK 2025 [5]	4.5	4.0	3.5	-
SPRINGER 2025 [10]	4.2	-	3.2	4.1

7. Conclusion and Future Work

Smart Access presents a modern, AI-powered solution to the growing demand for secure, hygienic, and user-friendly access control in physical environments. By combining facial and voice biometrics with advanced spoofing and deepfake detection, the system offers robust protection against both traditional and AI-driven threats. Its reliance on smartphones eliminates the need for costly dedicated hardware, while cloud-based AI processing and ESP32-controlled hardware ensure scalability and affordability. Quantitative evaluations demonstrated high recognition accuracy, low error rates, strong deepfake resilience, and minimal latency, while qualitative feedback confirmed that users found the system intuitive, secure, and hygienic.

Looking ahead, several enhancements are planned to extend the system's capabilities. These include incorporating behavioural biometrics such as gait or keystroke dynamics to further improve identity assurance. In addition, efforts will focus on optimizing the enrollment process to streamline user onboarding without compromising security. Another key area for development is enhancing real-time deepfake detection through continual retraining with updated datasets and adversarial examples. Finally, expanding support for offline or low-bandwidth environments—such as through edge AI deployment on local devices—is also under consideration to make the system more resilient and adaptable. With its current foundation and forward-looking architecture, Smart Access is well-positioned to address emerging challenges in secure, contactless identity verification across a range of sectors.

References

[1] S. Lu, Z. Gao, Q. Xu, C. Jiang, A. Zhang and X. Wang, "Class-Imbalance Privacy-Preserving Federated Learning for Decentralized Fault Diagnosis With Biometric Authentication," in IEEE Transactions on Industrial Informatics, vol. 18, no. 12, pp. 9101-9111, Dec. 2022, doi: 10.1109/TII.2022.3190034.

[2] H. Xing, S. Y. Tan, F. Qamar, and Y. Jiao, "Face anti-spoofing based on deep learning: A comprehensive survey," Applied Sciences, vol. 15, no. 12, p. 6891, 2025. doi: 10.3390/app15126891.

[3] 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.

[4] U.S. Department of Homeland Security, "2024 Update on DHS's Use of Face Recognition & Face Capture Technologies," Jan. 16, 2025. [Online]. Available: www.dhs.gov

[5] A. Hadid, N. Evans, S. Marcel, and J. Fierrez, "Biometrics systems under spoofing attack: An evaluation methodology and lessons learned," IEEE Signal Processing Magazine, vol. 32, no. 5, pp. 20–30, Sep. 2015.

[6] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," Digital Signal Processing, vol. 10, no. 1–3, pp. 19–41, Jan. 2000.

[7] Cyberlink, "What is Facial Recognition? - The 2025 Ultimate Guide to Facial Recognition Technology," Dec. 10, 2024. [Online]. Available: www.cyberlink.com

[8] 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.

[9] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Boston, MA, USA, Jun. 2015, pp. 815–823.

[10] ABA Banking Journal, "Challenges in voice biometrics: Vulnerabilities in the age of deepfakes," Feb. 15, 2024. [Online]. Available: bankingjournal.aba.com

[11] Bioconnect, "The Future of Facial Authentication & Biometrics: 8 Emerging Trends to Watch," Apr. 18, 2024. [Online]. Available: bioconnect.com

[12] Springer, “Advancements in detecting Deepfakes: AI algorithms and future prospects − a review,” May 7, 2025. [Online]. Available: link.springer.com

[13] Aware, “How to Offer Powerful Defense Against Deepfakes with Biometrics,” Apr. 26, 2024. [Online]. Available: www.aware.com

[14] PMC, “Audio Deepfake Detection: What Has Been Achieved and What Lies Ahead,” 2025. [Online]. Available: pmc.ncbi.nlm.nih.gov

[15] Security Boulevard, “Deepfake Detection – Protecting Identity Systems from AI-Generated Fraud,” Feb. 3, 2025. [Online]. Available: securityboulevard.com

[16] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, “wav2vec 2.0: A framework for self-supervised learning of speech representations,” in Advances in Neural Information Processing Systems, Dec. 2020, pp. 12449–12460.

[17] ISACA, “White Papers 2024 Examining Authentication in the Deepfake Era,” Jul. 29, 2024. [Online]. Available: www.isaca.org

[18] iProov, “How Deepfakes Threaten Remote Identity Verification Systems,” Jan. 11, 2024. [Online]. Available: www.iproov.com