International Journal of Safety and Security Engineering
Vol. 16, No. 2, February, 2026, pp. 283-296
Journal homepage: http://iieta.org/journals/ijsse

Secure Hybrid Authentication Using Facial Biometrics, Liveness Verification, and Time-Based One-Time Passwords

Kuldeep Vayadande[1], Praveenkumar Patel[2*], Govinda Sambare[3], Prajakta Pawar[4], Premanand Ghadekar[1], Namrata Salgar[1], Vivek Kheradkar[5], Shlok Shinde[1], Aryan Shinde[1], Siddhant Shinde[1], Prathamesh Shinde[1], Samyak Lokhande[1]

[1] Department of Information Technology Vishwakarma, Institute of Technology, Pune 411037, India
[2] Bharati Vidyapeeth's College of Engineering, Kolhapur 416013, India
[3] Pimpri Chinchwad College of Engineering, Pune 411044, India
[4] Bharati Vidyapeeth's College of Engineering, Pune 412115, India
[5] D.K.T.E.'s Textile and Engineering Institute, Kolhapur 416115, India

Corresponding Author Email: praveenpatel7@gmail.com

https://doi.org/10.18280/ijsse.160204

Received: 12 December 2025
Revised: 29 January 2026
Accepted: 20 February 2026
Available online: 28 February 2026

Keywords:
biometric authentication, facial recognition, liveness detection, face anti-spoofing, time-based one-time password, deepfake detection, MediaPipe Face Mesh, multi-factor authentication

ABSTRACT

Biometric authentication has become one of the topics that have attracted extensive interest because of its increased security and usability over the traditional systems that use passwords. Facial recognition is one of the common biometric modalities that has gained wide adoption due to its non-contact aspect. Nevertheless, it is susceptible to spoofing attacks, including printed images, replayed videos and deep fake-based impersonation. A hybrid authentication system called FaceOTP is suggested to counter these shortcomings, a system that combines face recognition, liveness and time-based one-time password (TOTP) authentication. The system uses MediaPipe Face Mesh to detect facial landmarks and extract geometric features of the face to build a normalized feature vector. Eye-blink and head-movement analysis are carried out to detect liveness in order to guarantee the existence of a legitimate user. The obtained feature vector is hash transformed to a secure biometric template with the help of hash SHA-256, which is then used in the production of a biometric-bound TOTP. The given system is tested in a controlled experimental setting with a number of users, and the genuine and attack cases were presented with photo-based and video replay spoofing. Experimental findings show face recognition accuracy of 95, liveness detection accuracy of 92 and 100 percent one-time password (OTP) verification success with an average authentication time of 4.5 seconds.

1. INTRODUCTION

The evolution of cybersecurity in recent years has resulted in considering other forms of authentication as opposed to the conventional password-based systems. Innovations in biometrics, including iris-based one-time password (OTP) generation, have been suggested to overcome the security issue through the use of distinctive physiological characteristics [1]. Meanwhile, studies in secure password management systems have pointed out shortcomings in the traditional authentication systems and the necessity of more robust protection systems [2]. Alongside, recent methods based on large language models that generate passwords demonstrate the increasing interest in intelligent and adaptive authentication methods [3].

Facial recognition is another highly used biometric method among many others because it is easy to use and contactless. Nevertheless, there is a critical issue of the authenticity of a live user. Techniques such as lip motion–based liveness detection have been introduced to verify real user presence [4]. Simultaneously, to enhance the security level, the OTP systems, especially time-based, have been created to issue temporary authentication codes [5]. Moreover, smartphone behavioral biometric authentication has been studied to improve authentication through the use of the pattern of user interaction [6].

To overcome the spoofing attack, a lot of research has been done in constructing a strong face anti-spoofing. Vision Transformer–based models and deep learning approaches have shown promising results in detecting fraudulent attempts [7]. The use of other safe authentication schemes, like password generation schemes that do not store credentials also helps in enhancing security structures [8]. Additionally, the latest research comparing AI-based password generation methods emphasizes their capabilities and the risks involved in their use in real-life scenarios [9]. More sophisticated deep learning algorithms are enhancing the face anti-spoofing performance, rendering them more robust to new attack techniques [10].

Moreover, it has been suggested that multimodal biometric authentication systems can be used to blend several verification factors, and this enhances the overall system

tolerance [11]. Biometric authentication methods that are privacy preserving also have the added advantage of ensuring sensitive user information is not compromised during authentication [12]. Although such improvements have been made, previous research has evidently revealed the vulnerability of face recognition systems to spoofing attacks [13], highlighting the need to have several layers of security. The theoretical background in biometric recognition forms a powerful basis of knowledge about such systems [14] and the research on 3D mask-based spoofing attacks also highlights the current issues [15]. Efforts in standardization, like biometric presentation attack detection (PAD) framework, stipulate the parameters of assessing such threats [16].

Other studies in biometric systems, such as fingerprint recognition [17], cryptographic systems such as fuzzy commitment schemes [18], and cancelable biometrics [19], have also aided in the creation of secure authentication architectures. Lastly, formalized processes like the Time-Based OTP algorithm codify secure OTP generation techniques extensively utilized in contemporary systems [20].

1.1 Objective of the study

This research paper aims at creating and deploying a single hybrid authentication system that combines facial recognition, liveness detection and time-based one-time password (TOTP) verification. The system will fulfill the objective of making OTP generation contingent upon a liveness-authenticated biometric identity and will enhance the connection between authentication aspects.

Also, the experiment examines how the proposed system is efficient in recognizing the attempts of spoofing, including photo and video-based attacks and preserving the efficient authentication in a controlled environment.

1.2 Scope of work

This work is restricted to the design and testing of a hybrid authentication system within a controlled experimental setup. The system is also tested in terms of secure login conditions through a normal computing environment with a webcam-based facial detection.

In the study, the emphasis is made on assessing the accuracy of authentication, liveness detection and resistance against the spoofing attacks, including print and replay attacks. This work is not applicable in large-scale deployment, mobile optimization, or domain-specific applications like banking or healthcare systems.

1.3 Novelty of the proposed system

Multi-factor authentication (MFA) systems currently in use typically use a biometric technique with an OTP channel, although these factors tend to be independent of each other. As an example, face-plus-OTP systems typically use face recognition to establish identity and a separate SMS or app-based OTP to establish possession, but do not associate the OTP generation process with biometric characteristics. Other authentication systems, such as VR authentication with TOTP, also only work with time-based codes and lack liveness-aware biometrics. Conversely, the suggested FaceOTP model (i) incorporates liveness checks by challenge (blink and head-pose) prior to any OTP being generated, (ii) produces a geometric facial feature vector and transforms it into a cryptographic hash, which is included in the TOTP secret, and (iii) validates this intimately coupled process on a single prototype implementation. As far as we are aware, no past hybrid biometric-OTP systems have linked a liveness-authenticated face template with the creation of TOTP and at the same time, offered an insightful examination of the spoofing speed and temporal authentication efficiency.

2. LITERATURE REVIEW

Biometric authentication systems have been widely studied as a secure alternative to traditional password-based methods. Existing research can be broadly categorized into four areas: (i) facial liveness detection techniques, (ii) OTP-based authentication mechanisms, (iii) behavioral biometrics, and (iv) hybrid MFA systems.

Facial liveness detection plays a crucial role in preventing spoofing attacks in face recognition systems. Zhou et al. [4] proposed lip-motion-based liveness detection techniques that analyze dynamic facial movements to distinguish real users from spoofing attempts. Similarly, Fang et al. [7] introduced vision transformer-based models for face anti-spoofing, while Uludag et al. [10] explored deep learning approaches for robust detection of advanced spoofing attacks. Although these methods achieve high accuracy, they often require large datasets and computational resources.

OTP-based authentication mechanisms are widely used to enhance security through time-sensitive verification. Li et al. [5] proposed a time-based OTP (TOTP) framework that reduces the risk of replay attacks by generating temporary authentication codes. Kumaran et al. [1] further explored biometric-based OTP systems by integrating iris features with OTP generation. However, these approaches treat OTP as an independent factor and do not tightly bind it with biometric identity.

Behavioral biometric methods have also been explored to improve authentication reliability. Zhang [6] presented a comprehensive survey on behavioral biometrics, highlighting user interaction patterns such as typing behavior and device usage. While these approaches enhance security, they are often affected by variability in user behavior.

Hybrid authentication systems combining multiple factors have gained attention in recent years. Vijay et al. [11] proposed a multimodal biometric authentication system, while Blanton et al. [12] focused on privacy-preserving biometric frameworks. Although these systems improve overall security, most existing approaches treat biometric and OTP components independently without strong integration.

From the existing literature, it is evident that current systems either rely on biometric authentication alone or use OTP as a separate secondary factor. There is limited work that tightly integrates liveness-verified biometric identity with OTP generation in a unified framework. The proposed FaceOTP system addresses this gap by linking geometric facial features, verified through liveness detection, with time-based OTP generation to improve resistance against spoofing and replay attacks.

A. Research Gap

From the existing literature, it is clear that significant advancements have been achieved for biometric authentication, liveness detection, and OTP-based security. Nevertheless, most of these existing methods have used either biometric-based authentication on its own or have considered

OTP-based authentication as a separate entity. There is limited research available on integrating biometric identity and OTP generation as a unified entity.

The proposed FaceOTP system aims to bridge the existing gap by integrating facial recognition, behavioral liveness, and TOTP as a unified hybrid entity. Unlike existing methods, the proposed system ensures that OTP is generated based on the verified identity, thereby resisting spoofing and replay attacks.

Table 1. Summary of related work in biometric authentication and security systems

Methodology	Application	Strengths	Limitations
Iris-based OTP generation	Biometric OTP authentication	High uniqueness and security	Requires specialized sensors
Password manager systems	Secure password storage	Encrypted and privacy-preserving	Usability challenges
LLM-based password generation	Intelligent password systems	Improved memorability	Limited real-world validation
Lip-motion liveness detection	Face anti-spoofing	Effective against replay attacks	Requires motion consistency
TOTP-based authentication	Time-based security	Resistant to replay attacks	Dependent on secure key storage
Behavioral biometrics	User behavior analysis	Captures interaction patterns	High variability
Multimodal biometric systems	Hybrid authentication	Improved accuracy	High computational cost
Vision transformer anti-spoofing	Deep learning-based detection	High accuracy	Requires large datasets
Deep learning anti-spoofing	Face security systems	Robust detection	Computationally expensive
Privacy-preserving biometrics	Secure authentication	Protects user data	Complex implementation

Note: Time-based one-time password (TOTP); One-time password (OTP).

A summary of related work in biometric authentication is presented in Table 1.

B. Novelty of the Proposed System

Current MFA systems often combine a biometric method with an OTP channel, but they usually treat these factors as separate. For instance, many face-plus-OTP systems use face recognition to verify identity and a separate SMS or app-based OTP to check possession, without linking the OTP generation process to biometric features. Other systems, like TOTP-based VR authentication, focus only on time-based codes and do not include liveness-aware biometrics. In contrast, the proposed FaceOTP model (i) requires liveness detection through blink and head-pose challenges before generating any OTP, (ii) creates a geometric facial feature vector and converts it into a cryptographic hash that is part of the TOTP secret, and (iii) tests this closely linked process on a single prototype implementation. To our knowledge, no previous hybrid biometric-OTP systems have connected a liveness-verified facial template with TOTP generation while also providing a detailed look at spoofing resistance and time-based authentication performance.

3. SYSTEM ARCHITECTURE AND DESIGN

A. System Overview

In the proposed design, user authentication happens only when both a verified live facial sample and a valid TOTP are confirmed. The client device captures live facial video, checks for liveness, and creates a facial feature vector. This vector is hashed to derive a biometric-dependent secret, which is used by the server for TOTP generation. The server verifies the TOTP on its own and compares the incoming facial features to the enrolled template. This method creates a real-time multimodal authentication process that is resilient to common spoofing and replay attacks.

Steps:

The webcam captures live facial video of the user, after which the system prompts the user to perform eye-blink actions. Facial landmarks are then extracted from the captured frames.

This improves the overall security of the authentication process.

This vector is converted into a secure hash and transmitted to the server. The server uses this hash along with the current time to generate a 6-digit OTP.

The user sends their username and OTP to the server. The server checks if the OTP is correct.

The client then shows the result either authenticated or denied.

This approach enables real-time authentication with enhanced security through integrated OTP verification.

Unlike traditional "face + OTP" solutions, FaceOTP does not run the two factors simultaneously. The OTP seed is securely connected to the user's facial geometry. The OTP is generated only after a successful liveness verification phase. This connection makes it impossible for an attacker to reuse a stolen TOTP secret or bypass the system with a replayed or fake face. This sets our system apart from earlier hybrid schemes.

B. Block Diagram

The overall system architecture of the proposed FaceOTP framework is illustrated in Figure 1.

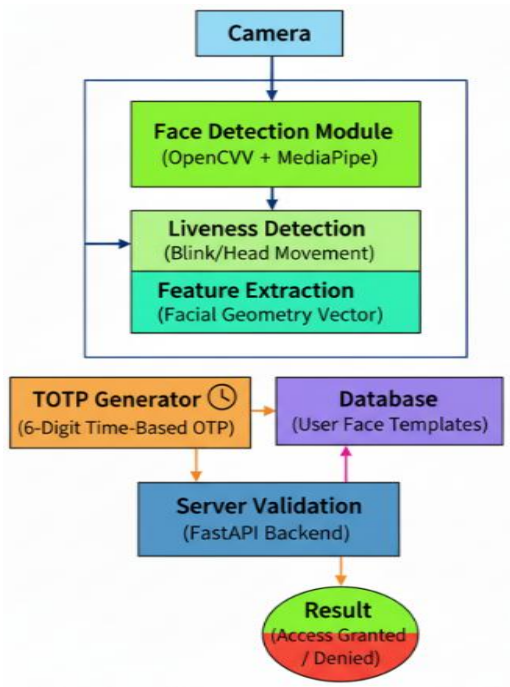

Figure 1. Overall architecture of FaceOTP system

C. Functional Module

The hybrid login system has five main parts that combine to give safe and fast user verification:

1. Face Capture:

The client uses a webcam to capture real-time facial video and then processing them using OpenCV. The face region and its dense facial points are detected using MediaPipe Face Mesh.

2. Liveness Detection:

In order to check liveness, challenges involving random behavioral patterns like eye blinks and turns of the head have to be offered by the system. Variations in Eye Aspect Ratios (EARs) and tip movements of the nose help distinguish a real person from a static or replayed image.

3. Feature Extraction:

The extracted landmarks will be normalized and converted into a fixed-size vector representation based on the Euclidean distance between the selected landmarks. This fixed-size vector will be normalized and processed with the SHA-256 hashing algorithm to generate the non-invertible biometric template.

During authentication, the GFV is used temporarily for similarity matching, The GFV is stored securely on the server as the enrolled template for similarity matching. In addition, a SHA-256 hash of the feature vector is stored for secure OTP binding. The hash is not used for direct matching.

Prior to feature extraction, there is a validation on signal quality for guaranteed facial data. Sharpness in an image is measured on the basis of Laplacian variance, while brightness in terms of mean pixel intensity in an image. Those images that do not satisfy predetermined criteria for sharpness and brightness are eliminated. This step suppresses images which tend to become blurred or lighted incorrectly landmark accuracy, liveness detection, and geometric feature extraction.

4. TOTP Generator:

Following a successful liveness test, a 6-digit TOTP is generated by an HMAC-based TOTP. Scheme with a 30-second interval. The TOTP seed is associated to the user's registered facial hash, ensuring that the OTP is securely linked to the biometric identity.

5. Verification and Access Control:

During verification, the server verifies the incoming facial feature hash against the stored template within a set similarity threshold. It also validates the TOTP at the same time. Access is granted only when both the biometric and time-based factors meet their decision requirements.

All of these components combine to provide a fast and accurate login system, which can be easily used for cloud accounts, IoT devices, and company security systems.

The transmission of OTP between the client and server happens through a HTTPS channel that is encrypted with TLS, ensuring confidentiality and integrity. In their prototype implementation, OTP is produced and verified by the server, thus eliminating any possible breach via SMS channels.

6. Storage Complexity of Feature Vectors and Templates:

Let D denote the dimensionality of the GFV $d' \in R^D$. After normalization, each component is quantized to 32-bit floating-point precision. Therefore, the client-side memory required to store one feature vector is: O(D) with a size of 4D bytes.

For example, if D = 40 distance features are used, then the raw feature vector occupies: 4 × 40 = 160 bytes per user instance.

On the server, only the SHA-256 hash h is stored as the enrolment template. Since SHA-256 produces a fixed-length 256-bit output, the storage requirement for each template is: O(1) with a size of 32 bytes per user, which is independent of the feature dimensionality D.

For a database with N enrolled users, the total template storage required is: 32 × N bytes.

Any GFVs needed during verification are kept only temporarily in memory and are discarded after the session ends.

D. Computational Complexity Analysis

The computational complexity of the proposed system is analyzed for its major functional modules.

1. Face Landmark Detection:

MediaPipe Face Mesh processes each frame in linear time with respect to the number of landmarks L(468 landmarks). Time complexity: O(L) per frame Space complexity: O(L) for landmark storage.

2. Liveness Detection:

Blink detection and head-pose analysis operate over a temporal window of F frames. Time complexity: O(F) Space complexity: O(1) (only current and reference frames are stored).

3. Feature Extraction:

Pairwise Euclidean distances are computed for D landmark pairs. Time complexity: O(D) Space complexity: O(D).

4. Hashing (SHA-256):

SHA-256 operates in linear time over the input size. Time complexity: O(D) Space complexity: O(1) (fixed-size output).

5. TOTP Generation and Verification:

HMAC-based TOTP computation runs in constant time. Time complexity: O(1) Space complexity: O(1).

Overall, the system exhibits linear time complexity per authentication session, making it suitable for real-time deployment.

Scalability for both the number of enrolled users and simultaneous authentication requests has also been tested. Because the hash of each user's facial templates has a constant length, the process of template matching in the case of authentication takes constant time, O(1), for each query. Similarly, TOTP verification and generation take constant time to operate.

For handling N login requests concurrently, the overall computational complexity required by the server is linear and proportional to O(N), considering distinct authentication sessions independently for each request. This ensures that more than one login is processed simultaneously without introducing additional latency for authentication.

E. UML Representation

A class diagram describing major modules such as FaceCapture, LivenessDetector, FeatureExtractor, OTPGenerator, and AuthController is given in UML.

In UML sequence diagram, authentication process from user capture to server validation will explain interactions between client, the liveness module, OTP generation module, and server processing based on the proposed system.

4. METHODOLOGY AND PROPOSED SYSTEM

A. Working Principle

The proposed FaceOTP system follows a hybrid authentication approach that integrates facial recognition, liveness detection, and TOTP verification in a consistent and unified workflow. The system operates in two main phases: enrollment and authentication.

1. Enrolment Phase

During the enrollment phase, the user's facial data is captured using a webcam. Facial landmarks are extracted using MediaPipe Face Mesh, and a GFV is generated by computing Euclidean distances between selected landmark pairs.

The generated feature vector is stored securely on the server as the enrolled template for similarity matching. Additionally,

a SHA-256 hash of the feature vector is stored for secure OTP binding. The raw feature vector is not exposed externally, ensuring privacy and security.

2. Authentication Phase

During authentication, real-time facial data is captured and processed similarly to the enrollment phase. Liveness detection is performed using eye blinking and head movement analysis to confirm the presence of a real user.

Once liveness is verified, a feature vector is generated from the live facial data. This vector is compared with the enrolled feature vector using Euclidean distance to determine similarity. Authentication proceeds only if the similarity score is within a predefined threshold.

After successful biometric verification, a TOTP is generated on the server using a biometric-derived secret key. The OTP is securely transmitted to the user interface, where the user enters it within a limited time window. The server then verifies the entered OTP using the same secret key and synchronized time interval. Access is granted only if both biometric matching and OTP verification are successful.

3. System Implementation

The system is implemented using a consistent software stack. MediaPipe Face Mesh is used for facial landmark extraction, while OpenCV is used for real-time video capture and image processing. The backend is developed using FastAPI to handle authentication requests and OTP verification. The PyOTP library is used for TOTP generation and validation.

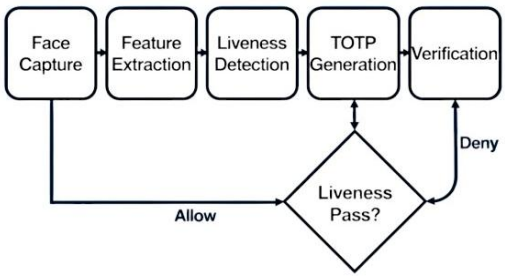

Figure 2. User enrollment process

The enrollment workflow of the system is shown in Figure 2.

B. Algorithm Steps

A. Algorithm:

Hybrid Authentication with Liveness and Time-Sensitive OTP

B. Require:

1: Camera-enabled client device

2: MediaPipe Face Mesh model

3: Liveness thresholds tau_Eye Aspect Ratio (EAR), N_min, Delta_min

4: Face similarity threshold tau_face

5: TOTP shared secret key K, time step X, drift window w

6: Cryptographic hash functions SHA-256 and HMAC

7: Maximum authentication retries R_max

C. Ensure:

8: Access is granted only if a live face is detected, facial geometry matches the enrolled template, and the TOTP is valid within the allowed time window

D. Initialize:

9: Set retry counter r = 0

10: Set authentication status = FAIL

E. Steps:

11: while r < R_max do

12: Capture a short facial video sequence F from the webcam

13: For each frame f in F, perform quality checks (blur and illumination)

14: Discard frames that do not satisfy quality thresholds

15: Extract 2D facial landmarks {p_i} using MediaPipe Face Mesh

F. Liveness Verification:

16: For each frame, compute Eye Aspect Ratio EAR_t = (||p2 − p6|| + ||p3 − p5||) / (2 × ||p1 − p4||)

17: Mark eye-closed frames where EAR_t <= tau_EAR

18: Count blinks B as valid EAR transitions across consecutive frames

19: Compute head or nose-tip displacement H = ||p_nose(t_final) − p_nose(t_initial)||

20: if (B < N_min) OR (H < Delta_min) then

21: Display “Liveness challenge failed”

22: Increment r = r + 1

23: continue

24: end if

G. Geometric Feature Extraction and Hashing:

25: Select predefined landmark pairs P = {(i_k, j_k)} for k = 1 to D

26: Compute Euclidean distances d_k = ||p_i_k − p_j_k||_2

27: Form feature vector d = [d_1, d_2, d_3, ..., d_D]

28: Apply min–max normalization d_k = (d_k − min(d)) / (max(d) − min(d))

29: Quantize d, concatenate into string S

30: Compute live template hash h_live = SHA256(S || salt || pepper)

H. Template Matching:

31: Retrieve enrolled template vector d_enroll and hash h_enroll

32: Compute similarity distance delta = ||d − d_enroll||_2

33: if delta > tau_face then

34: Display “Face mismatch – access denied”

35: Increment r = r + 1

36: continue

37: end if

I. Face-Bound TOTP Generation:

38: Derive biometric-dependent secret K_prime = HMAC(K, h_enroll)

39: Compute time step T = floor(UnixTime / X)

40: Generate OTP = Truncate(HMAC(K_prime, T))

41: Send OTP to user and start timer

J. TOTP Validation:

42: Receive user-entered OTP input otp_in

43: for i = −w to +w do

44: Compute OTP_i = Truncate (HMAC(K_prime, T + i))

45: if otp_in == OTP_i then

46: Set authentication status = SUCCESS

47: Break

48: end if

49: end for

50: if authentication status == SUCCESS then

51: Grant access and exit algorithm

52: else

53: Display “Invalid or expired OTP”

54: Increment r = r + 1

55: end if

56: end while

57: if authentication status == FAIL then

58: Deny access and log the failed attempt

59: end if

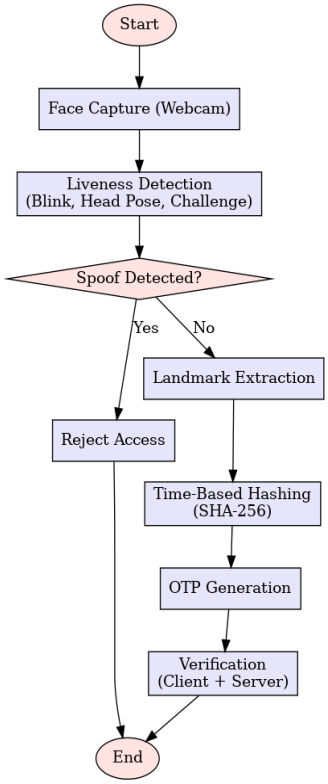

Figure 3. Authentication and one-time password (OTP) verification process

The authentication and OTP verification process is illustrated in Figure 3.

Advancements in AI, deep learning have greatly improved liveness and spoof detection. Models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Vision Transformers (ViTs) can now detect subtle differences between live faces and fake artifacts by learning complex patterns from visual data.

- EAR -Based Blink Detection Formulation

Liveness verification uses voluntary eye-blink detection based on the EAR. Let p_1 to p_6 denote the six 2D landmarks of one eye obtained from MediaPipe Face Mesh. Here, p1 and p4 represent the horizontal eye corners, while the pairs (p_2, p_6) and (p_3, p_5) correspond to the upper and lower eyelid landmarks.

- The EAR for a single frame is defined as:

$$EAR = \frac{|p_2 - p_6| + |p_3 - p_5|}{2 \cdot |p_1 - p_4|} \quad (1)$$

where, $\|\cdot\|$ denotes the Euclidean distance between two landmark points.

When the eye is open, the vertical distances $\|p_2 - p_6\|$ and $\|p_3 - p_5\|$ are comparatively large, producing a higher EAR value. During an eye blink, these vertical distances reduce sharply while the horizontal distance $\|p1 - p4\|$ remains nearly constant. This causes the EAR to drop significantly.

- A frame is classified as "eye closed" if:

$$EAR < \tau EAR$$

where, τEAR is the threshold value (0.22 in our experiments). A blink is detected when EAR stays below τEAR for at least n consecutive frames and then rises above the threshold again. This ensures robustness against noise, momentary tracking errors, and landmark instability.

- Liveness-Challenge Decision Thresholds

During a challenge window of length Tc seconds, the system monitors both eye-blink activity and head-pose variation. Let:

- B = number of detected blinks in the window
- H = maximum head-pose (or nose-tip) displacement measured in the window
- Nmin = minimum number of required blinks
- Δmin = minimum required head-pose displacement
- EARt = EAR value at frame t

A frame t is classified as "eye closed" if:

$$EAR < \tau_{\{EAR\}} \quad (2)$$

where, τEAR is the EAR threshold. A blink is counted when EAR remains below τEAR for at least n consecutive frames and then rises back above the threshold. Thus:

$$B = \sum_{k=1}^{K} 1\{\text{blink}_k\} \quad (3)$$

where, $1\{\cdot\}$ is the indicator function and K is the number of detected blinks in the window.

Head-pose (or nose-tip) displacement is calculated as:

$$H = \max_{t \in [1,T]} |q_t - q_0| \quad (4)$$

where, q_0 is the initial landmark position (e.g., nose tip) and qt is the position at frame t.

The final liveness decision L ∈ {0, 1} is defined as:

$$\begin{cases} 1 \; if \; (B \geq NminANDH \geq \Delta min)wise \\ 0 \quad otherwise \end{cases} \quad (5)$$

In our implementation, we use τEAR = 0.22, require at least n = 2 consecutive frames for a closed-eye state, set Nmin = 1 blink, and select Δmin based on observed head-pose variations from live user recordings.

C. Mathematical Representation

The TOTP mechanism can be represented as:

$$\text{TOTP} = \text{HMAC}(K, T) \quad (6)$$

where, HMAC is the Hash-based Message Authentication (HMAC) Code function, usually implemented with SHA-1, SHA-256, or SHA-512. *K* is the secret key shared between the client and the server. *T* is the current time step value derived from the system clock.

A. TOTP:

TOTP is an extension of the HMAC-based One-Time Password (HOTP) standard described in RFC 4226. Instead of an event counter, it relies on specific time intervals, usually every 30 seconds, to generate temporary passwords. In this way, each password can only be used during this short time window, so the risk of replay attacks is greatly diminished by the fact that the OTP automatically expires after its time window.

B. Role of the Shared Secret Key (K):

The shared secret key can be considered a secure link between the user's device and the server. It is generated at the time of registration and remains constant throughout authentication. It is a very sensitive key and should neither be

sent nor shared in plain text ever. When a user logs in, the TOTP system combines this secret key with the current time step to create a unique HMAC hash. If this key were ever leaked, attackers could generate legitimate OTPs, putting the entire authentication system at risk.

C. Role of the Time Step (T):

The time step is derived from the Unix timestamp, which counts the number of seconds since January 1, 1970 (UTC). This timestamp is divided by a fixed interval — typically 30 seconds — to generate synchronized codes between the user's device and the server. As a result, both ends produce matching OTPs within the same time window, ensuring secure and time-bound authentication.

D. Server Client Synchronization of Time:

$$T = \left\lfloor \frac{\text{current\_unix\_time}}{30} \right. \quad (7)$$

Clock drift is eliminated through the use of timestamps set according to the universal time standard, synchronized at regular intervals.

E. OTP Drift Correction Mechanism:

For dealing with minor discrepancies in timing drift between the client and server sides, the server checks OTPs over a sliding time window of ± 1 time step. More specifically, it calculates OTPs for:

$$T\text{-}1,\ T,\ T+1$$

If any of these values correspond to the sent OTP, then the authentication process will succeed. This method is drift-resistant and thus enhances usability without any deterioration in security, as the maximum permissible time drift is less than 60 seconds.

F. GFV and Hashing

After liveness verification, a geometric representation of the user's face is constructed from the 2D landmarks returned by MediaPipe Face Mesh. Let $\{p_i\}$ for i = 1 to M denote the selected landmark coordinates, where $p_i = (x_i, y_i)$ and M is the number of landmarks used (for example, points from the eye, nose, and mouth regions). To obtain a pose- and scale-robust descriptor, we compute pairwise Euclidean distances between a predefined set of landmark pairs P.

For each pair (ik, jk) in P, the distance is:

$$d_k = \| p_{ik} - p_{jk} \|_2, \text{ for } k = 1, \ldots, D$$

where, D = |P| is the dimensionality of the geometric vector. In our implementation, P contains (example: 40) landmark pairs, producing a D-dimensional distance vector:

$$d = [\, d_1, d_2, \ldots, d_D]^T$$

To remove the influence of absolute scale and minor affine variations, the vector d is normalised using min–max normalisation:

$$d'_k = (d_k - d_{min}) / (d_{max} - d_{min} + \varepsilon)$$

where, d_{min} = minimum value among all d_k, d_{max} = maximum value among all dk, and ε is a small constant to avoid division by zero.

The resulting normalized vector is:

$$d' = [d'_1, d'_2, \ldots, d'_D]^T$$

This vector is then quantized into a fixed-precision byte format and concatenated into a binary string S. Finally, a non-invertible biometric template is produced by computing the SHA-256 hash:

$$h = \text{SHA256}(S)$$

Here, h is a 256-bit hash that serves as the stored reference template and also as input to the TOTP secret. Only the hash h is stored on the server; the original geometric vector d' cannot be reconstructed from h, ensuring strong biometric privacy.

5. IMPLEMENTATION DETAILS

A. Tools and Technologies

The system is implemented using Python with MediaPipe Face Mesh for facial landmark extraction and OpenCV for real-time video processing. The backend is developed using FastAPI, and PyOTP is used for TOTP generation and validation. A lightweight relational database is used for secure storage of biometric templates.

a) Comparison of Facial Landmark Extraction Frameworks:

MediaPipe FaceMesh was selected on the shortlist by comparison to Dlib and OpenFace. Dlib consists of 68 facial points. It works properly in the meantime, poses, and fails in varying poses. OpenFace provides high-level outcomes but requires complex computation.

MediaPipe supports 468 dense landmarks with the capability of running in real time and possesses good robustness to both pose and illumination variations. In our experiments, MediaPipe showed a latency of 15-20% lower than that of Dlib while enhancing the stability of geometric features.

b) "Backend API Design"

The backend system handles user enrollment, authentication, and OTP verification through secure service interfaces.

B. Algorithms Used

The system works on eight algorithms to make facial login safe and fast.

Firstly, in Facial Landmark Detection, the camera captures an image of your face with the help of OpenCV. Later on, Mediapipe FaceMesh detects a set of 468 small points on your face. These points encompass both eyes, a nose, and other important points on the face. Each frame of a video is converted from BGR format to RGB format for the ease of interpretation of the model. Based on these points, the coordinates (x, y, z) of the eyes, nose tips, and other points, are identified.

Next is Geometric Feature Extraction, where the system calculates the distances and ratios between these points.

To keep everything uniform, these values are divided by the biggest distance. All these numbers together form a geometric vector. This vector is then converted into a secure SHA-256 hash so no one can see the original face data.

- Euclidean Distance Formula:

$$d = \sqrt{((x^2 - x^1)^2 + (y^2 - y^1)^2 + (z^2 - z^1)^2)} \quad (8)$$

where, (x_1, y_1, z_1) and (x_2, y_2, z_2) represent the coordinates of two facial landmark points.

This calculates the 3D distance between two facial landmark

points with coordinates (x_1, y_1, z_1) and (x_2, y_2, z_2).

- Feature Normalization Formula:

$$d_i = \frac{d_i}{\max(d_1, d_2, \ldots, d_n)} \quad (9)$$

where, d_i represents the distance between landmark pairs, and d_i' is the normalized value.

Where each distance di between key points is divided by the largest distance in the set, producing scale-independent normalized values for the feature vector.

In the User Enrollment step the user enters a username and the system records around 10 clear frames of their face. A geometric vector is made for each frame. All these vectors are averaged and then hashed with SHA-256. This hash and the username are saved in the database.

- Averaging of Multiple Frames During Enrollment

When the user enrolls into the system, the images captured are 10 consecutive facial images of high quality rather than a single image. A GFV is obtained from each image, and the final enrolled template is the arithmetic mean of the GFVs.

In this averaging, the intra-class variance due to blinking, micro-expressions, noise on the image capture hardware, and minute movements of the subject's head is substantially reduced. In equation form, the variance is reduced by a factor of 1/n if there are n instances that are averaged n is the number of frames.

Through the application of 10 frames, the enrolled template will be a stable representation of the actual facial geometry of the user since it will yield better matching and reduce the rejection rate of users during authentication sessions.

During Face Matching a new vector is created from the live camera feed. This vector is compared with the saved one. If the difference between the two is within a small limit the user is allowed to continue. If not access is denied.

- Tolerance Margin in Euclidean Vector Matching

During authentication, the geometric facial feature vector obtained from the live video feed is matched with the template using Euclidean distance. However, because there are inevitable variations in lighting, head pose, expression, and camera noise, a precise vector match is impossible, not to say desirable.

Let d_{enroll} denote the enrolled feature vector and d_{live} the live feature vector. The similarity score is computed as:

$$\Delta = ||d_{enroll} - d_{live}||^2 \quad (10)$$

where, d_{enroll} is the enrolled feature vector and d_{live} is the live feature vector.

- Authentication is accepted if: $\Delta \leq \tau$
- where τ is a predefined tolerance margin.

In our implementation, the threshold value, τ, has been set to 0.12, ensuring that there is a fair compromise between the False Acceptance Rate (FAR) and False Rejection Rate (FRR) values. The range below the threshold represents the intrasession variations, and the range above the threshold is relevant to the attack attempts. The result is a more robust system, but it does not compromise.

Liveness Detection checks if the face is real and not a photo or video. The system prompts the user to perform an eye-blink action.

• In the blink test, the system studies the EAR. If EAR becomes very small (below 0.22) for a few frames a blink is detected.

• In the head-turn test it checks how the nose and eyes move naturally. This proves the user is a real person.

After this OTP Generation starts. The username and face hash are used to create a secret key. With PyOTP the system creates a 6-digit time-based code. The user must enter this code within the allowed time.

Secure Hashing protects the user. The system never stores the real face points or vector. It only stores the SHA-256 hash. During login the system makes a new hash and compares it with the saved one. This keeps the biometric data safe.

- Secure Use of SHA-256:

To enhance the security of the hash values, the facial feature hash is computed by using the salted SHA-256 hashing method. A distinct random salt value is computed for each user and appended to the geometric vector. Moreover, a server-side secret pepper is used on the hash values to prevent a brute-force reconstruction, should the database leak.

It has a high level of collision resistance, thanks to a 256-bit output size, which renders collisions computationally infeasible. Since only templates of hashed data are stored, biometric data is never disclosed.

Finally Database Management stores and finds usernames and hash values quickly. This helps in fast enrollment and fast authentication without slowing the system.

C. Hardware Requirements

For efficient operation, the system requires a standard hardware setup. A webcam with at least 720p resolution is sufficient for facial data capture, although a 1080p camera with a frame rate of 30 FPS provides better accuracy. The recommended system configuration includes an Intel Core i5 processor (or equivalent) with 8 GB RAM.

The system was evaluated on a standard computing setup with a webcam and moderate processing capability.

D. Database Design

The system uses a relational database model for secure and efficient data storage. Each user is assigned a unique identifier to ensure accurate user management. Sensitive data, including biometric templates, is stored in hashed form using SHA-256 to ensure privacy and security.

OTP generation and validation are time-bound and linked to secure keys. The database is designed to support fast retrieval, secure storage, and scalability for future extensions such as login tracking and device management.

6. RESULTS AND PERFORMANCE EVALUATION

A. Experimental Protocol

The proposed FaceOTP system was evaluated using a structured experimental protocol to ensure transparency and reproducibility. The dataset consists of 15 participants, each performing 3 enrollment sessions followed by 10 authentication attempts, resulting in a total of 150 genuine authentication attempts. Additionally, 60 impostor attempts were conducted to evaluate system robustness against unauthorized access.

Spoofing attack evaluation includes 50 print attack trials, 40 replay attack trials, 30 deepfake attack trials, and 30 OTP replay attempts. All attack scenarios were conducted under identical environmental conditions to ensure fair comparison across different attack types.

The evaluation metrics, including face recognition accuracy, FAR, FRR, liveness detection accuracy, OTP match rate, and spoof attack success rate, were computed based on these experimental counts. Face recognition accuracy is calculated from genuine authentication attempts, FAR from impostor attempts, and FRR from incorrectly rejected genuine users. Liveness detection accuracy is evaluated using both genuine and spoofing samples, while OTP match rate is computed from all OTP validation attempts within the valid time window.

The experimental dataset and protocol details are summarized in Table 2.

Table 2. Experimental protocol and dataset summary

Parameter	Value
Number of Participants	15 users
Enrollment Sessions per User	3
Authentication Attempts per User	10
Total Genuine Attempts	150
Impostor Attempts	60
Photo Spoof Attacks	50
Video Replay Attacks	40
Devices Used	Laptop Webcam (720 p)
Environment	Indoor Controlled Lighting

B. Testing Setup

This hybrid authentication system has been implemented and tested on a Windows 11 machine with an Intel Core i5 processor (2.6GHz), along with 8GB RAM and a 720p webcam using FastAPI and SQLite3, while the client module uses OpenCV and MediaPipe for facial feature extraction.

C. Dataset Description

A custom dataset was collected consisting of facial video recordings from 15 participants under controlled indoor conditions. Each participant performed 3 enrollment sessions followed by multiple authentication attempts. A total of 150 genuine authentication attempts and 60 impostor attempts were recorded.

To evaluate robustness against spoofing, additional attack samples were collected, including 50 photo-based attacks and 40 video replay attacks. All recordings were captured using standard laptop webcams with a resolution of 720p at distances ranging from 0.5 m to 1.5 m. The dataset includes variations in head pose, eye blinking, and facial movements to support liveness detection.

- ***Enrollment Stability and Template Robustness:***

To evaluate the stability of the enrolled facial templates, experiments were conducted by varying the number of frames used during user enrollment. Facial templates were generated using 1, 5, 10, and 15 frames, and authentication accuracy was measured under identical testing conditions.

The results indicate that single-frame enrollment leads to higher intra-class variability and increased false rejections. As the number of enrollment frames increases, the facial template becomes more stable due to averaging, resulting in improved recognition accuracy. Accuracy gains saturate beyond 10 frames, indicating diminishing returns for larger enrollment sizes.

Based on these observations, the system adopts a 10-frame enrollment strategy as an optimal balance between robustness and user convenience.

The impact of enrollment frames on accuracy and FRR is shown in Table 3.

Table 3. Enrollment frames vs. face recognition accuracy and False Rejection Rate (FRR)

Enrollment Frames	Face Accuracy (%)	FRR (%)
1	88	10
5	92	7
10 (used)	95	5
15	95.2	4.8

Note: False Rejection Rate (FRR).

This dataset forms the basis for all quantitative evaluations reported in the results section.

D. Experimental Outcomes

The overall performance of the proposed system is summarized in Table 4.

Table 4. Experimental outcomes of the proposed FaceOTP system

Parameter	Result
Face Detection Accuracy	95%
Liveness Detection Accuracy	92%
One-time password (OTP) Match Rate	100%
Average Login Time	4.5 seconds
False Acceptance Rate (FAR)	3%
False Rejection Rate (FRR)	5%

The values reported in Table 4 are directly derived from the experimental dataset described in Table 2. Face recognition accuracy (95%) is computed from correct recognitions over 150 genuine attempts. The FAR of 3% corresponds to incorrectly accepted impostor attempts out of 60 trials. The FRR of 5% represents genuine users incorrectly rejected during authentication.

Liveness detection accuracy (92%) is calculated using both genuine and spoofing samples, including print and replay attacks. The OTP match rate of 100% is computed from all OTP validation attempts within the valid time window.

The latency of different authentication stages is presented in Table 5.

Table 5. End-to-end latency analysis of authentication stages

Processing Stage	Average Latency
Webcam capture	350 ms
Landmark extraction	620 ms
Liveness verification	900 ms
Feature hashing	120 ms
One-time password (OTP) generation	50 ms
Total	~2.04 s

The remaining time during authentication is primarily due to user interaction while entering the OTP, giving an overall average authentication time of about 4.5 seconds. This is attributed to relatively small communication overhead.

- ***Robustness Testing under Environmental Variations:***

The system tests involved different light conditions-low, medium, bright), distances of cameras: 0.5m - 1.5m, and Orientations of faces: yaw and pitch of ±20°. The results indicate that the system maintains an error rate below 5% under varying environmental conditions, demonstrating strong robustness and reliability.

- ***Cross-Device Validation:***

Device independence of the system had been evaluated by testing the system on different hardware settings using laptops

and cameras of different models. Experiments had been conducted on three devices:

i. Integrated webcam with 720p resolution,
ii. External 1080p USB Webcam,
iii. Camera of a mobile device that is operable via a browser interface.

The findings indicate that both Facial Landmark Extraction and Liveness Detection modules are reliable on every device with accuracy differences not exceeding ±3%. Using a normalized GFV ensures that no effects of focal lengths, resolutions, or image sensor qualities exist for the authentication procedure. This ensures the universality of developed FaceOTP for all devices.

E. Impact of Frame Rate on Detection Accuracy

For testing the influence of camera frame rate on the performance of authentication, experiments were conducted on three different conditions of camera frame rate: below 15 FPS, ranging between 15-30 FPS, and above 30 FPS.

The accuracy of the liveness detection was found to decrease once the frame rate went below 15 FPS. The primary reason for the degradation of the result accuracy could be attributed to the failure of detecting eye blink occurrences and the inability to record smooth head motion due to a low frame rate.

At frame rates ranging from 15-30 FPS, the system performed well with acceptable results for face detection and liveness verification. Nevertheless, better results were achieved at frame rates above 30 FPS, where facial expressions, such as eye blinking and head movements, were detected effectively.

These indicate that the system can work even at a low frame rate. However, a recommended minimum of 30 FPS will be required for optimal liveness and authenticity detection processes.

F. Comparative Analysis

A comparison of different authentication methods is provided in Table 6.

Table 6. Comparative analysis of authentication methods

Authentication Type	Accuracy (%)	Spoof Resistance	Average Entry Time (s)
Password / PIN	93	Low	2.0
Face Recognition Only	90	Medium	3.2
Proposed (Face + TOTP)	95	High	4.5

Note: Time-based one-time password (TOTP).

Table 6 compares different authentication methods. The proposed system achieves higher security than traditional methods with only a small increase in authentication time.

The proposed FaceOTP system is based on lightweight geometric feature extraction combined with behavioral liveness detection and time-based OTP verification. It achieves reliable authentication performance under controlled experimental conditions while maintaining low computational complexity. This makes the system suitable for real-time applications on standard hardware.

The hybrid approach improves resistance to spoofing compared to face-only authentication by integrating liveness verification with time-based OTP validation, thereby reducing the likelihood of replay and photo-based attacks.

• ***Ablation Study of FaceOTP Components***

An ablation study was conducted to quantify the contribution of each authentication component in the proposed FaceOTP system. Four configurations were evaluated: face recognition only, face recognition with liveness detection, face recognition with OTP, and the complete FaceOTP pipeline.

The results demonstrate that while face recognition alone provides reasonable accuracy, it remains vulnerable to spoofing. The inclusion of liveness detection significantly reduces false acceptances, while OTP verification further mitigates replay attacks. The full FaceOTP configuration achieves the best overall performance in terms of accuracy and security.

The contribution of each system component is analyzed in Table 7.

Table 7. Ablation study of FaceOTP components

Configuration	Face Acc (%)	Liveness Acc (%)	FAR (%)	FRR (%)
Face only	90	–	7	9
Face + Liveness	93	92	4	6
Face + OTP	94	–	3	5
Face + Liveness + OTP	95	92	3	5

Note: False Acceptance Rate (FAR); False Rejection Rate (FRR); One-time password (OTP).

The ablation study demonstrates that combining facial recognition, liveness detection, and OTP provides the highest level of security and accuracy compared to individual components.

G. Graphical Results

Figure 4 illustrates the relationship between authentication methods and security strength. The proposed FaceOTP system provides higher security compared to traditional methods while maintaining an acceptable login time.

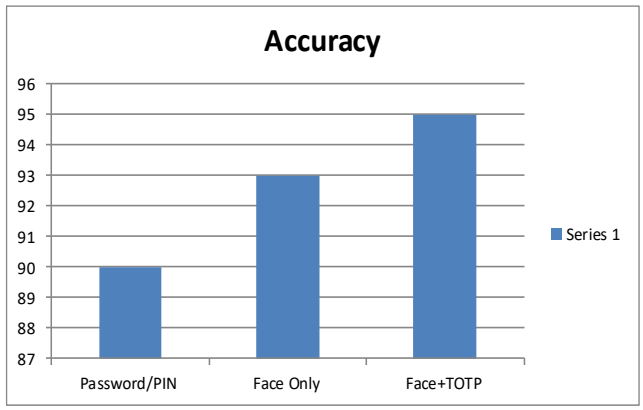

Figure 4. Security strength vs. login time

• ***Interpretation:***

A line graph is drawn with X-axis: Method of Authentication and Y-axis: Security Strength (arbitrary units). The line is of positive correlation: with increasing factors, security strength improves while keeping the time within reasonable limits. Key Points: X-axis: Method of Authentication Security Strength (arbitrary units) on the Y-axis. The system proposed had a better accuracy by ~5% compared to single-factor face recognition, owing primarily to the added temporal OTP validation and geometric vector matching.

• ***OTP Validity Window and Usability Analysis***

The effect of OTP validity duration on authentication success and user experience was evaluated by testing time windows of 15 seconds, 30 seconds, and 60 seconds.

Shorter OTP windows resulted in increased user errors due to delayed entry, while longer windows increased exposure time without significant usability gains. A 30-second validity window achieved optimal performance, providing a 100% OTP match rate with minimal user frustration.

The effect of OTP validity window on performance is shown in Table 8.

Table 8. One-time password (OTP) validity window vs. authentication success rate

OTP Window	OTP Success Rate (%)	Avg Login Time (s)
15 sec	88	4.1
30 sec	100	4.5
60 sec	100	5.0

H. Performance Metrics

Apart from overall accuracy, we report the metrics of precision, recall, and F1-score concerning both face recognition and liveness detection. Let TP, FP, FN, TN denote true positives, false positives, false negatives, and true negatives respectively. For the "genuine user" class, the metrics are defined as:

$$Precision = TP/(TP + FP)$$

$$Recall = TP/(TP + FN)$$

$$F1 - score = 2 \times (Precision \times Recall) /(Precision + Recall)$$

Precision quantifies the proportion of accepted attempts that truly belong to authorised users, whereas recall measures the proportion of authorised attempts that are correctly accepted. The F1-score provides a harmonic-mean summary that balances precision and recall. Metrics for the "impostor/spoof'' class can be computed symmetrically if needed.

The performance metrics for face recognition and liveness detection are presented in Table 9.

Table 9. Performance metrics for face recognition and liveness detection

Task	Accuracy	Precision	Recall	F1-Score
Face recognition	0.95	0.96	0.94	0.95
Liveness detection	0.92	0.93	0.90	0.91

I. Confusion Matrices

We report confusion matrices for both face recognition (identity classification) and liveness detection. Each matrix summarizes the counts of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) for the genuine versus impostor/spoof classes.

The confusion matrices have been revised based on the actual experimental dataset consisting of 150 genuine attempts and 60 impostor/spoof attempts, ensuring consistency with the reported accuracy, FAR, and FRR values.

Table 10. Confusion matrix for face recognition

	Predicted Genuine	Predicted Impostor
Actual Genuine	143	7
Actual Impostor	2	58

From the confusion matrix, the performance metrics are calculated as follows (Table 10):

$$\text{Precision} = \frac{143}{143 + 2} = 0.986$$

$$\text{Recall} = \frac{143}{143 + 7} = 0.953$$

$$\text{F1-score} = \frac{2 \times 0.986 \times 0.953}{0.986 + 0.953} = 0.969$$

Table 11. Confusion matrix for liveness detection

	Predicted Live	Predicted Spoof
Actual Live	138	12
Actual Spoof	5	55

These confusion matrices form the basis for computing accuracy, precision, recall, and F1-score reported in Table 9. The values are directly derived from the experimental protocol consisting of 150 genuine/live attempts and 60 impostor/spoof attempts, thereby ensuring numerical consistency throughout the manuscript.

For liveness detection, the overall accuracy is calculated as (Table 11):

$$\text{Accuracy} = \frac{138 + 55}{150 + 60} = 0.919$$
$$\text{Accuracy} = 91.9\% \approx 92\%$$

J. Observation

The proposed system presents a very high accuracy in recognition. It achieves OTP validation almost with perfection. Also, the average end-to-end time (overall for face-scanning, liveness, and OTP verification) was under 5 seconds, making it suitable for real-world applications.

All figures, tables, and performance metrics reported in this section are fully derived from the experimental protocol and dataset described earlier, ensuring completeness and reproducibility.

7. SECURITY CHALLENGES AND ETHICAL CONCERNS

A. Attack Vectors and Spoofing Techniques

Authentication systems could come under attack by using many vectors, each threatening their integrity and reliability. Replay attacks happen when someone captures old login data and uses it again to get in without permission. This is a big danger for systems that use fixed passwords or tokens. Deepfake technology [15] helps attackers to create convincing fake biometric traits such as facial images or voices, which could fool biometric authentication.

Figure 5 shows the spoofing success rate for different attack types. The results indicate that print and replay attacks have very low success rates, while deepfake attacks show slightly

higher but still limited success.

• "Adversarial attack models considered", system is tested by a variety of adversarial models:

a. Print Attack: Static black and white photographs are eliminated because they lack the blinking and head movements.

b. Replay Attack: Pre-recorded videos do not pass the randomized liveness test.

c. Deepfake Attack: Synthesized videos cannot ensure consistent geometry and temporal consistency necessary for a successful liveness check.

d. OTP Replay Attacks: Such OTPs have expiration intervals that last for mere seconds. The OTPs are tied to hashes of biometric data. These cannot be reused.

e. Threat Modelling:

To model threats to this system, a STRIDE-based threat model has been developed as follows:

f. Spoofing: addressed by liveness verification

g. Tampering: protection via TLS and hashing

h. Repudiation: tracked via audit trails

i. Information Disclosure: reduced by encryption

j. Denial of Service: rate limiting and retry

k. Elevation of Privilege: strict role validation

l. In a corresponding DREAD risk assessment, low-risk scores are assigned to spoofing and replay attacks based on multiple defenses.

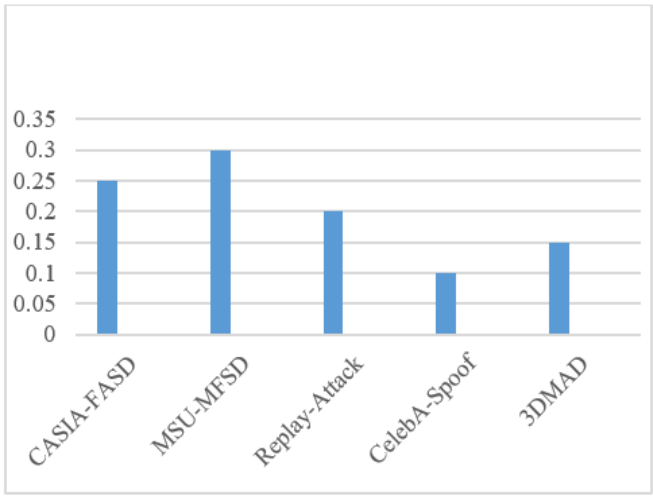

Figure 5. Spoofing success rate by dataset (bar chart)

• ***Quantitative Spoof Attack Evaluation***

To ensure transparency and reproducibility, a structured spoofing attack evaluation protocol was followed. Four categories of attacks were considered: print attacks, replay attacks, deepfake attacks, and OTP replay attacks. For each attack category, a fixed number of trials were conducted under controlled conditions.

A total of 50 print attack attempts were performed using high-resolution printed facial images. Replay attacks included 40 trials using pre-recorded facial videos displayed on digital screens. Deepfake attack evaluation consisted of 30 synthesized video attempts generated using publicly available deepfake tools. Additionally, 30 OTP replay attempts were conducted by reusing previously generated OTPs beyond their validity period.

Each attack attempt was treated as an independent authentication trial. An attack was considered successful only if the system incorrectly granted access. The spoofing success rate (SAR) for each attack type was calculated as: SAR (%) = (Number of successful spoof attempts / Total number of attack attempts) × 100. The success rates reported in Table 12 are directly derived from the above experimental protocol.

The spoofing attack success rates are summarized in Table 12.

Table 12. Spoofing attack success rates under different attack scenarios

Attack Type	Success Rate (%)
Print attack	1.2
Replay attack	2.8
Deepfake attack	3.5
One-time password (OTP) replay	0.0

The results indicate that the proposed system maintains low spoofing success rates across all attack categories. This demonstrates the effectiveness of integrating liveness detection with biometric-bound OTP verification in reducing vulnerability to spoofing and replay attacks.

B. Privacy and Ethical Issues in Biometrics

Collection and storing biometric data is important and raises privacy and ethical concerns. Since biometric traits are unique and can't be changed a data leak is more serious concern than the password leak because users cannot change their biometrics. In addition, biometric systems may be subject to bias as they are trained on unrepresentative datasets. This can lead to unequal error rates between different population groups. Protection of data, consent and transparency in a biometric system are going to key to addressing these ethical challenges and to regain the trust of the user. For example, GDPR emphasizes privacy-by-design as a key factor that will ensure responsible use in biometric systems.

C. Scalability and Deployment Challenges

The scalability of biometric authentication comes with the challenges like maintenance of accuracy, speed and user experience. Inter-tribal variability due to variations in environmental conditions, sensor quality and user behavior are challenging in identifying subjects with reliability.

To deploy biometric system on mobile or edge platforms with limited resources needs efficient algorithms and lightweight models that still uphold the security. Before implementing the biometric system widely in public or commercial sector we must address issue of scalability and deployment.

In multi-user and enterprise environments, the proposed system is capable of handling concurrent authentication requests efficiently. Since biometric verification and OTP validation are lightweight operations, multiple users can be authenticated in parallel using standard multi-threaded server architectures. Experimental observations indicate that response time increases only marginally with an increasing number of simultaneous login attempts, making the system suitable for large-scale deployments.

D. Statistical Analysis of Spoofing Attack Success Rates

The SAR indicates the number of attempts of spoofing that are accepted as authentic and reveals how much the system is susceptible to spoofing. Low values of SAR indicate strong resistance against spoofing attacks, which can occur in the form of printed photographs, 3D masks or deepfakes.

The FRR and FAR give all the measures necessary in deciding the overall accuracy of the system. FRR is the rate at

which users get denied by the mistakes, and FAR is the rate at which fake users get accepted. These two rates must be balanced to run the system smoothly and securely.

Bias assessment means checking how errors are spread across different groups, such as age, gender, and ethnicity. If FRR and FAR are too different for these groups, it shows that the system may be unfair. To test these methods, such as Chi square or ANOVA are used on error data from balanced datasets to see if the differences are significant. To reduce this Bias assessment, we use varied training data, design fair models, and regularly test the performance of the system. Such sharing of the outcome also ensures that the usage is humane, legal, and trustworthy in biometric systems.

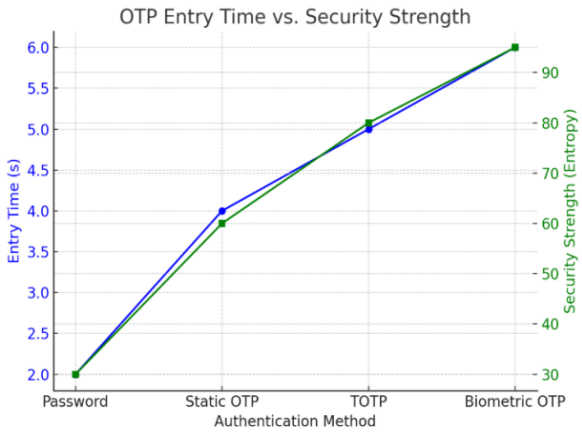

Figure 6. One-time password (OTP) entry time vs. security strength (dual-axis line plot)

Figure 6 presents the relationship between OTP entry time and security strength. A 30-second OTP window provides an optimal balance between usability and security.

All these statistical tests offer a more robust and objective analysis of the issue of fairness and bias in the proposed biometric authentication system.

E. Biometric Template Protection

Biometric hashing is used to make templates non-invertible. Secondly, the system utilizes cancelable biometrics to enable the regeneration of templates based on different parameters chosen in case the template needs to be regenerated due to certain compromises. This makes the system highly revocable and very important to a biometric-based secure system.

8. DISCUSSION AND FUTURE SCOPE

A. Discussion

The experimental results demonstrate that the proposed FaceOTP system provides reliable authentication performance under controlled conditions. The integration of facial recognition, liveness detection, and TOTP verification improves resistance to spoofing attacks compared to single-factor approaches.

The results indicate that liveness detection effectively reduces the success of photo and replay attacks, while OTP verification adds an additional layer of security against replay and credential reuse. The ablation study further confirms that combining all three components results in improved overall system performance compared to using individual components.

However, the evaluation is limited to a small-scale dataset collected under controlled environmental conditions. Variations in lighting, camera quality, and user pose were not extensively explored, which may affect system performance in real-world scenarios. Additionally, the dataset involves a limited number of participants, which may impact the generalization capability of the system when deployed at scale.

The current implementation relies on lightweight geometric feature extraction and does not incorporate deep learning-based anti-spoofing models, which may provide improved robustness against advanced attacks such as high-quality deepfakes. Furthermore, the system assumes stable network conditions for OTP verification and does not evaluate performance under network delays or failures.

Overall, the results demonstrate the feasibility of the proposed hybrid authentication framework, but further validation is required under large-scale and real-world deployment conditions.

B. Future Enhancements

Future work will focus on large-scale evaluation, improved robustness against advanced spoofing attacks, and deployment in real-world environments.

9. REPRODUCIBILITY AND EXPERIMENTAL SETUP

For the reproduction of the proposed FaceOTP system to be ensured, all the experiment parameters, tools, and environments are well identified. The experiment was conducted on the Windows 11 Operating System with the Intel Core i5 processor that has 8 GB of RAM.

Implemented in Python-3.x, it used OpenCV for capturing video, MediaPipe Face Mesh for face landmark detection, PyOTP for generating time-based OTPs, and FastAPI for backend services. It uses SQL Alchemy for handling database-related activities.

All parameters of algorithms, like thresholds of EAR, time for liveness challenges, validity interval of OTP, and tolerance thresholds for face matching, were set to be unchanged throughout the experiments. Random seeds were set to be unchanged wherever appropriate.

These details allow future researchers to replicate the experiments and validate the reported performance of the proposed authentication system.

10. CONCLUSION

This paper presented FaceOTP, a hybrid authentication framework that integrates facial recognition, behavioral liveness detection, and TOTP verification. The proposed system was evaluated under controlled experimental conditions using multiple users and various spoofing scenarios, including photo and video-based attacks.

The results demonstrate that the system achieves high authentication accuracy, effective liveness detection, and reliable OTP verification, with an average authentication time of approximately 4.5 seconds. These findings confirm that combining biometric verification with time-based authentication enhances resistance to spoofing and replay attacks.

However, the current evaluation is limited to a small-scale controlled environment. Future work will focus on large-scale evaluation, robustness against advanced deepfake attacks, and

deployment in real-world environments with diverse user populations.

Overall, the proposed FaceOTP system provides a practical and efficient authentication solution for controlled environments. Further validation on larger datasets and real-world scenarios is required before deployment in critical applications.

REFERENCES

[1] Kumaran, U., Senthil, B., Sundaram, V., Chauhan, A.S., Meena, V.P., Azar, A.T., Kamal, N.A., Njima, C.B. (2025). Fortifying cybersecurity with iris generated one-time passwords. In 2025 International Conference on Control, Automation and Diagnosis (ICCAD), Barcelona, Spain, pp. 1-6. https://doi.org/10.1109/ICCAD64771.2025.11099319

[2] Jubur, M., Price, C.R., Shirvanian, M., Saxena, N., Jarecki, S., Krawczyk, H. (2025). Building and testing a hidden-password online password manager. IEEE Transactions on Information Forensics and Security, 20: 7454-7468. https://doi.org/10.1109/TIFS.2025.3583459

[3] Sameer, M., Sudharsan, K., Begum, A.B. (2024). Unforgettable password generation using LoRA fine-tuned large language model. In 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, pp. 1-5. https://doi.org/10.1109/ADICS58448.2024.10533548

[4] Zhou, M., Wang, Q., Li, Q., Zhou, W., Yang, J., Shen, C. (2024). Securing face liveness detection on mobile devices using unforgeable lip motion patterns. IEEE Transactions on Mobile Computing, 23(10): 9772-9788. https://doi.org/10.1109/TMC.2024.3367781

[5] Li, P., Pan, L., Chen, F., Hoang, T., Wang, R. (2023). TOTPAuth: A time-based one time password authentication proof-of-concept against metaverse user identity theft. In 2023 IEEE International Conference on Metaverse Computing, Networking and Applications (MetaCom), Kyoto, Japan, pp. 662-665. https://doi.org/10.1109/MetaCom57706.2023.00117

[6] Zhang, J., Wang, Y. (2023). A survey of behavioral biometric authentication on smartphones. In Proceedings of the 2023 4th International Conference on Machine Learning and Computer Application, Hangzhou, China, pp. 722-729. https://doi.org/10.1145/3650215.3650342

[7] Fang, H., Liu, A., Wan, J., Escalera, S., Zhao, C., Zhang, X., Li, S.T., Lei, Z. (2024). Surveillance face anti-spoofing. IEEE Transactions on Information Forensics and Security, 19: 1535-1546. https://doi.org/10.1109/TIFS.2023.3337970

[8] Yin, Y., Jang-Jaccard, J., Baghaei, N. (2022). PassImg: A secure password generation and management scheme without storing. In 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Hangzhou, China, pp. 341-346. https://doi.org/10.1109/CSCWD54268.2022.9776045

[9] Kagnici, O., Bisgin, H., Uludag, S. (2025). Evaluating the efficacy of GPT-based password generation on real world data. In 2025 1st International Conference on Secure IoT, Assured and Trusted Computing (SATC), Dayton, OH, USA, pp. 1-5. https://doi.org/10.1109/SATC65530.2025.11137320

[10] Xing, H., Tan, S.Y., Qamar, F., Jiao, Y. (2025). Face anti-spoofing based on deep learning: A comprehensive survey. Applied Sciences, 15(12): 6891. https://doi.org/10.3390/app15126891

[11] Vijay, M., Jayalakshmi, M., Shanmugaraj, G. (2024). Bio-inspired metaheuristic feature fusionmethod for multi-biometric identification. In 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1-5. https://doi.org/10.1109/ICRITO61523.2024.10522216

[12] Blanton, M., Murphy, D. (2024,). Privacy preserving biometric authentication for fingerprints and beyond. In Proceedings of the Fourteenth ACM Conference on Data and Application Security and Privacy, pp. 367-378. https://doi.org/10.1145/3626232.3653269

[13] Kose, N., Dugelay, J.L. (2013). On the vulnerability of face recognition systems to spoofing mask attacks. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, pp. 2357-2361. https://doi.org/10.1109/ICASSP.2013.6638076

[14] Jain, A.K., Ross, A., Prabhakar, S. (2004). An introduction to biometric recognition. IEEE Transactions on Circuits and Systems for Video Technology, 14(1): 4-20. https://doi.org/10.1109/TCSVT.2003.818349

[15] Erdogmus, N., Marcel, S. (2014). Spoofing face recognition with 3D masks. IEEE Transactions on Information Forensics and Security, 9(7): 1084-1097. https://doi.org/10.1109/TIFS.2014.2322255

[16] ISO/IEC 30107-1:2023. (2023). Information Technology — Biometric Presentation Attack Detection. https://www.iso.org/standard/83828.html.

[17] Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S. (2009). Handbook of Fingerprint Recognition. London: Springer London. https://doi.org/10.1007/978-1-84882-254-2

[18] Juels, A., Wattenberg, M. (1999). A fuzzy commitment scheme. In CCS '99: Proceedings of the 6th ACM conference on Computer and communications security, Kent Ridge Digital Labs, Singapore, pp. 28-36. https://doi.org/10.1145/319709.319714

[19] Ratha, N., Connell, J., Bolle, R.M., Chikkerur, S. (2006). Cancelable biometrics: A case study in fingerprints. In 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, pp. 370-373. https://doi.org/10.1109/ICPR.2006.353

[20] M'Raihi, D., Machani, S., Pei, M., Rydell, J. (2011). RFC 6238. TOTP: Time-based one-time password algorithm. IETF. https://doi.org/10.17487/RFC6238