PeerJ Computer Science

A keyless multimodal-based user authentication scheme using generative adversarial networks

Mayada Tarek[1,2], Eslam Hamouda[1,2], Amjad Alsirhani[2,3], Abdullah Alomari[4] and Ayman Mohamed Mostafa[5,6]

[1] Computer Science Department, Faculty of Computers & Information, Mansoura University, Mansoura, Egypt
[2] Computer Science Department, College of Computers and Information Sciences, Jouf University, Sakaka, Jouf, Saudi Arabia
[3] Faculty of Computer Science, Dalhousie University, Halifax, Canada
[4] Department of Computer Science, Al-Baha University, Al-Baha, Saudi Arabia
[5] Information Systems Department, Faculty of Computers & Informatics, Zagazig University, Zagazig, Egypt
[6] Information Systems Department, College of Computers and Information Sciences, Jouf University, Sakaka, Jouf, Saudi Arabia

ABSTRACT

Biometrics are increasingly used for access control, fraud detection, and authentication systems. Nevertheless, attackers can deceive such systems using forged biometrics. This research proposes a novel method that makes biometric security systems more resilient to such attacks. The proposed method transforms the user's biometric data into an irreversible code to protect the original data. This code combines data from multiple biometric modalities, making fabricating a false biometric harder. Additionally, the proposed method does not depend on any secret keys, which helps avoid cases of stolen tokens. The proposed method utilizes the generative adversarial network (GAN) to generate synthetic biometric templates from multiple modalities, which is considered a transformation function for biometric data. Three fusion levels are presented; features from multiple biometric modalities are extracted first in each fusion level. Subsequently, the features train a generative adversarial network to produce synthesized biometric templates. These synthesized templates serve as secure substitutes for the original biometrics during authentication, preventing direct exposure of raw biometric data. We evaluated our methods on the CASIA-V3-Internal and MMU1 iris datasets and the AT&T (ORL) and FERET face datasets. The results showed that our proposed methods can achieve higher accuracy, usability, and improved security compared to a single biometric modality. The proposed feature-level, GAN-based, and decision-level fusion schemes achieved 2.03%, 0.82%, and 0.0297% error rates, respectively, for CASIA and ORL datasets and 1.53%, 0.80%, and 0.0313% error rates, respectively, for MMU1 and FERET datasets. Moreover, we have demonstrated that our method resists pre-image and correlation attacks.

Subjects Artificial Intelligence, Data Mining and Machine Learning, Security and Privacy, Neural Networks

Keywords Multimodal cancelable biometrics, Template protection, Generative adversarial network, Biometric fusion

Submitted 17 March 2025
Accepted 14 October 2025
Published 18 November 2025

Corresponding author
Mayada Tarek,
mayada_tarek@mans.edu.eg

Academic editor
Giovanni Angiulli

Additional Information and Declarations can be found on page 25

DOI **10.7717/peerj-cs.3360**

How to cite this article Tarek M, Hamouda E, Alsirhani A, Alomari A, Mostafa AM. 2025. A keyless multimodal-based user authentication scheme using generative adversarial networks. **PeerJ Comput. Sci. 11:e3360 DOI 10.7717/peerj-cs.3360**

INTRODUCTION

Multimodal biometrics is a type of biometric authentication that uses multiple biometric traits to identify an individual by combining different types of biometrics (*Jain & Ross, 2004*). This article presents a new method of multimodal biometrics for secure authentication.

Multimodal biometrics has many merits over unimodal biometrics (*Adusumalli & Bhuvaneswari, 2018*). Firstly, it can increase accuracy by leveraging the advantage of several biometric modalities. For instance, face recognition is usually more accurate at a distance, whereas fingerprint recognition is usually more accurate up close (*Adusumalli & Bhuvaneswari, 2018*). Secondly, combining biometrics can diminish spoofing and fraud by securing the system and preventing it from being fooled. For instance, an attacker must create a fake fingerprint and voice to spoof a multimodal biometric system based on finger and voice recognition (*Aditya & Kaur, 2021*). Moreover, multimodal biometrics increase usability as the authentication process becomes more user-friendly. For example, users can identify themselves using their faces and fingerprints, which can be more convenient than remembering PINs or passwords (*Sheena & Mathew, 2014*). There are numerous challenges associated with multimodal biometrics (*Patil, 2012*). Users' privacy should be a key consideration in the biometric template design. Templates require removing all the personal information that could be used to identify the users. Moreover, the biometric traits used in the system should be compatible with each other. For example, these two traits must be linked if that system uses face and iris recognition. The fusion of multimodal biometric data may be challenging because the data quality may differ; data may be of various types. There are three primary levels of fusion in multimodal biometrics (*Gupta, 2015*), utilizing the features that have been obtained from various biometric modalities (*Govindarajan, 2004*), including scores that have been assigned to each biometric modality (*Feifei & Gonging, 2011*), or accumulating the decisions that have been made by multiple biometric systems (*Prabhakar & Jain, 2002*).

Cancellable biometrics (CB) is a biometric that can be transformed into a non-reversible form, which ensures its resistance to attack (*Patel, Ratha & Chellappa, 2015*). This is obtained by distorting the inherent biometric data in a way that keeps the information required for authentication. However, reconstructing the original data is impossible due to such distortion (*El-Hameed et al., 2021*). Cancelable biometrics offers many advantages over traditional biometrics. It can enhance security through the increase in the complexity of hacking a user's biometric data. Moreover, it can improve flexibility by permitting users to revoke their biometric records without involving a change in their appearance. Besides, it can improve usability by simplifying the authentication process for users on multiple devices (*Manisha & Kumar, 2020*). Several methods have been developed to generate cancelable biometric templates, including random projection, cancelable filters, biohashing, permutation, bio-convolving, and Bloom filters (*Choudhury et al., 2018*). Random projection techniques map features into a random subspace while preserving distances, with variants such as the Johnson–Lindenstrauss lemma improving system

performance (*Soliman, Amin & Abd El-Samie, 2018*). Cancelable filters apply random convolution kernels to produce secure templates (*Savvides, Kumar & Khosla, 2004*), while biohashing extends random projection by incorporating user-specific random keys (*Teoh, Ngo & Goh, 2004*). Permutation methods randomize biometric features using auxiliary data, and bio-convolving generates sequences of cancelable templates from sequential data (*Maiorana et al., 2010*). Bloom filter approaches, meanwhile, provide efficient biometric querying through adaptive filtering (*Rathgeb et al., 2014*). Recently, one-factor cancelable authentication schemes based on Generative Adversarial Networks (GANs) have been proposed, in which cancelable keys are derived from permuted biometric traits. Although these schemes demonstrate improved recognition performance, their security remains a critical concern and warrants further enhancement (*Tarek, Hamouda & El-Metwally, 2021*).

Integrating multimodal and cancelable biometrics is particularly promising for many application areas, such as access control, border security, and law enforcement. Moreover, it offers many advantages (*Paul & Gavrilova, 2012*), including:

- **Improved accuracy**: Using multiple biometric traits can help to reduce the chances of a false match or rejection.
- **Enhanced security**: Cancellable biometrics complicates the theft or presentation forgery of the biometric template.
- **Increased usability**: In terms of human-computer interactions, multimodal biometric systems are more user-friendly than ones that use a single modality.
- **Improved robustness**: Systems employing multiple biometric modes are designed to be more robust to changes in lighting and pose and other environmental parameters.

GANs are deep learning algorithms that can produce realistic synthetic data (*Saxena & Cao, 2020*). The application of GANs can produce biometric templates that can be used for authentication without compromising the user's privacy. GANs are an emerging technique for creating a more reliable, keyless, secure, and user-friendly biometric access system (*Tarek, Hamouda & El-Metwally, 2021*; *Alqahtani, Kavakli-Thorne & Kumar, 2019*). GANs consist of a pair of generative and discriminative networks trained to compete against each other. One network generates a synthetic biometric template, while the other distinguishes a genuine and synthetic template from the given biometrics. As they train, the generator gets better at generating templates that are indistinguishable from real ones. GANs are effective at generating synthetic biometric templates that are both realistic and secure (*Saxena & Cao, 2020*).

This article suggests a novel approach for cancelable multimodal biometrics that uses GANs to protect multimodal biometrics (left and right iris and face) at different fusion levels. The proposed approach first extracts feature from multiple biometric modalities. These features then train a GAN to generate synthetic biometric templates. The generated

templates are non-reversible and can be used for authentication without compromising the user's privacy. The proposed approach presents three fusion levels:

- **Feature-level fusion**: The features from multiple biometric modalities are combined to generate the training input features for the GAN models.
- **GAN-based fusion**: Each biometric modality features trains its own GAN model. The generated synthetic biometric templates from each GAN model are then combined.
- **Decision-level fusion**: Each biometric modality features train its GAN model to generate synthetic biometric templates. A voting process is then used to decide the final result.

The suggested approach uses GANs, which generate synthetic biometric templates, making it more difficult for hackers to spoof the system. The non-reversibility of the generated templates protects the user's privacy. Employing a system based on various biometric modalities can improve the system's accuracy, and the proposed approach is relatively simple to implement. This work relies on the difficulty of an attacker simultaneously impersonating multiple biometric traits for a genuine user (*Merkle, Kevenaar & Korte, 2012*). The GAN models work as a keyless salting scheme for the input biometric modality (*Tarek, Hamouda & El-Metwally, 2021*). Accordingly, the performance is increased using multiple biometric modalities instead of single or unimodal. In addition, multimodal provides a cost-saving way to enhance performance as it does not require any data sensor extraction, feature modules, or matching units (*Ross & Jain, 2004*).

Despite their success in feature generation and data augmentation, GANs exhibit several vulnerabilities when applied to authentication systems. A major limitation is mode collapse, wherein the generator fails to capture the full variability of the underlying biometric distribution and instead produces highly similar outputs. This lack of diversity reduces the representativeness of synthetic samples, thereby compromising the robustness of the authentication process and increasing susceptibility to spoofing attacks (*Saxena & Cao, 2020*). The proposed framework addresses these limitations through a multi-level fusion strategy that integrates feature-level, GAN-based, and decision-level components. This design alleviates the impact of mode collapse by diversifying the representation space and ensuring that authentication does not rely solely on the variability captured by the generator. In addition, the adoption of cancelable biometric transformations enhances privacy protection by concealing original templates, thereby mitigating risks associated with overfitting and potential template leakage.

The rest of the article is organized as follows: related works about multimodal cancelable biometrics are summarized in "Background and Related Works". The "Problem and Mathematical Formulation" introduces the problem and mathematical formulation for the presented multibiometric models. "Materials and Methods" presents the proposed cancellable multimodal security schemes. "Results" presents the experimental results for the proposed schemes. "Discussion" discuss the security perspectives and explores the

operational considerations and real-world application of the proposed method. Finally, "Conclusion" concludes the main article's work.

BACKGROUND AND RELATED WORKS

El Rahman & Alluhaidan (2024) proposed two convolutional neural networks (CNN)-based multimodal biometric systems that combine fingerprint and an electrocardiogram (ECG) data. The authors explored different feature fusion levels and employed various feature extraction and classification methods. They claimed that a sequential multimodal CNN system outperforms a parallel one. In the same year, *Salturk & Kahraman (2024)* published a method that combined signature dynamics and static and face-based data. They implement CNNs, long short-term memory (LSTMs), and gated recurrent units (GRUs), as well as the temporal convolutional networks (TCNs) that are based on the foundation of deep learning algorithms. Their research proved that the system could achieve outstanding performance by merging kinetic and static biometric features. The multimodal biometric system proposed by *Haider et al. (2023)* incorporates finger vein patterns and fingerprint recognition. A fuzzy system fuses biometric information of multi-traits. Initially, the data on finger texture is classified using a Support Vector Machine (SVM). Subsequently, transfer learning is applied by plugging in the pre-trained CNNs for finger vein recognition. With the help of a fuzzy rules-based inference system, the Multimodal system performs the task more accurately than just the unimodal system. *Balraj & Abirami (2022)* combined facial recognition and iris scanning with score-level fusion and an extended ant colony optimization (ACO) approach. *El-Rahiem et al. (2021)* presented a multibiometric CB system using CNN features, feature map fusion, and the DeepDream algorithm. *Abdellatef et al. (2020)* proposed a CNN-based scheme for face recognition that achieves cancelability through bio-convolving encryption. *Sudhakar & Gavrilova (2020)* utilized CNN features and random projection for finger vein and iris recognition. Table 1 summarizes the recent techniques that have been noticed for their significant contributions to multibiometric authentication systems. However, the recognition performance for existing multimodal cancelable biometrics systems is promising, and factors such as security, privacy, and usability should be considered.

PROBLEM AND MATHEMATICAL FORMULATION

The main challenge for this work is how to generate a cancelable biometric template from multibiometric traits without any external key by using GAN model. A GAN model designed with two competed networks was considered here as a transformed function where the inputs are multibiometric features and a driven key from this feature and the output is the transformed template. The foundation of this proposed work draws inspiration from the established GAN model, which serves as a keyless biometric salting scheme (*Tarek, Hamouda & El-Metwally, 2021*). The notations used by the proposed multimodal cancelable biometric scheme are shown in Table 2.

Table 1 Recent multibiometric authentication systems.

Authors	Year	Biometric trait	Algorithm	Fusion level
Abdellatef et al. (2020)	2020	• Different regions of a face	CNN & bio-convolving encryption	Feature fusion
Sudhakar & Gavrilova (2020)	2020	• Finger vein • Iris	CNN & SVM &random projection	Feature fusion
El-Rahiem et al. (2021)	2021	• Fingerprint • Finger vein • Iris	CNN & DeepDream algorithm	Feature fusion
Balraj & Abirami (2022)	2022	• Face • Iris	Extended ACO	Score fusion
Haider et al. (2023)	2023	• Finger texture • Finger Vein	SVM & CNN	Score fusion
El Rahman & Alluhaidan (2024)	2024	• Fingerprint • ECG Signal	CNN	Feature fusion
Salturk & Kahraman (2024)	2024	• Signature • Face	CNN & LSTM & GRU & TCN	Feature fusion

Table 2 Notations used by the proposed multimodal cancelable biometric scheme.

Notation	The meaning of the notation
$\boldsymbol{x}$	The input biometric data
n	The length of the iris binary template
f	The length of the face binary template
$\boldsymbol{k}$	The salting key
$\boldsymbol{S}$	The generator output data
$\boldsymbol{C}$	The cancelable template
$\boldsymbol{G}$	The generator of the GAN model
$\boldsymbol{D}$	The discriminator of the GAN model
θ	The decision threshold
$\boldsymbol{Sim}\ (\boldsymbol{C})$	The similarity score of the cancelable template C

Let the biometric system contain p biometric samples $[x_1, x_2, \ldots, x_P]$ for each enrolled user. The GAN model is used to create a biometric salting data S, which is then combined with the mean of the input biometric data to form the cancelable template C. The decision of whether the user is genuine or impostor is made based on the cancelable template C. The GAN model contains two compotator networks: a forger (the generator, G) and a detective (the discriminator, D). The generator G tries to generate biometric salting data indistinguishable from actual data. In contrast, the discriminator D tries to distinguish between real and generated data. This competition is formulated mathematically in the equations (1:6) (*Saxena & Cao, 2020*), Eq. (1) presents GAN loss function which is obtained from the binary cross-entropy formula

$$L\ (x, k) = [k\ .\log(x) + \ (1 - k).\log(1 - x)] \tag{1}$$

where the discriminator loss function LD remaining the same with a variation of Generator 's cost function LG. While training discriminator, the output will be 1 for real data and 0 for fake data. Then, by substituting this into Eq. (1), we have:

$$L\ (D(k), 1) = \log(D\ (k)) \tag{2}$$
$$L\ (D(G(x), 0) = \log(1 -\ D\ (G\ (x))). \tag{3}$$

The final discriminator loss function is denoted in Eq. (4) as the discriminator's goal is to classify accurately its input as fake or real. Therefore, the given G & D loss functions have to be maximized.

$$L^D\ = \max\ [\log(\ D\ (k))\ + \log(1 -\ D\ (G\ (x))). \tag{4}$$

The final generator loss function is denoted in Eq. (5) as the generator is competing against discriminator. So, the generator aims to minimize the optimization problem given in Eq. (4).

$$L^G\ = \min\ [\log(\ D\ (k))\ + \log(1 -\ D\ (G\ (x)))]. \tag{5}$$

Therefore, Eq. (6) denoted the combination of the generator and discriminator loss functions as combining between Eqs. (4), (5).

$$L\ = \min_G \max_D [\log(\ D\ (k))\ + \log(1 -\ D\ (G\ (x)))]. \tag{6}$$

Loss function in Eq. (6) is valid only for a single data point. Therefore, to consider this equation to entire dataset, Eq. (7) considered the expectation of combined loss function as:

$$\min_G \max_D V(G, D) = E_k[\text{Log } D(k)] + E_k[\log(1 - D(G(x))]. \tag{7}$$

Here, V (G, D) represents the overall performance of both G and D. The lower this value is for G, the better it is at creating realistic fakes. Conversely, the higher it is for D, the better it is at spotting fakes. The first part of the equation calculates the average confidence D has when evaluating actual data. In contrast, the second part of the equation calculates D's average confidence when evaluating the generator's creations. K represents the salting key, a permutated version for input biometric sample x_i; it is generated by rearranging the bits in a randomly selected template x_i for the enrolled user.

The cancelable template for each enrolled user, C, is calculated using Eq. (8):

$$C = Y \oplus\ S \tag{8}$$

where $\oplus$ is the XOR operation applied to the generator output, S, and the binarized mean of all input samples for the enrolled user, Y, it can be calculated using Eqs. (9) and (10) as follows:

$$T = \text{mean}(x_1,\ x_2,\ \ldots, x_p) \tag{9}$$
$$Y\_i = \begin{cases} 1, & \textit{if } (T_i \geq 0.5) \\ 0, & \textit{otherwise} \end{cases} \tag{10}$$

The decision of whether the user is genuine or impostor is made based on the cancelable template C using a decision threshold θ. If the similarity score of cancelable template C is

greater than θ, then the user is classified as genuine. Otherwise, the user is classified as an impostor. It can be calculated using Eq. (11):

$$\text{Decision} = \begin{cases} 1, & if\ (Sim(C) \geq \theta) \\ 0, & otherwise \end{cases}. \tag{11}$$

Sim(C) is the similarity score of the cancelable template C, which is calculated based on the hamming distance measure. This metric is applied to check the variation between a test biometric template and the reference template; it refers to the number of positions where the two biometric templates differ. This represents the general mathematical formation of the proposed methodology.

MATERIALS AND METHODS

The primary contribution of this research lies in enhancing the security and performance of the biometric system. This improvement is achieved by utilizing a multimodal biometric modality instead of relying solely on a single instance from one biometric modality (*Adusumalli & Bhuvaneswari, 2018*). The actual implementation of the methodology varies depending on the applied fusion scheme, as discussed in the later subsections. Unlike methods that handle transformation keys, the suggested approach only takes biometric data and avoids the risk of key storage vulnerabilities. Security enhancement stems from the increased complexity impostors face when attempting to breach the system using multiple biometric traits. Additionally, performance enhancement is anticipated by applying fusion principles within the context of multimodal biometric traits. In light of these considerations, this work presents three strategies for combining iris and face data using a multimodal fusion approach based on the standard GAN model: feature-level fusion, GAN-based fusion, and decision-level fusion. By feature-level fusion, the unique feature of shared feature values among the biometric modalities is of utmost help in unifying the related feature values. It prioritizes a concise selection of certain key features that can enhance recognition accuracy (*Govindarajan, 2004*). In GAN-based fusion, the generated templates from various GAN models are integrated to produce a new template. This resultant template can then be utilized by verification or identification modules to make informed decisions about an individual's identity (*Feifei & Gonging, 2011*). Decision-level fusion involves consolidating multiple hypotheses into a single decision. This technique is frequently employed to enhance decision precision (*Prabhakar & Jain, 2002*).

Feature-level fusion

Given that humans possess two iris instances (Left and right) and one face, two binary iris templates (left and right, each of length $[1 \times n]$) and one binary face template of length $[1 \times f]$ are concatenated into a single binary feature vector $[1 \times (2n + f)]$. This concatenated vector is then input into a GAN model, where the generator produces a transformed template while the discriminator is trained using both generated and permuted samples. To enhance security and non-invertibility, the binarized output of the generator is XORed

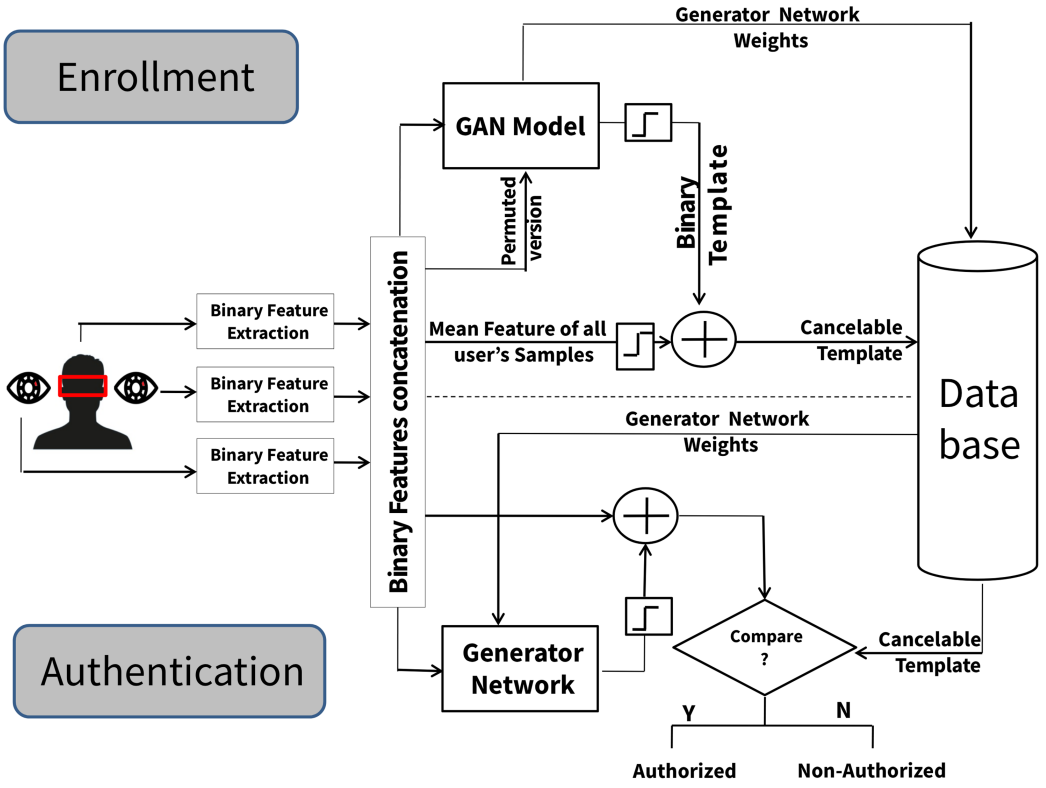

Figure 1 **The multimodal cancelable feature level fusing.** Full-size DOI: 10.7717/peerj-cs.3360/fig-1

with the binarized mean of the concatenated samples, producing the final cancelable template stored in the system database, as illustrated in Fig. 1.

First, the GAN model generates a transformed template from this concatenated sample. Second, this transformed template is subjected to XOR (exclusive OR) operation with the mean of the concatenated samples to enhance system performance and security concerns. As displayed in Fig. 1, the enrollment phase for each individual comprises three core stages. The first stage aims to execute feature-level fusion between the features of the two binary iris instances and one binary face instance. These results in a definitive binary feature achieved through a straightforward concatenation procedure. A total of p binary concatenated samples are generated for each person. Each sample has a length of ([1 × 2n] + [1 × f]) binary template; each left and right iris instance has a template of length 1 × n binary features, and the face has a template of length 1 × f binary features. In the second stage, the produced p binary concatenated samples serve as training inputs for the generator networks within the GAN model. In contrast, a single selected binary concatenated sample is randomly permuted and introduced as an additional input to the discriminator network in the same GAN model. This input to the discriminator serves as a salting key for the GAN transformation model.

During GAN model training, the generator's output, a template of size 1 × (2n + f), is also input for the discriminator network. Across each training epoch, a conventional

Scheme 1 **The proposed feature-level fusion scheme.**

Input:
- Left iris samples $\{L_1, \ldots, L_p\}$, length n
- Right iris samples $\{R_1, \ldots, R_p\}$, length n
- Face samples $\{F_1, \ldots, F_p\}$, length f
- N (training epochs), α (learning rate)
- h_1 (generator hidden size), h_2 (discriminator hidden size)

Procedure:
1. For each sample $i \in \{1, \ldots, p\}$:
   Create template x_i = concatenate (L_i, R_i, F_i) of size $(2n + f)$
2. Randomly select template x_i
3. Generate salting key k = permute(x_i)
4. Compute reference template T using Eq. (9)
5. Compute binarized template Y using Eq. (10)
6. Build generator network G: $[(2n + f) \rightarrow h1 \rightarrow (2n + f)]$
7. Build discriminator network D: $[(2n + f) \rightarrow h2 \rightarrow 1]$
8. Initialize weights W_g(G), W_d (D)
9. For epoch = 1 to N:
   Train G and D alternately with backpropagation algorithm: Input: $\{x_1, \ldots, x_p\}$
10. Retrieve generator output S
11. Compute cancelable reference template C_{ref} using Eq. (8)
12. Store $\{W_g, C_{ref}\}$ in system database

Output:
During authentication, compute decision for test template C_{test} using Eq. (11).

back-propagation learning algorithm is applied to update the weights of the discriminator network (*Goodfellow et al., 2014*). The weights of the generator network are updated using the same learning algorithm by the output of the discriminator network. After a specific number of training epochs, the final output of the generator network is employed to construct a transformed template, and the final generator network weights are stored in the system database. Because of the reversible nature of the GAN model (*Creswell & Bharath, 2018*), the original concatenated iris and face features can be recovered if the transformed template is publicly stored in the system database, along with the weight values of the generator network. A third stage is introduced to generate a non-invertible cancelable template to circumvent this security concern while enhancing system performance. This is achieved through an additional transformation layer involving XOR between a binarized version of the transformed template (generator's output) and another binarized version derived from the mean computation of all concatenated iris and face samples (generator's inputs). The resultant binary output from the XOR operation constitutes a template of the size $1 \times (2n + f)$ and is proposed as the final stored cancelable template within the system database. The steps for the feature-level fusion scheme are illustrated in Scheme 1. During the authentication phase, the proposed system solely necessitates tested samples from the left and right iris instances along with the face to authorize the identity of the tested individual. As depicted in Fig. 1, a tested sample is generated by concatenating the features extracted from both irises and face. This tested template is input into the generator networks, employing their stored weight values. A binarized variant of the generator networks' output is XORed with the tested sample, culminating in creating the final tested

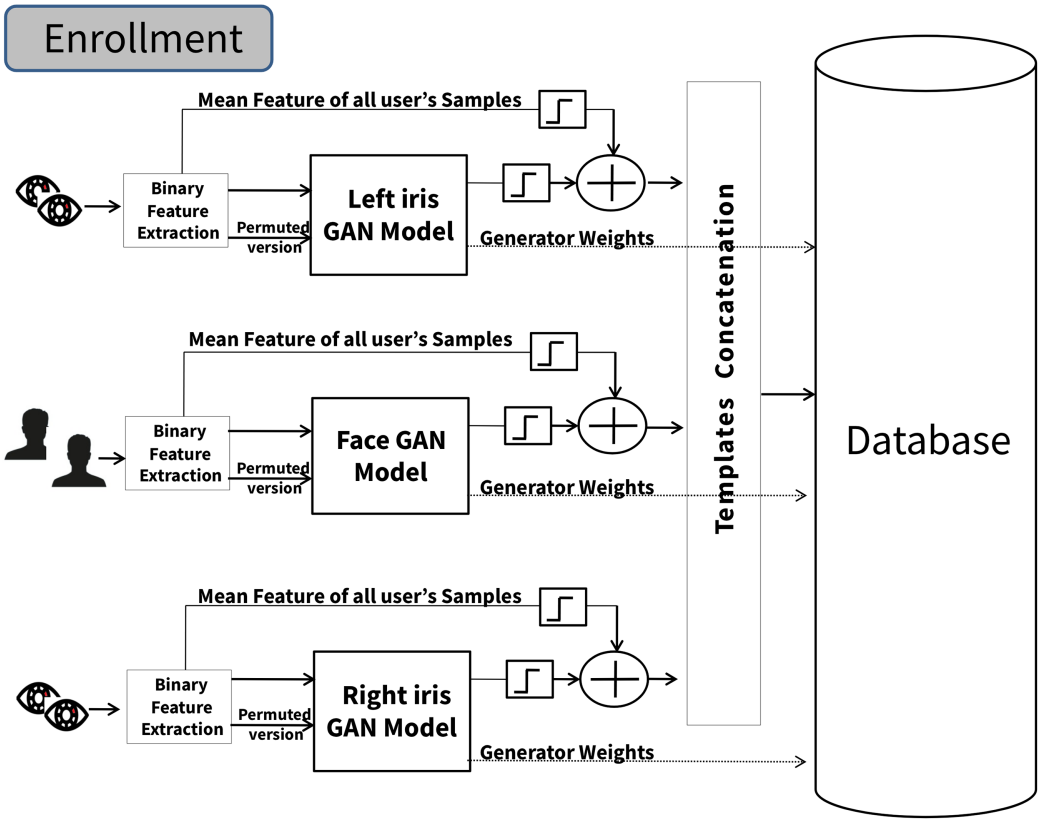

Figure 2 **The multimodal cancelable GAN-based fusion scheme (enrollment phase).**
Full-size DOI: 10.7717/peerj-cs.3360/fig-2

cancelable template. Ultimately, comparing these tested and stored cancelable templates for the claimed individual facilitates the ultimate authentication decision.

GAN-based fusion

The second proposed scheme is centered around merging the outcomes of each cancelable instance model. The three cancelable templates generated from each cancelable iris and face instance model are fused through a concatenation process, as depicted in Fig. 2. In this approach, each cancelable iris and face model operates independently. As depicted in Fig. 2, enrollment phase functions independently for each iris and face instance, with the outputs of each iris and face instance model concatenated to yield the fusion system's cancelable template. The process begins by treating each iris and face instance model distinctly. For every instance model, a set of p binary feature samples from each iris and face instance model are utilized as training inputs for each generator network within the GAN model. The output of the generator networks—a template of size $1 \times n$ for the iris models and $1 \times f$ for the face model, is used as an input for their discriminator network and, in combination with a randomly permuted version of a chosen binary input sample from their generator, as seen in Fig. 2. A back-propagation learning algorithm (*Goodfellow et al., 2014*) is employed during the GAN training process for each instance

Scheme 2 **The proposed GAN-based/decision-level fusion schemes.**

Input:
- *Left iris samples $\{L_1, \ldots, L_p\}$, length n*
- *Right iris samples $\{R_1, \ldots, R_p\}$, length n*
- *Face samples $\{F_1, \ldots, F_p\}$, length f*
- *N (training epochs), α (learning rate)*
- *h_1 (generator hidden size), h_2 (discriminator hidden size)*

Procedure:
1. *Randomly select $l_i \in \{L_1, \ldots, L_p\}$*
2. *Randomly select $r_i \in \{R_1, \ldots, R_p\}$*
3. *Randomly select $f_i \in \{F_1, \ldots, F_p\}$*
4. *Generate salting keys:*
   k_l = permute(l_i), kr = permute(r_i), k_f = permute(f_i)
5. *Compute reference templates T_l, T_r, T_f using Eq. (9)*
6. *Compute binarized templates Y_l, Y_r, Y_f using Eq. (10)*
7. *Build generator networks:*
   G_l: [n → h1 → n], G_r: [n → h1 → n], G_f: [f → h1 → f]
8. *Build discriminator networks:*
   D_l: [n → h2 → 1], D_r: [n → h2 → 1], D_f: [f → h2 → 1]
9. *Initialize weights for all generator networks*
10. *Initialize weights for all discriminator networks*
11. *For epoch = 1 to N:*
    Train $\{G_l, G_r, G_f\}$ and $\{D_l, D_r, D_f\}$ alternately
    Inputs: $\{L_1, \ldots, L_p\}$, $\{R_1, \ldots, R_p\}$, $\{F_1, \ldots, F_p\}$
    Keys: k_l, k_r, k_f
12. *Retrieve outputs S_l, S_r, S_f from generators*
13. *Compute cancelable templates C_{l_ref}, C_{r_ref}, C_{f_ref} using Eq. (8)*
14. *For GAN-based fusion:*
    Concatenate C_{l_ref}, C_{r_ref}, C_{f_ref} → C_{ref} (size: 2n + f)
15. *Store generator weights and reference template(s) in database*

Output:
- *GAN-based: authenticate test template C_{test} using Eq. (11)*
- *Decision-level: authenticate using majority voting on $\{C_{l\_test}, C_{r\_test}, C_{f\_test}\}$ with Eq. (5)*

model, updating the discriminator and generator network weights based on the output of the discriminator network for each GAN network. Following a specific number of epochs, the final generator network weights are stored in the system database. Each generator network output is binarized and subsequently X-ORed with a binarized version derived from the mean computation of their iris/face instance samples (generator's inputs). The resultant binary output from the XOR operation yielding a template of size 1 × n for each iris model and 1 × f for the face model is proposed as the cancelable template. The last step involves concatenating the three cancelable templates generated from each iris and face instance model to yield the stored fused cancelable template (a template of the size 1 × (2n + f)) within the system database and the generator networks' weights for each iris and face instance model. The steps for a GAN-based fusion scheme are illustrated in Scheme 2.

Subsequently, in the authentication phase, the proposed system requires tested samples from both the left and right iris and face instances to validate the individual's identity. As illustrated in Fig. 3, each tested template is forwarded to its respective generator networks as input, leveraging the stored weight values for each iris/face instance generator network. After being binarized, each generator network's output is X-ORed with the corresponding trait instance's tested sample to generate an instance-level cancelable template. Ultimately,

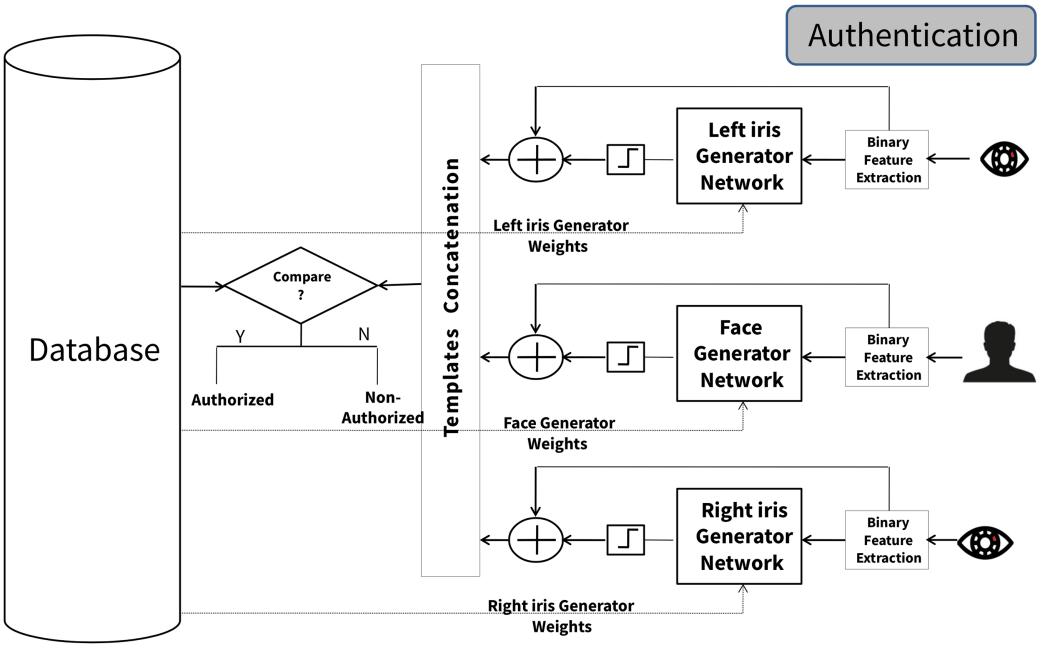

Figure 3 **The multimodal cancelable GAN-based fusion scheme (authentication phase).**
Full-size DOI: 10.7717/peerj-cs.3360/fig-3

the fused tested cancelable template results from concatenating the three instance-level cancelable templates. To finalize the authentication process, these fused tested cancelable templates are matched against the stored fused cancelable template for the purported individual, culminating in the authentication decision.

Decision-level fusion

The third proposed scheme's main idea depends on applying the majority voting mechanism to the final decisions of each cancelable GAN model for each instance (two iris and face). This approach aims to improve overall decision accuracy and reliability by leveraging the diversity of information provided by different sources. Each instance evaluates its input data using this approach and makes decisions. These individual decisions are then aggregated or "voted" upon to determine the outcome. Each instance's decision is considered as a "vote". The decision that receives the most votes is selected as the final decision. As shown in Fig. 4, the method works by first training three independent GAN models. The models are trained using the same algorithm described earlier. During the authentication, the proposed system operates with the requirement of verified samples originating from both the left and right iris and face instances.

As depicted in Fig. 5, each tested template is directed to its corresponding generator networks as input. This process involves utilizing the weight values stored for each generator network associated with iris instances. After being transformed into a binary format, each generator network's outcome is subjected to an X-OR operation with the respective tested sample of the trait instance. Each tested cancelable template is matched against the stored cancelable template for its instance model, leading to an authentication

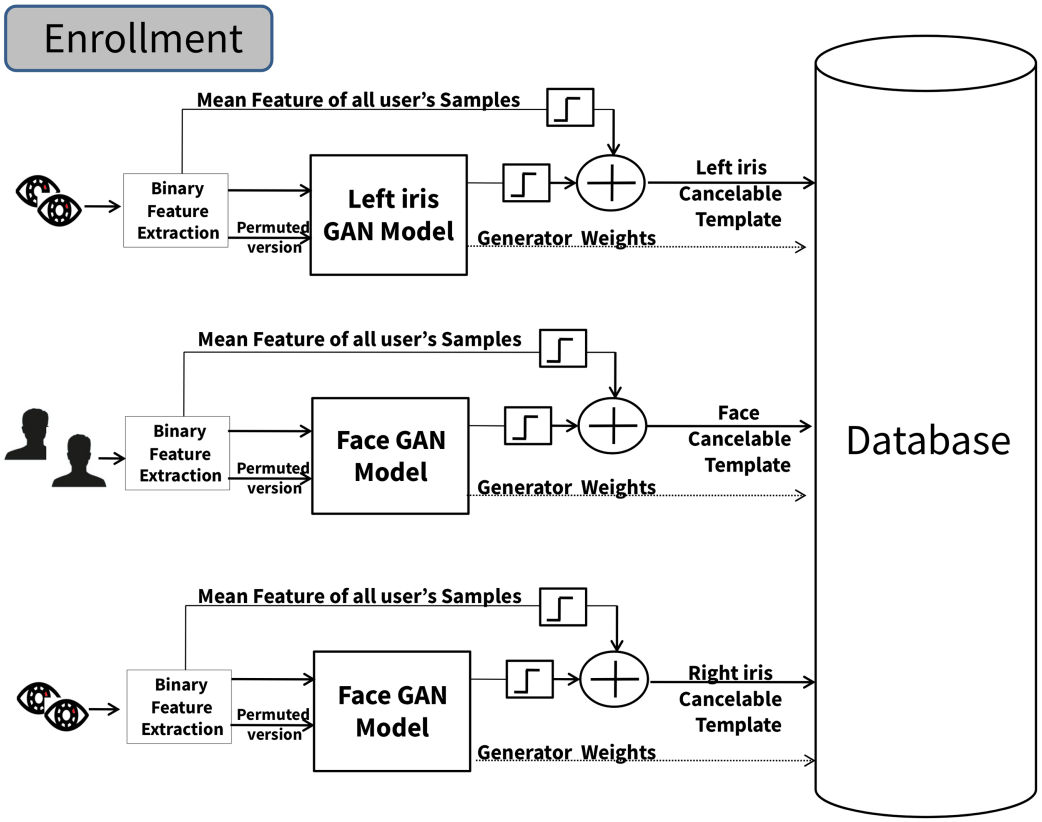

Figure 4 The multimodal cancellable decision fusion scheme (enrollment phase).
Full-size DOI: 10.7717/peerj-cs.3360/fig-4

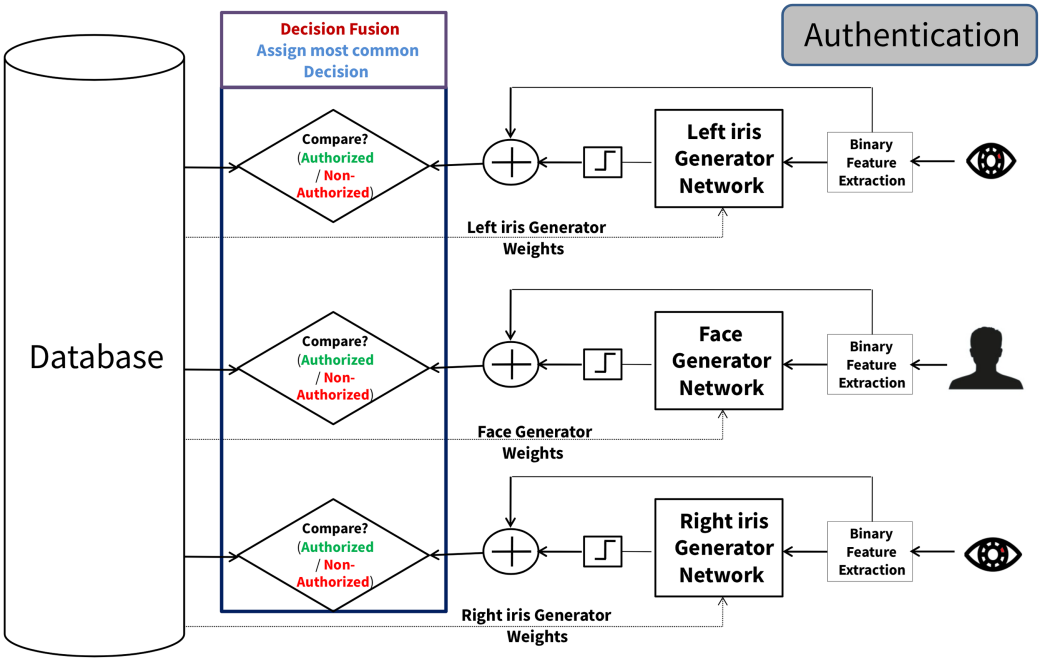

Figure 5 The multimodal cancelable decision fusion scheme (authentication phase).
Full-size DOI: 10.7717/peerj-cs.3360/fig-5

decision for each instance model. Finally, the final authentication decision is formed from the majority voting decision for the three-authentication decision for each instance model. The steps for the Decision-Level fusion scheme are illustrated in Scheme 2.

Data preprocessing

Experiments were conducted employing four publicly available datasets (CASIA, MMU1, ORL, and FERET), which are anonymized benchmark datasets provided for research purposes and were used in strict accordance with their usage policies and established ethical guidelines. *Institute of Automation, Chinese Academy of Sciences (CASIA) (2020)* comprises 146 subjects, each with left and right eye instances. *AT&T Laboratories Cambridge (2001)* comprises 400 images encompassing 40 distinct individuals. *MMU1 (2008)* contains 460 iris images for 46 subjects; each individual has five images for the left and right iris, respectively. *National Institute of Standards and Technology (NIST) (2003)* has 994 subjects; each individual has sets of five to 11 images. While each face database does not have corresponding iris images, iris samples are paired with an arbitrary (but fixed) face sample. In this study, 40 randomly selected subjects from the CASIA-V3-Internal iris dataset are paired with the ORL face dataset. At the same time, a randomly selected 40 subjects from the MMU1 iris dataset are paired with the FERET face dataset. This pairing strategy is justified by the fact that all samples originate from publicly available benchmark datasets widely used in biometric research. While the samples do not belong to the same individuals across datasets, their fusion enables a realistic simulation of multimodal authentication systems, where complementary modalities are combined to strengthen security and robustness. To mitigate concerns of dataset-induced bias, we ensured balanced sampling across genders and age groups where possible, and performance was evaluated on each dataset independently as well as in fused settings. For iris datasets, iris images were first subject to segmentation, normalization, and encoding into binary iris codes using the Libron Mask code (*Libor & Peter, 2003*). The Libron Mask algorithm is a complex sequence of procedures that results first from the Hough circular Transform and later Hough Transform linear for iris separation and identification, and after signal standardization by implementing Daugman's rubber sheet method, and, finally, conversion of the iris area in 1D Log-Gabor filtering and phase quantization, hence, embedding them into binary iris templates. Every image of each iris template went through the conversion into the same binary iris code vector, which was then utilized to feed the generator and the discriminator networks. On the other hand, binary facial features were extracted for face datasets (*Turk & Pentland, 1991*) utilizing an optimized genetic algorithm transformation (*Hamouda et al., 2016*). During the experiments, the samples are divided into 60% for the training set and 40% for the test set. The generator and discriminator networks have the same number of neurons in the input layer, while the discriminator network is crowded with a single neuron within its output layer. In each fused scheme, random weight initialization was applied to the generator and discriminator networks. The algorithms have been implemented using MATLAB R2023a (The MathWorks Inc., Natick, MA, USA). All the experiments are performed on the same computing infrastructure with attributes shown in Table 3.

Table 3 Attributes of used machine.

Processor	AMD Ryzen 7, CPU 3.20 GHz
Memory	16 GB
Operating system	Windows 11

Evaluation methods

In biometric authentication systems, three critical performance metrics are the Equal Error Rate (EER), False Acceptance Rate (FAR), False Rejection Rate (FRR). FAR represents the probability that an unauthorized individual is incorrectly accepted by the system, indicating a security vulnerability. Conversely, FRR measures the likelihood that a legitimate user is mistakenly rejected, leading to usability concerns. These rates are inversely related; reducing one typically increases the other. The EER is the point at which FAR and FRR are equal, serving as a balanced measure of system performance. A lower EER signifies a more accurate and reliable biometric system, making it a widely used metric for comparing different authentication methods (*Jain, Ross & Prabhakar, 2004*). Equations (12), (13), and (14) define the False Acceptance Rate (FAR), False Rejection Rate (FRR), and Equal Error Rate (EER), respectively, True Positive (TP) refers to correctly accepted genuine users, while True Negative (TN) represents correctly rejected impostors. False Positive (FP) occurs when an unauthorized individual is mistakenly accepted, leading to a security breach. Conversely, False Negative (FN) happens when a legitimate user is wrongly rejected, affecting system usability.

$$FAR = \frac{FP}{FP + TN} \tag{12}$$

$$FRR = \frac{FN}{FN + TP} \tag{13}$$

$$EER = \frac{FAR + FRR}{2}. \tag{14}$$

In parallel, performance can also be assessed using precision, recall, and the F1-score. Precision reflects the proportion of correctly identified positive instances among all positive predictions, while recall (or sensitivity) captures the proportion of actual positives correctly recognized by the system. The F1-score, computed as the harmonic mean of precision and recall, integrates these two complementary measures into a single metric. This ensures that the evaluation considers both predictive accuracy and completeness, thereby offering a robust assessment of system performance without bias toward one measure at the expense of the other. Equations (15), (16), and (17) define the precision, recall, and F1-score, respectively.

$$\text{Precision} = \frac{TP}{FP + TP} \tag{15}$$

$$\text{Recall} = \frac{TP}{FN + TP} \tag{16}$$

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{17}$$

Table 4 GAN control parameter effects on EER.

Exp_a		Exp_b		Exp_c		Exp_d	
h_1	EER_a	h_2	EER_b	N	EER_c	alpha	EER_d
16	1.95	16	2.10	10	2.15	0.1	2.18
32	**1.80**	32	**1.89**	50	**1.82**	0.01	2.17
64	1.99	64	1.93	100	1.96	0.001	2.16
128	1.90	128	2.08	200	2.04	0.00001	**1.92**

Note:
The best-found performance is highlighted in bold.

RESULTS

In this section, we present the substantiation of how our proposed approaches significantly enhance recognition performance, highlighting their primary contribution.

Experimental setup

Identifying the optimal combination of control parameters (hidden layer size for networks, learning rate, and number of training epochs) requires an extensive number of possible combinations. While optimization techniques could efficiently identify the best parameter configuration for GAN; this work mainly focuses on utilizing GAN to protect multibiometric schemes. Four preliminary experiments were conducted to investigate the effect of GAN control parameters. The first experiment (Exp_a) explores the effect of hidden layer size for generator networks while fixing other parameters (h2 = 32, N = 50, and alpha = 0.00001). The second experiment (Exp_b) explores the effect of hidden layer size for discriminator networks while fixing other parameters (h1 = 32, N = 50, and alpha= 0.00001). The third experiment (Exp_c) explorers the effect of the number of training epochs while fixing other parameters (h1 = 32, h2 = 32, and alpha = 0.00001). Eventually, the fourth experiment (Exp_d) explores the effect of the learning rate while fixing other parameters (h1 = 32, h2 = 32, and N = 50). Table 4 shows the effect of the GAN control parameter on EER. As shown in Table 4, the best-found performance is achieved when h1 equals 32, h2 equals 32, N equals 50, and alpha equals 0.00001. The values of these control parameters are fixed for the following experiments. The experimental parameters settings are presented in Table 5.

Each GAN model is designed with a lightweight yet effective architecture tailored for binary biometric feature transformation. The generator consists of three fully connected layers: an input layer, a hidden layer activated by the Rectified Linear Unit (ReLU) function, and an output layer of the same size as the input, activated using the Sigmoid function to ensure binary-like outputs. The discriminator is composed also of one hidden using Leaky ReLU activation, followed by a final Sigmoid layer that outputs the probability of the input being real or generated. The model is trained with an effective batch size of 64, a learning rate of 0.00001, Training is conducted for 50 epochs, with binary cross-entropy as the loss function. This configuration balances computational efficiency with the ability to learn discriminative feature transformations while avoiding overfitting.

Table 5 Experimental parameters.

Parameters	Feature-level fusion scheme	GAN-based fusion scheme	Decision-level fusion scheme
Number of GAN's model	**1**	**3**	**3**
Feature size	19,400	9,600 (left iris GAN model) 9,600 (right iris GAN model) 200 (face GAN model)	
Generator structure	19,400 × 32 × 19,400	9,600 × 32 × 9,600 (left iris GAN model) 9,600 × 32 × 9,600 (right iris GAN model) 200 × 32 × 200 (face GAN model)	
Discriminator structure	19,400 × 32 × 1	9,600 × 32 × 1 (left iris GAN model) 9,600 × 32 × 1 (right iris GAN model) 200 × 32 × 1 (face GAN model)	

Figure 6 highlights the encouraging recognition accuracy demonstrated by the curves depicting the FAR and FRR. Notably, the point of intersection between these curves signifies the hamming distance threshold (θ) employed during the verification phase. This threshold results in a minimal EER value of 2.03% and 0.82% for the feature-level and GAN-based fusion schemes, respectively, for CASIA and ORL datasets, where it achieves EER values of 1.53% and 0.80% for MMU1 and FERET datasets. While the decision-level fusion scheme, FAR and FRR are computed in terms of (the number of FN trails, the number of FP trails, the number of TN trails, and the number of TP trail) as in Eqs. (12), (13), and (14). The EER for the decision-level fusion scheme was 0.0297% for CASIA and ORL datasets and 0.0313% for MMU1 and FERET datasets.

To analyze our proposed schemes further, we compared recognition accuracy against the original unimodal with a multi-model for iris and face instances. The ROC curves are depicted in Fig. 7 for the original iris instances represented by the left and right iris curves, original face instances, and multimodal iris and face instances. Both schemes outperform alternative unimodal iris/face instance models. Furthermore, Fig. 8 represents a histogram diagram comparative assessment of recognition accuracy values against unimodal GAN instances represented by the left and right iris, unimodal GAN face with the multimodal iris and face proposed schemes. It can be concluded that the proposed fusion schemes outperform alternative unimodal (iris/face) instance models. Also, it is worth noting that the proposed feature-level fusion scheme exhibits a minor decrease in recognition accuracy compared to the other proposed schemes. On the other hand, the decision-level fusion scheme achieves the best recognition accuracy.

Eventually, the proposed multimodal biometric schemes are compared to recent multimodal biometric schemes published in the literature. Table 6 shows the results of this comparison, including the type of biometric data, the fusion type, and EER. Our proposed schemes improve the recognition accuracy of multimodal biometric systems compared to existing schemes. However, the feature-level fusion scheme reduces the recognition accuracy compared to other proposed schemes. This is because of the curse of

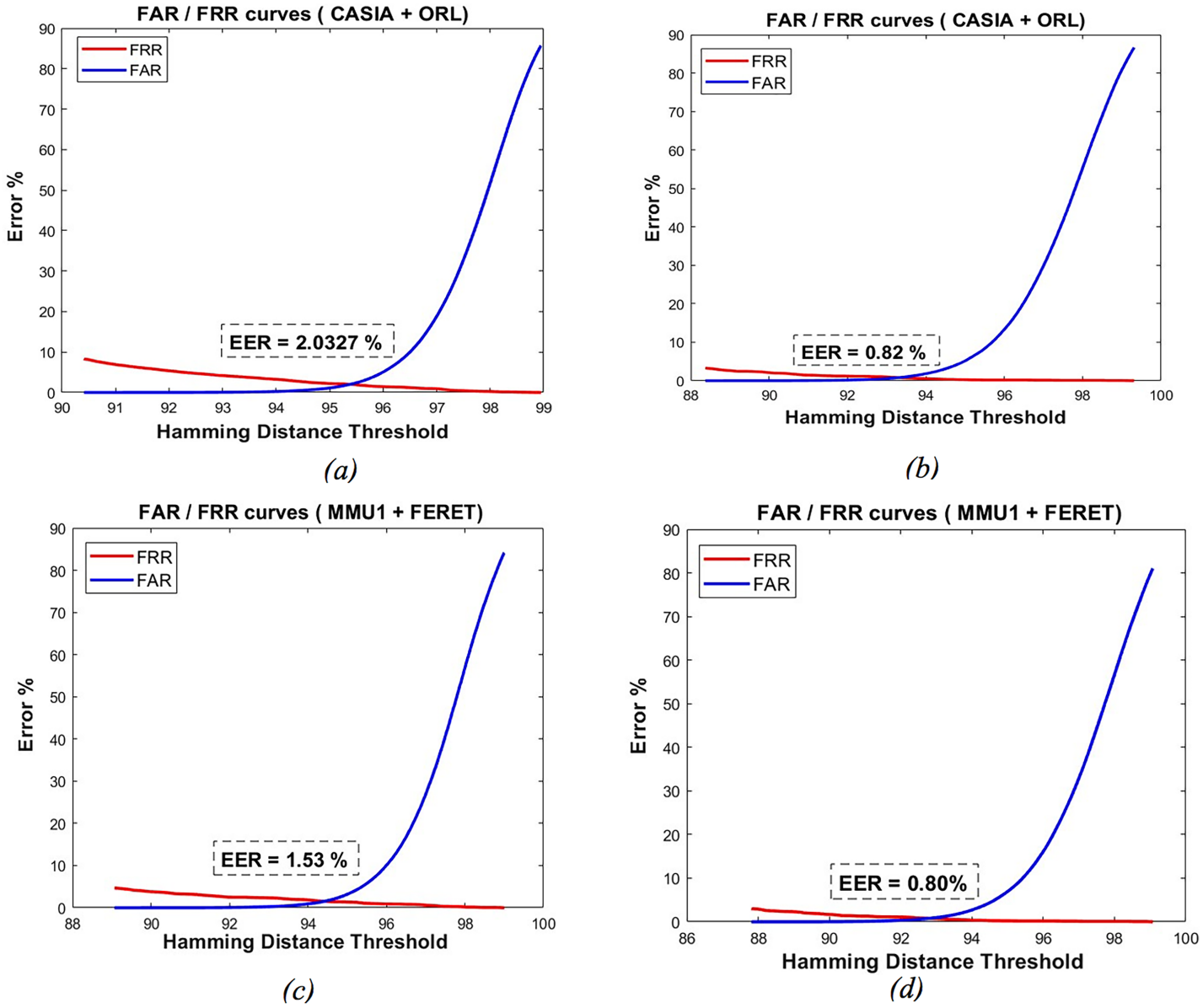

Figure 6 **FAR and FRR curves for the proposed schemes: feature-Level fusion scheme (A, C) and GAN-based fusion scheme (B, D).**
Full-size DOI: 10.7717/peerj-cs.3360/fig-6

dimensionality, which occurs when combining features from multiple sources create a new feature space with much higher dimensionality than any original. This can make it challenging to learn a good predictive model.

On the other hand, the decision-level fusion approach outperforms the other proposed schemes, achieving a F1-score of 98.76%. This is because the classifiers are more likely to agree on the correct prediction than to disagree. Since it operates only on final decisions (*e.g.*, accept/reject or 1/0), it enhances security and reduces the risk of data leakage. Moreover, this strategy simplifies the integration of different biometric modalities or

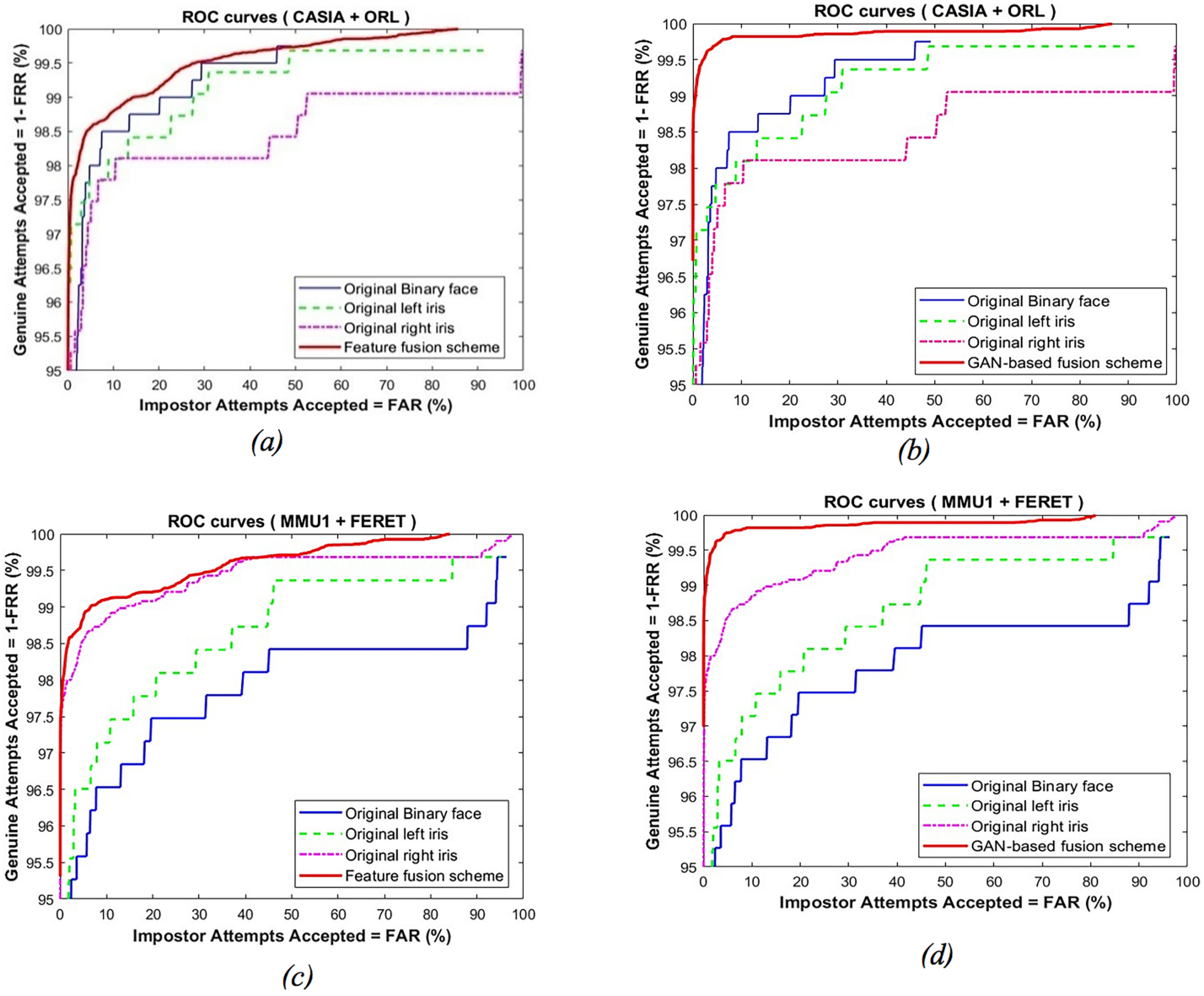

Figure 7 **ROC curves comparison among feature-leve fusion scheme (A, C) and GAN-based fusion scheme (B, D) compared to its unimodal instance.**
Full-size DOI: 10.7717/peerj-cs.3360/fig-7

vendor systems, which is one of the key reasons why decision-level fusion achieves higher recognition performance compared to other fusion schemes.

The promising performance results come at a specific computational cost. The proposed method utilizes GANs for synthetic template generation, the most computationally intensive step. The training process has a time complexity of O ($N * P * L * (h1 + h2)$), where N represents the number of training epochs, P represents the number of training samples per epoch, L represents the length of the input biometric template, and $h1$ and $h2$ represent hidden layer sizes of generator and discriminator networks, respectively. The

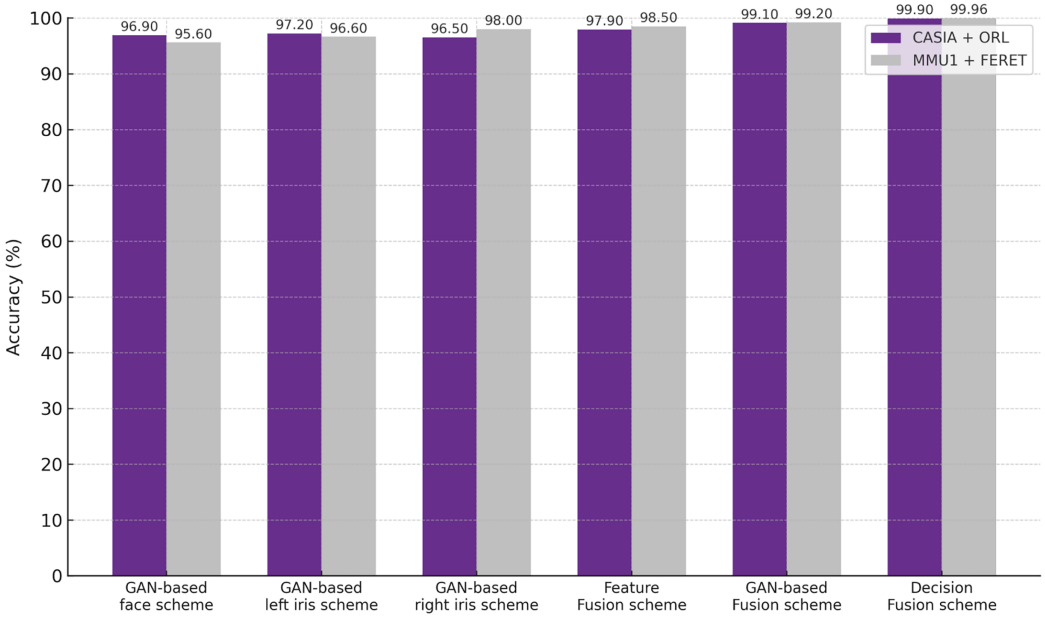

Figure 8 **Recognition accuracy comparison among the multimodal schemes and its unimodal instance.** Full-size DOI: 10.7717/peerj-cs.3360/fig-8

Table 6 **Comparative study analysis, a indicates (CASIA and ORL) datasets and b indicates (MMU1 and FERET) datasets.**

Method	Year	Biometric modality	Fusion type	EER (%)
Hili et al. (2016)	2016	Face + iris	Score-level	0.63
Miao et al. (2017)	2017	Face + iris	Score-level	0.39
Mostafa et al. (2020)	2020	Face expressions	Score-level	4.20
Sudhakar & Gavrilova (2020)	2020	Multi-instance iris + finger vein	Feature-level	0.12
Balraj & Abirami (2022)	2022	Multi-instance iris	Score-level	3.58
Salturk & Kahraman (2024)	2024	Signature + Face	Feature-level	1.99
Proposed feature-level fusion	–	Multi-instance iris + face	Feature-level	2.032
				1.53
Proposed GAN-based fusion	–	Multi-instance iris + face	Feature-level	0.820
				0.800
Proposed decision-level fusion	–	Multi-instance iris + face	Decision-level	0.0297
				0.0313

value of L depends on the applied fusion level; it represents the sizes of iris/face data. The feature extraction step also adds complexity; however, its computational load in comparison to the training phase is relatively marginal, especially for larger input sizes n and f. It is important to note that training typically occurs offline during system setup. Although the complexity is a consideration, the improved security offered by the multimodal, GAN-based approach outweighs the computational cost in the targeted application. The computational time for training the schemes, are approximately 1,500 s

for the proposed feature-level fusion, 420 s for the proposed GAN-based fusion, and 300 s for the proposed decision-level fusion.

DISCUSSION

From a security standpoint, the proposed multimodal cancelable fusion schemes are regarded as secure schemes if they meet the following criteria:

Recoverability, diversity, and unlinkability: Recoverability is the ability to generate multiple versions of a cancelable template from the same original biometric data. The proposed schemes address this by introducing a new salting key based on a random permutation of the biometric data. Because this permutation process has many possible combinations, it allows us to create variations of the salting key for a single person. The recoverability property is fulfilled with a total number of different salting keys equals $(2n + f)!$ for the first proposed scheme and $(n! * n! * f!)$ for the second and third proposed schemes, where n and f represent the iris and face feature sizes, respectively. The number of possible permutations grows very quickly as the size of the biometric data increases. This vast number guarantees recoverability. By changing the salting key, each user's template becomes unique and unpredictable, making it harder for attackers to exploit (diversity) and cannot be linked together for the same user (unlinkability).

Non-invertibility: The schemes are non-invertible when the original characteristics of the biometric instance cannot be deduced from the parameters stored in the authentication system database. For the proposed schemes, these parameters are the weight values of the generators and the stored cancellable template. Prior mathematical analysis in *Tarek, Ouda & Hamza (2016)* has established that recovering the input or output of a neural network model using only the network's weights is computationally challenging. Consequently, the final stored cancelable template results from the X-OR operation involving a binarized version of the network generator's output and another template. This design ensures no direct stored information regarding the inputs or outputs of the GAN networks. Furthermore, the inherent irreversibility property of the XOR function, which prevents the retrieval of its inputs based only on its output, significantly hampers any potential attacker's ability to computationally recover the network generator's inputs (*e.g.*, the mean of input biometric templates) using the stored cancelable template. Consequently, the proposed schemes adhere to the non-invertibility property. This compliance stems from the irreversible nature of the stored network weights, which makes it arduous to deduce the exact network input or output, and the irreversibility inherent to the XOR process makes it exceptionally difficult to unveil its inputs using only the cancellable template. Moreover, experimental evaluation shows that attempting to reconstruct the original biometric from the stored cancelable templates results in an average reconstruction error above 95%, confirming that the stored data does not reveal meaningful information about the original inputs.

Resistance to pre-image and correlation attacks: A pre-image attack involves crafting synthetic biometric features based on the parameters stored in a biometric system

database, aiming to pass them off as genuine features for authentication (*Tarek, Ouda & Hamza, 2017*). As previously mentioned, the stored parameters within the proposed schemes are effectively useless to potential attackers. Consequently, attempting to construct pre-image biometric features would prove unsuccessful. Thus, the pre-image attack becomes computationally daunting, similar to a brute force attack. The attacker would require $2^{(2n+f)}$ trials to generate a disclosed biometric feature in the feature-level fusion scheme and $2^{(2n+2n+f)}$ trials in the GAN-based and decision-level fusion schemes. Consequently, attacking a biometric system with a larger feature size would be even more challenging. On the other hand, a correlation attack seeks to recover the original biometric template by correlating various cancelable templates derived from the same biometric traits (*Cai & Jiankun, 2014*). As previously explained, the schemes ensure distinct unlinkable cancelable templates for the same biometric instances across diverse biometric systems through distinct permutations for feature instances and distinct initializations of GAN networks' weights. In conclusion, the proposed fusion schemes are secure against various attack types. To further support unlinkability, we computed correlation scores between cancelable templates generated from the same biometric under different permutations and GAN initializations. The results showed near-zero correlation values, demonstrating that the templates cannot be linked back to the same user across systems.

Real-world applications and operational considerations

This subsection discusses the possible practical implementations of the proposed multimodal user authentication schemes. In addition, the operational aspects of implementation are analyzed. The introduced technology is very convenient for real-life applications requiring robust and reliable security. It can be summarized as follows:

- **Access control systems:** In high-security environments (*e.g.*, data centers, government buildings, and research laboratories), multimodal biometric systems are heavily required. The proposed method can contribute to developing access control systems by making them more secure, private, user-friendly, and flexible. The irreversible templates generated by GANs make forgery much harder for fraudsters. In addition, the system ensures users' privacy as it does not save the original biometric data. Eventually, the multimodal features will make biometric recognition user-friendly, even with lighting or pose changes.
- **Border security:** Accurate and secure identification of persons is a critical part of any border security application. The proposed method can be helpful for border control as it would allow border control to guarantee privacy. The advantage of using multiple modalities and synthetic templates will significantly limit the risk of unauthorized access and increase accuracy. Furthermore, privacy issues are addressed as the system does not store original biometric data.
- **Law enforcement:** Law enforcement situations often require robust and spoof-resistant identification. Law enforcement personnel can use the proposed method to secure authentication using multimodal biometrics with synthetic templates. Additionally, the

proposed method can be used for accurate identification of suspects in sensitive situations.

Despite the mentioned benefits, the proposed method encounters some implementation operational issues, such as computational cost, data collection privacy, and system integration. The training phase of the proposed GAN-based models can be too computationally intense for real-world-based applications. When it happens during offline system setup, this issue can be resolved through cloud training offering that reduces any processing onboard. Furthermore, model optimization and resource reduction, for instance, could be used along the way. On the one hand, consent should be obtained for data collection and anonymization of GAN training. However, this problem may be overcome by establishing rigorous user consent procedures for collecting biometric data. In addition, differential privacy methods for anonymity in the data used for training should be used. Eventually, integrating the proposed technique with current security programs could require further improvement. The integration problem can be overcome with modularity to allow future integration and cooperation with security system manufacturers on integration protocol development. The advantages of security and privacy are the main focus of this research. However, multimodal biometrics with enhanced efficiency and robustness may also bring economic benefits and cost savings in the future for the companies that employ this technology.

CONCLUSIONS

This article introduces keyless multimodal cancelable biometrics. Three fusion-level schemes are presented: fusion at the feature level for the first scheme, GAN-based level for the second, and the decision level for the third. The generative adversarial network is adopted as a cancelable transformation function to secure the biometrics data. Furthermore, the schemes employ a random permutation salting key extracted from the input biometric data, eliminating the need for external storage of keys and avoiding possible security breaches. The proposed method effectively tackles various security challenges of biometric salting methods, such as non-invertibility, recoverability, and diversity. In addition, their resistance to pre-image and correlation attacks is also addressed. Experimental simulations based on several iris and face datasets show promising recognition performance for the proposed multimodal schemes compared to the unimodal schemes. The slight decrease in accuracy observed from the feature-level fusion scheme, compared to the GAN-based and decision-level fusion schemes, is considered acceptable, given the overall performance improvement. However, despite these advantages, the proposed system has certain limitations. The computational complexity introduced by the GAN-based transformation may impact real-time performance, making it less suitable for resource-constrained environments. Additionally, while the system demonstrates improved security, its robustness against sophisticated adversarial attacks requires further investigation.

Future work will focus on several directions. Comprehensive empirical simulations of adversarial and correlation attacks will be conducted to further assess robustness under

practical scenarios. Strategies for scaling the approach to efficiently handle significantly larger datasets will also be explored, facilitating deployment in real-world biometric systems. While the implementation of optimization methods, such as pruned or lightweight generator architectures, is beyond the scope of the current study, we recognize their potential to reduce computational overhead. These strategies are therefore highlighted as a future research direction, aimed at enabling deployment in resource-constrained environments. While these studies are extensive and beyond the scope of the current work, they are considered essential for a dedicated future investigation.

ACKNOWLEDGEMENTS

An AI tool (ChatGPT, Gemini) was used only to improve the language and readability of this manuscript. All ideas, research, analysis, and conclusions are the authors' own, and the final version has been fully reviewed and approved by the authors.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Mayada Tarek conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Eslam Hamouda conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Amjad Alsirhani performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Abdullah Alomari performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Ayman Mohamed Mostafa conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The code is available in the Supplemental File.

The CASIA dataset is available at: Chinese Academy of Sciences (2020) CASIA iris image database, v3.

http://english.ia.cas.cn/db/201610/t20161026_169399.html.

The ORL dataset is available at: AT&T Laboratories Cambridge (2001) ORL face image database, v1,

http://cam-orl.co.uk/facedatabase.html.

The MMU1 dataset is available at Kaggle: https://www.kaggle.com/datasets/naureenmohammad/mmu-iris-dataset.

The FERET dataset is available at NIST: Facial Image Database (2003) Image Group, Information Access Division, ITL, National Institute of Standards and Technology.

https://www.nist.gov/itl/products-and-services/color-feret-database.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.3360#supplemental-information.

REFERENCES

Abdellatef E, Ismail NA, Abd Elrahman SE, Ismail KN, Rihan M, Amin M, Eisa AA, El-Samie FE. 2020. Cancelable multibiometric recognition system based on deep learning. *The Visual Computer* **36(2)**:1097–1109 DOI 10.1007/s00371-019-01715-5.

Aditya A, Kaur S. 2021. Novel methods for multimodal biometric systems to strengthen the security. In: *2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST)*, 99–105 DOI 10.1109/ICCMST54943.2021.00031.

Adusumalli B, Bhuvaneswari S. 2018. A study of multimodal biometric system recognition using different modalities and fusion techniques. *International Journal of Research Studies in Computer Science and Engineering (IJRSCSE)* **5(4)**:38–41.

Alqahtani H, Kavakli-Thorne M, Kumar G. 2019. Applications of generative adversarial networks (GANs): an updated review. *Archives of Computational Methods in Engineering* **28(2)**:525–552 DOI 10.1007/s11831-019-09388-y.

AT&T Laboratories Cambridge. 2001. The database of faces. *Available at http://cam-orl.co.uk/facedatabase.html*.

Balraj E, Abirami T. 2022. Performance improvement of multibiometric authentication system using score level fusion with ant colony optimization. *Wireless Communications and Mobile Computing* **2022(1)**:453 DOI 10.1155/2022/4145785.

Cai L, Jiankun H. 2014. Attacks via record multiplicity on cancelable biometrics templates. *Concurrency and Computation: Practice and Experience* **26(8)**:1593–1605 DOI 10.1002/cpe.3042.

Choudhury B, Then PH, Issac B, Raman V, Haldar MK. 2018. A survey on biometrics and cancelable biometrics systems. *International Journal of Image and Graphics* **18(1)**:1850006 DOI 10.1142/S0219467818500067.

Creswell A, Bharath AA. 2018. Inverting the generator of a generative adversarial network. *IEEE Transactions on Neural Networks and Learning Systems* **30(7)**:1967–1974 DOI 10.1109/TNNLS.2018.2875194.

El Rahman SA, Alluhaidan AS. 2024. Enhanced multimodal biometric recognition systems based on deep learning and traditional methods in smart environments. *PLOS ONE* **19**:e0291084 DOI 10.1371/journal.Pone.0291084.

El-Hameed HAA, Ramadan N, El-Shafai W, Khalaf AAM, Ahmed HEH, Elkhamy SE, El-Samie FEA. 2021. Cancelable biometric security system based on advanced chaotic maps. *The Visual Computer* **38(6)**:2171–2187 DOI 10.1007/s00371-021-02276-2.

El-Rahiem BA, Amin M, Sedik A, Samie FE, Iliyasu AM. 2021. An efficient multibiometric cancellable biometric scheme based on deep fusion and deep dreams. *Journal of Ambient Intelligence and Humanized Computing* **3(4)**:2177–2189 DOI 10.1007/s12652-021-03513-1.

Feifei C, Gonging Y. 2011. Score level fusion of fingerprint and finger vein recognition. *Journal of Computational Information Systems* **7(16)**:5723–5731.

Goodfellow IJ, Abadie JP, Mirza M, Xu B, Farley DW, Ozair S, Courville A, Bengio Y. 2014. Generative adversarial nets, NIPS. ArXiv DOI 10.48550/arXiv.1406.2661.

Govindarajan RK. 2004. Feature level fusion in multimodal biometric. MS thesis. Statler College of Engineering and Mineral Resources, Morgantown, WV, USA. DOI 10.33915/etd.1489.

Gupta D. 2015. Multimodel biometric system: fusion techniques and their comparison. In: *The 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS)*, 1–4 DOI 10.1109/RAECS.2015.7453287.

Haider SA, Ashraf S, Larik RM, Husain N, Muqeet HA, Humayun U, Yahya A, Arfeen ZA, Khan MF. 2023. An improved multimodal biometric identification system employing score-level fuzzification of finger texture and finger vein biometrics. *Sensors* **23**:9706 DOI 10.3390/s23249706.

Hamouda E, Ouda O, Yuan X, Hamza T. 2016. Optimizing discriminability of globally binarized face templates. *Arabian Journal for Science and Engineering* **41(8)**:2837–2846 DOI 10.1007/s13369-015-2020-3.

Hili NK, Montagne C, Lelandais S, Hamrouni K. 2016. Quality-dependent multimodal fusion of face and iris biometrics. In: *2016 the 6th International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 1–6 DOI 10.1109/IPTA.2016.7820954.

Institute of Automation, Chinese Academy of Sciences (CASIA). 2020. CASIA iris image database, v3. *Available at http://english.ia.cas.cn/db/201610/t20161026_169399.html*.

Jain AK, Ross A, Prabhakar S. 2004. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology* **14(1)**:4–20 DOI 10.1109/TCSVT.2003.818349.

Jain AK, Ross A. 2004. Multibiometric systems. *Communications of the ACM* **47(1)**:34–40 DOI 10.1145/962081.962102.

Libor M, Peter K. 2003. MATLAB source code for a biometric identification system based on Iris patterns. In: *The School of Computer Science and Software Engineering*. Perth, Australia: The University of Western Australia.

Maiorana E, Campisi P, Fierrez J, Ortega-Garcia J, Neri A. 2010. Cancelable templates for sequence-based biometrics with application to on-line signature recognition. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* **40**:525–538 DOI 10.1109/TSMCA.2010.2041653.

Manisha, Kumar N. 2020. Cancelable biometrics: a comprehensive survey. *Artificial Intelligence Review* **53(5)**:3403–3446 DOI 10.1007/s10462-019-09767-8.

Merkle J, Kevenaar TA, Korte U. 2012. Multimodal and multi-instance fusion for biometric cryptosystems. In: *2012 BIOSIG - Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG)*. Piscataway: IEEE, 1–6 DOI 10.1109/biosig52210.2021.9548288.

Miao D, Zhang M, Sun Z, Tan T, He Z. 2017. Bin-based classifier fusion of iris and face biometrics. *Neurocomputing* **224(11)**:105–118 DOI 10.1016/j.neucom.2016.10.048.

MMU1. 2008. MMU1 Iris image databases, 2008 MMU1 Iris image database. Kaggle. *Available at https://www.kaggle.com/datasets/naureenmohammad/mmu-iris-dataset*.

Mostafa AH, El-Sayed HA, Mohamed A, Belal F. 2020. Benchmarking of convolutional neural networks for facial expression recognition. *Journal of Theoretical and Applied Information Technology* **98(18)**:3104–3115.

National Institute of Standards and Technology (NIST). 2003. Color FERET Database. *Available at https://www.nist.gov/itl/products-and-services/color-feret-database*.

Patel VM, Ratha NK, Chellappa R. 2015. Cancelable biometrics: a review. *IEEE Signal Processing Magazine* **32(5)**:54–65 DOI 10.1109/MSP.2015.2434151.

Patil S. 2012. A study of biometric, multimodal biometric systems: fusion techniques, applications and challenges. *International Journal of Computer Science and Technology* **3(1)**:524–526.

Paul PP, Gavrilova ML. 2012. Multimodal cancelable biometrics. In: *2012 IEEE 11th International Conference on Cognitive Informatics and Cognitive Computing*, 43–49.

Prabhakar S, Jain A. 2002. Decision-level fusion in fingerprint verification. *Pattern Recognition* **35(4)**:861–874 DOI 10.1016/S0031-3203(01)00103-0.

Rathgeb C, Breitinger F, Busch C, Baier H. 2014. On application of Bloom filters to iris biometrics. *IET Biometrics* **3**:207–218 DOI 10.1049/iet-bmt.2013.0049.

Ross A, Jain A. 2004. Multimodal biometrics: an overview. In: *The 12th European Signal Processing Conference*, 1221–1224.

Salturk S, Kahraman N. 2024. Deep learning-powered multimodal biometric authentication: integrating dynamic signatures and facial data for enhanced online security. *Neural Computing and Applications* **36(19)**:11311–11322 DOI 10.1007/s00521-024-09690-2.

Savvides M, Kumar BV, Khosla PK. 2004. Cancelable biometric filters for face recognition. In: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*, Vol. 3, 922–925.

Saxena D, Cao J. 2020. Generative adversarial networks (GANs): challenges, solutions, and future directions. *ACM Computing Surveys (CSUR)* **54(3)**:1–42 DOI 10.1145/3446374.

Sheena S, Mathew S. 2014. A study of multimodal biometrics system. *International Journal of Research in Engineering and Technology* **3(15)**:93–98 DOI 10.15623/ijret.2014.0327018.

Soliman RF, Amin M, Abd El-Samie FE. 2018. A modified cancelable biometrics scheme using random projection. *Annals of Data Science* **6**:223–236 DOI 10.1007/s40745-018-0172-1.

Sudhakar T, Gavrilova M. 2020. Deep learning for multi-instance biometric privacy. *ACM Transactions on Management Information Systems* **12(1)**:1–23 DOI 10.1145/3389683.

Tarek M, Hamouda E, El-Metwally S. 2021. Unimodal-Bio-GAN: keyless biometric salting scheme based on generative adversarial network. *IET Biometrics* **10(6)**:654–663 DOI 10.1049/bme2.12034.

Tarek M, Ouda O, Hamza T. 2016. Robust cancelable biometrics scheme based on neural networks. *IET Journal on Biometrics* **5(3)**:220–228 DOI 10.1049/iet-bmt.2015.0045.

Tarek M, Ouda O, Hamza T. 2017. Pre-image resistant cancelable biometrics scheme using bidirectional memory model. *International Journal of Network Security* **19(4)**:498–506.

Teoh A, Ngo DC, Goh A. 2004. Biohashing: two factor authentication featuring fingerprint data and tokenised random number. *Pattern Recognition* **37(11)**:2245–2255 DOI 10.1016/j.patcog.2004.04.011.

Turk M, Pentland A. 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience* **3(1)**:71–86 DOI 10.1162/jocn.1991.3.1.71.