IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 September 2024, accepted 18 November 2024. Date of publication 00 xxxx 0000, date of current version 00 xxxx 0000.

Digital Object Identifier 10.1109/ACCESS.2024.3512003

For Your Eyes Only: A Privacy-Preserving Authentication Framework Based on Homomorphic Encryption and Retina Biometrics

DAVID PALMA AND PIER LUCA MONTESSORO, (Member, IEEE)
Polytechnic Department of Engineering and Architecture, University of Udine, 33100 Udine, Italy
Corresponding author: David Palma (david.palma@uniud.it)

ABSTRACT Securing personal information and data has become an imperative challenge, especially after the introduction of legal frameworks, such as, in Europe, the General Data Protection Regulation (GDPR). Traditional authentication methods, such as PINs and passwords, are vulnerable to cyber threats, underscoring the need for more robust biometric systems. These systems offer improved security by accurately verifying a user's identity, reducing the risk of impersonation. The human retina has demonstrated remarkable reliability as a biometric trait mainly because of its unique and stable patterns, even though the adoption of these systems gives rise to significant concerns regarding the confidentiality of biometric data. This study presents a groundbreaking approach to address these concerns by integrating homomorphic encryption into retina-based authentication. The combination of homomorphic encryption and retina biometrics within the proposed framework offers a comprehensive solution that ensures both privacy and security with no loss in accuracy. The proposed approach mitigates the risks associated with possible unauthorised access and security breaches by keeping the data encrypted throughout the entire procedure. Furthermore, it preserves the individual's privacy by preventing the exposure of sensitive biometric information. We evaluated the proposed system through extensive experiments and simulations, demonstrating its effectiveness in terms of both security and privacy when the system operates in normal (ideal) and abnormal (under attack) conditions. Experimental results indicate that the combined approach offers robust resistance to various attacks, including replay attacks and data exposure, providing a robust and privacy-centric authentication solution.

INDEX TERMS Authentication, biometrics, cryptography, homomorphic encryption, information security, pattern recognition, privacy, private biometrics, retinal recognition, security attacks.

I. INTRODUCTION

In today's digital landscape, where fraud, cybercrime, and theft are on the rise, the importance of algorithms and solutions in the biometric field is growing along with security awareness. Conventional/traditional authentication methods, relying on token-based systems that use something the user has (e.g., smart card) and knowledge-based systems that use something the user knows (e.g., password or PIN), have demonstrated their vulnerability to various security threats, including brute-force attacks, data leaks, and phishing attempts [1]. Contrary to these types of authentication methods, biometric-based approaches are inherently tied to the person themselves, making it much harder for unauthorised individuals to impersonate someone else. As a result, biometric authentication methods have gained prominence in ensuring the identity and security of users in various applications, ranging from personal devices to critical infrastructures [2]. Biometrics makes use of the unique physiological/biological or behavioural characteristics of individuals, such as facial features, fingerprints,iris patterns, keystroke dynamics, and voice for authentication purposes. The biometric data usage

The associate editor coordinating the review of this manuscript and approving it for publication was M. Sabarimalai Manikandan.

for authentication offers several advantages, including the elimination of the need to remember complex passwords, reducing the risk of identity theft, and enhancing user's experience [3]. Among the diverse biometric modalities, retina-based recognition has emerged as a promising frontier owing to its high accuracy and reliability. The retina, located at the back of the eye, contains a unique and intricate pattern of blood vessels that remains stable throughout an person's life. This pattern, known as the retinal vascular network, has garnered significant attention in recent years for its potential in biometric authentication. By integrating advanced imaging technologies and pattern recognition algorithms, retina-based Biometric Authentication Systems (BASs) are able to detect and analyse the distinctive features of a person's retinal vasculature. This process results in the formation of a secure and virtually inviolable biometric template [4]. The necessity for a live individual's presence during the scanning procedure further strengthens the system defences against cyber threats such as spoofing and tampering, except when a sensor module is compromised [5]. Unlike facial recognition, which can be fooled by photographs or videos, or fingerprint scanners, which can be deceived with artificial replicas, the retina is virtually impossible to replicate convincingly. Capturing a retinal image requires specialised equipment that emits near-infrared light and accurately captures the reflected patterns, making this technology difficult to deceive. Furthermore, many modern retina scanners feature mechanisms to verify that the retina being scanned belongs to a live person by detecting physiological responses such as blood flow, thus preventing the use of artificial or static images. Additionally, retina-based recognition systems have demonstrated exceptional accuracy, achieving False Acceptance Rates (FAR) and False Rejection Rates (FRR) that are orders of magnitude lower than the ones of other biometric methods. This high level of accuracy makes the retina an ideal choice as a biometric trait for user-centric security-critical applications [6].

A conventional retinal recognition system comprises two essential phases: enrolment and authentication. In both phases, the process starts with capturing a digital image of the user's retina using a specialised scanner that employs low-intensity infrared light to illuminate the retina. This non-invasive procedure is safe for the user's eyes, as the infrared light passes through the eye lens and the reflected patterns are captured by the scanner. The captured image is then pre-processed to remove artifacts and enhance contrast, ensuring the clarity of the retinal blood vessel patterns. These patterns are used to extract distinctive features and create a retinal template. During the enrolment phase, this template is stored as a reference for future authentication attempts. When authentication is required, the newly acquired retinal image undergoes the same pre-processing steps as in the enrolment phase. The system then proceeds to compare the features extracted from this image with the stored template in the database [7]. Storing these biometric templates without proper security exposes them to a range of threats, making them susceptible to unauthorised access and potential misuse. However, it's essential to recognise that a biometric security solution comprises multiple components, with the recognition module addressing authentication as just one aspect (see [5] for a comprehensive overview of BAS vulnerabilities).

Common attacks against template protection include [8]: (i) masquerade attack, which occurs when a malicious actor attempts to impersonate a legitimate user by presenting a fake biometric sample; (ii) correlation attack, where a malicious actor tries to correlate different templates to establish a link between them, potentially revealing sensitive information; (iii) replay attack, here a malicious actor captures a user's biometric data during authentication and later replays it to gain unauthorised access; (iv) stolen-token attack, that involves the theft of a biometric token or representation, which can be used to impersonate the genuine user; (v) hill-climbing forged data attack, where a malicious actor may exploit the search space of a biometric template to find a close match, potentially bypassing security measures; (vi) model inversion attack, which involve machine learning or statistical techniques to infer information about a user's biometric data based on the responses or decisions made by a biometric system.

Hence, although retina-based recognition offers remarkable accuracy in identity authentication, concerns regarding the storage and transmission of sensitive biometric data persist. Privacy preservation in biometric authentication is a critical issue that must be faced to guarantee the widespread adoption of these technologies. The need to protect biometric data from unauthorised access, misuse, and potential breaches has led to the development of privacy-preserving techniques and protocols [9]. A straightforward method for achieving privacy in biometric recognition involves the use of standard encryption algorithms like AES and DES. However, because environmental conditions invariably alter biometric templates, these traditional ciphers require decryption before comparing biometric data, thereby exposing sensitive information during the process. This is due to the influence of the avalanche effect on these cryptographic standards [10], which cannot be directly employed in practical applications despite their robust theoretical foundations, as they only apply to precise data. The primary goal of many systems discussed in literature, which focus on biometric data privacy protection, is to alter the stored biometric templates to prevent unauthorised individuals from accessing this information. Some of these methods are summarised in Figure 1.

Cancelable features allow the transformation of retina biometric templates into a non-invertible format, ensuring that the original biometric data remains protected even if the template is compromised. In general, the design criteria for cancelable biometrics are:

1) diversity: it is imperative to avoid cross-matching between templates of the same user across various applications;

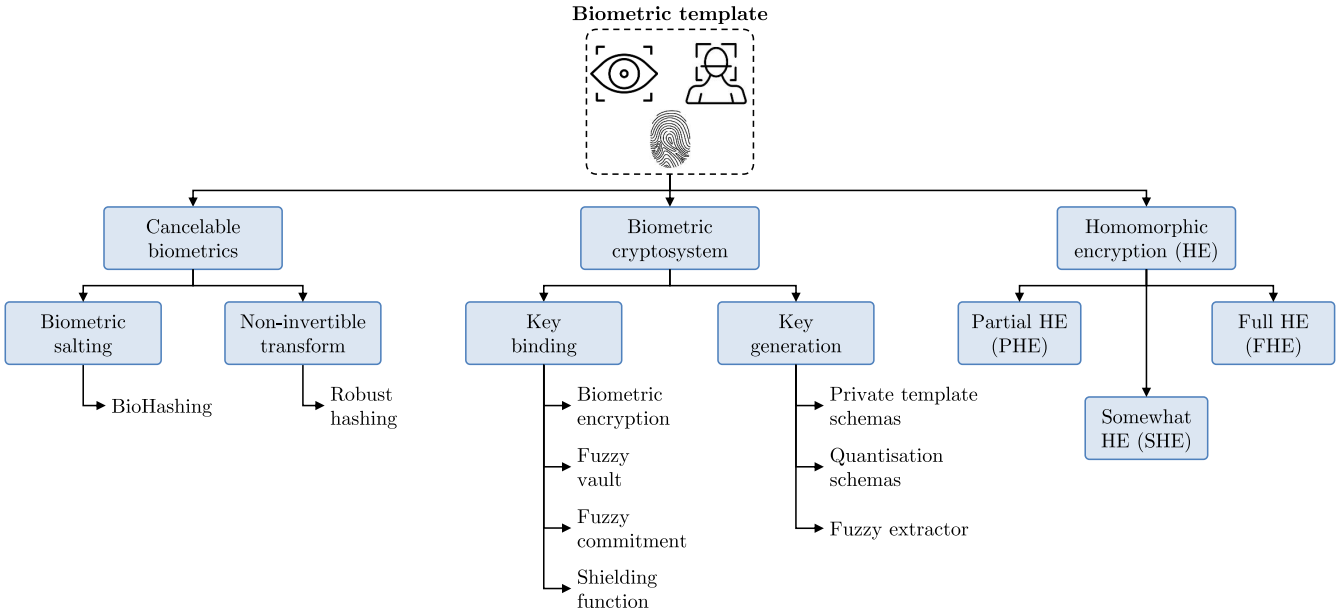


FIGURE 1. Classification of biometric template protection schemes.

2) revocability: a new protected template can be re-issued once the previous one is compromised;
3) non-invertibility: it should be computationally infeasible to obtain the original biometric data from the protected template so that the unprotected biometric features can never be reconstructed;
4) performance: the accuracy performance in terms of recognition of an unsecured system should either be upheld or improved with respect to the secured one.

Most biometric cryptosystems, instead, rely on storing biometric-dependent public information, known as helper data, to facilitate key retrieval or generation. Due to the variability in biometric traits, it is generally not possible to directly extract keys from them. Helper data, while ensuring it doesn't disclose significant details about the original biometric templates, plays a vital role in the key reconstruction process. Biometric comparisons are carried out indirectly by confirming key validity, resulting in either the release of a key or a failure message during the authentication process. As the verification of keys involves biometric comparisons within an encrypted domain, biometric cryptosystems are employed to secure biometric templates and enable biometric-dependent key release. Depending on the method used to derive helper data, biometric cryptosystems are categorised as either key-binding or key-generation systems. Nevertheless, biometric cryptosystems and cancelable biometrics show limitations when applied to unprotected data or when they depend on helper data for verification. Compromising this helper data could lead to the exposure of sensitive information, thereby impacting both the system security and the individual's privacy. An alternative solution to these methods that rely on auxiliary data makes use of Homomorphic Encryption (HE), which enhances privacy by enabling secure operations such as additions and multiplications on the encrypted data. Homomorphic encryption can be categorised into three main types: (i) Partial Homomorphic Encryption (PHE) allows operations of either addition or multiplication within an encrypted domain, but not both simultaneously; (ii) Somewhat Homomorphic Encryption (SHE) enables both additions and multiplications within the encrypted domain, but there are limitations on the number of times these operations can be performed; (iii) Fully Homomorphic Encryption (FHE) empowers unlimited additions and multiplications within the encrypted domain. A number of HE schemes involving different mathematical manipulations have been proposed in literature within the above-mentioned categories. In particular, PHE-based approaches include: (i) RSA [11], which is one of the first public key encryption method and is considered the first multiplicative PHE [12]; (ii) ElGamal [13], derived from Diffie-Hellman key exchange algorithm, its security is guaranteed by the computational hardness assumption about the decisional discrete logarithm problem; Paillier [14], a probabilistic public key encryption scheme based on the composite residuosity problem that implements homomorphic addition. Notwithstanding several research studies in biometrics have made use of PHE-based approaches like ElGamal and Paillier, these methods have been proved to be not secure against quantum attacks, so that the scientific community have proposed alternative HE methods. One of the alternatives relies on SHE-based approaches, which include: (i) CKKS [15], that consists of a scheme that enables HE for approximate addition and multiplication operations on ciphertexts using plaintexts represented as vectors of real or complex values and where the ciphertext and plaintext spaces are essentially identical; (ii) BGN [16], which is the first scheme able to perform

addition and multiplication operations on ciphertexts while maintaining a constant size, thus allowing an unlimited number of additions and a single multiplication operation on ciphertexts of a predetermined length. An inherent property of homomorphic encryption is that with each homomorphic operation, errors pile up [17]. Consequently, beyond a specific number of multiplications or additions, ciphertexts become increasingly susceptible to decryption errors due to the accumulating error. To face this problem, Gentry [18] proposed a method named bootstrapping, that transforms a scheme that lacks full homomorphism (e.g., SHE) into a fully homomorphic one. FHE-based approaches, which are built on bootstrappable SHE, include (for a comprehensive review about FHE methods, please refer to [17]): (i) BGV [19], that consists of a scheme relying on Learning With Error (LWE) or its Ring-based variant (RLWE) named BFV [20], does not integrate the Gentry's bootstrapping technique. RLWE, as an algebraic extension of LWE, was introduced to provide more efficient and robust security for real-world applications; (ii) NTRU [21], instead, employs a lattice-based encryption scheme. LTV, a variant NTRU, integrates the bootstrapping and modulus switching techniques, and it uses a novel notion of HE scheme called MultiKey Homomorphic Encryption (MKHE) [22], that ultimately allows computation on ciphertexts encrypted under different keys. However, the specific application of HE to retina-based biometric authentication remains an unexplored area. This article aims to bridge this gap by presenting a novel framework that seamlessly integrates HE into retina-based biometric authentication, providing a comprehensive solution for privacy preservation and security enhancement in this context. The intricate and highly personalised nature of the retina biometric trait makes it an ideal candidate for the application of cancelable template without any helper data. As retinal scans provide a wealth of information about an individual's identity, ensuring the confidentiality and integrity of this data is a crucial aspect [23].

In summary, our research presents a robust and secure identity verification solution that, to our knowledge, represents the first-known integration of FHE with a retina-based BAS. The proposed privacy-preserving authentication framework, supported by a comprehensive discussion of both theoretical and practical aspects, utilises FHE to prevent user's sensitive data exposure throughout the authentication process. Additionally, we introduce a novel feature extraction technique tailored specifically for FHE, enhancing usability by overcoming the common computational hurdles associated with such encryption methods while achieving state-of-the-art accuracy. We rigorously tested the system using several parameter sets and performed comparative analyses to ensure optimal performance. The resilience of our framework has been thoroughly evaluated against multiple attacks, confirming its robustness and providing a high level of security compared to current systems. In a proactive stance against emerging quantum threats, we opted for a 192-bit security level over the conventional 128-bit, bolstering our system defences while meeting the requirements specified in the ISO/IEC 24745:2022 standard [24]. The paper is structured as follows: Section II reviews related work on retina-based biometric feature extraction and template protection. Section III covers preliminary concepts of homomorphic encryption schemes and basic notation used in the paper. Section IV describes our framework, exploring the fundamental principles underlying retina-based recognition and detailing the proposed algorithm tailored to this biometric data, highlighting its potential to address privacy and security concerns. Section V reports and discusses the experimental results of the developed system in terms of accuracy and robustness under normal (ideal) and abnormal (under attack) conditions. Finally, Section VI concludes the paper and draws final insights.

II. RELATED WORK

A. RETINA-BASED FEATURE EXTRACTION

The human retina is a complex and distinct biological structure located at the back of the eye, comprising intricate patterns of blood vessels and nerve fibers. These patterns, known as retinal vascular networks, are unique to each individual and remain stable throughout a person's lifetime, making the retina an ideal candidate for biometric identification [25]. An effective and trustworthy retina-based biometric recognition system is characterised by its accuracy. However, within such systems, the recognition rate is heavily influenced by the complexity of the vasculature in retinal images. Typically, a healthy retina exhibits a tidy arrangement of blood vessels. The isolation of these vessels through segmentation is relatively straightforward, as they exhibit no irregularities in their pattern and therefore have no impact on the recognition rate. Conversely, the vascular pattern in diseased retinal images becomes notably intricate due to the presence of pathological symptoms. Symptoms and signs related to certain disease or pathological condition, rely on visually observable patterns [26] that appear as gaps and clusters in the retinal vascular mapping. Hence, the retinal vasculature has been extensively explored as a distinctive feature for biometric recognition and various techniques have been developed to extract and analyse these patterns. However, the illumination in a retinal image is non-uniform due to the variation of the retina reaction or the non-uniformity of the imaging framework, which makes the task quite difficult. Certain approaches use properties of the vasculature to derive the features for person identification, whereas alternative methods rely on different characteristics of retinal images, such as the Optic Disc (OD) or image organisation properties. These varying types of features lead to a dichotomy in the feature extraction phase: one category involves the extraction of features based on vasculature, while the other focuses on non-vascular traits [27]. Among these approaches, Fatima et al. [28] employed a recursive supervised multilayered thresholding method to achieve accurate segmentation. Then, vascular

ending and bifurcation have been used as key features using Mahalanobis distance as the similarity measure for identification. Emary et al. [29] make use of an automated model based approach for vessel segmentation, introducing an Artificial Bee Colony (ABC) optimisation model along with Fuzzy C-Means clustering (FCM). In the work of Sadikoglu and Uzelaltinbulat [30], they used the feature vector extracted from the segmented image with a neural network trained by backpropagation. Wang et al. [31] devised a supervised hybrid method for segmenting blood vessels, which involved the use of a Convolutional Neural Network (CNN) classifier to extract hierarchical features. Subsequently, a Random Forests (RFs) classifier, combined with the 'winner-takes-all' technique, was employed to categorise pixels into vessel and non-vessel regions. Jiu et al. [32] applied the Gabor wavelet transform for blood vessels enhancement in the retinal images. They generated the feature vector by computing the distance and angle between a feature point and its four nearest neighbors. Authentication was then tested using the Euclidean distance measure. BahadarKhan et al. [33] introduced a technique aimed at extracting vascular skeletons using a morphological hessian based approach and region based Otsu thresholding that enhances the distinctiveness of retinal vascular patterns, especially for individuals with minor variations. Imani et al. [34], instead, presented an improved method on vessels detection using Morphological Component Analysis (MCA) developed based on sparse representation of signals, each of which is a linear combination of several morphologically distinct components.

B. TEMPLATE PROTECTION

In order to provide a secure and privacy-preserving authentication, a possible solution is to protect the biometric templates with a conventional encryption scheme such as the Rivest-Shamir-Adleman cryptosystem (commonly known as RSA). However, the similarity between samples before encryption cannot be preserved in the encrypted domain. At the same time, the template must not be decrypted for matching because this may otherwise lead to exposure of the template. As a result, template protection techniques are developed to preserve template secrecy while allowing a similarity assessment to be carried out concurrently [35]. To incorporate biometrics into cryptographic protocols, a frequently suggested approach involves substituting conventional matching algorithms with error-correction procedures and replacing traditional templates with secure schemas. Within these methodologies, biometric features are regarded as confidential and are employed, for example, to derive cryptographic keys. While this approach significantly streamlines the matching process, it does not comprehensively address all the security challenges posed by biometrics to date [36]. In particular, BioHashing involves the conversion of biometric features using a function determined by a user-specific key or password, with this transformation typically being reversible. While in the work of Teoh et al. [37] the system primarily focuses on facial recognition, comparable techniques can be extended to various other biometric traits like iris and fingerprint data. Various methods to protect the biometric template involve a non-invertible transformation function, even considering different biometric traits, such as fingerprint-based approaches in [38] and iris-based ones in [39]. The problem with such an approach is that it requires a trade-off between the transformation function discriminability and its non-invertibility. Nagar and Jain [40] provide an analysis of the measurement of non-invertibility in methods that make use of fingerprint data. In a key-binding biometric cryptosystem, the template security is ensured through the application of cryptographic algorithms. Typically, this system requires the computation of a transformation of the encrypted templates in the unencrypted domain, which often consumes a significant amount of time. Methods following this approach include the fuzzy vault [41] and the fuzzy commitment scheme [42]. In key-generating biometric cryptosystems, cryptographic keys are derived directly from biometric data [43], however, the problem lies in generating keys with sufficient stability and entropy. Additionally, there exist other studies focused on securely comparing data [44], [45], [46], [47]. Homomorphic encryption, a cryptographic technique that enables mathematical operations on encrypted data without exposing the underlying information, has been widely explored in the context of biometric authentication. The choice of homomorphic encryption technique in biometric authentication depends on the specific security and computational requirements of the system. Partial homomorphic encryption, somewhat homomorphic encryption, and fully homomorphic encryption each offer unique advantages and trade-offs in terms of security, efficiency, and functionality [19]. Early work by Gentry [18] laid the foundation for fully homomorphic encryption schemes, enabling secure operations on encrypted data. The potential of homomorphic encryption for protecting biometric data privacy was subsequently recognised by many researchers. The deployment of homomorphic encryption within biometric authentication systems has shown promise in addressing privacy and security concerns. For instance, Gilad-Bachrach et al. [48] introduced CryptoNets, a framework that applies neural networks to encrypted biometric data, allowing for high-throughput and accurate computations while preserving data privacy. Similarly, Juels and Ristenpart [49] proposed Honey Encryption, a method that enhances the security of biometric data by exceeding brute-force bounds. Moreover, the integration of homomorphic encryption with other cryptographic techniques, such as Secure Multi-party Computation (SMC), has been explored to enable secure collaborative authentication without revealing sensitive data to multiple parties simultaneously [50]. Abidin and Mitrokotsa [51] introduced an authentication system that involves a combination of Somewhat Homomorphic Encryption (SHE) and Ring-LWE (Learning With Errors on Rings). In this system, SHE is employed for template encryption, while Ring-LWE is harnessed for

the matching process. Barni et al. [52] used a Paillier cryptosystem that permits the addition of encrypted data. In their approach, the probe is encrypted, while the database remains unencrypted, resulting in a lack of security protection for the database. Kikuchi et al. [53] introduced additive homomorphic encryption and a cryptographic method for verifying that a committed value falls within a specified interval, all while maintaining the confidentiality of the actual value. The vector comparison is accomplished using cosine and Euclidean metrics, facilitated by a zero-knowledge proof of range mechanism.

III. HOMOMORPHIC ENCRYPTION

HE schemes are recognised for their ability to perform arithmetic operations on encrypted data, enabling computations such as homomorphic addition and/or multiplication directly on ciphertexts. These capabilities are indispensable for calculating similarity measures, such as the Hamming distance between encrypted vectors. Among various FHE schemes, our research utilises the levelled Brakerski/Fan-Vercauteren (BFV) scheme [20], [54], based on the RLWE hardness problem and renowned for its quantum-resistant security properties [55]. The rationale behind selecting the BFV scheme is underscored by its efficient noise management and ability to perform homomorphic operations without significant performance degradation. These aspects are relevant for systems requiring real-time processing capabilities, as in the case of authentication systems. This scheme is particularly advantageous as it eliminates the need for bootstrapping, a typically resource-intensive process, thereby ensuring exact results with enhanced computational efficiency. Additionally, it supports SIMD operations, enabling parallel processing within a single encrypted vector. In contrast, alternative schemes like the Gentry-Peikert-Vaikuntanathan (GPV) scheme [56], the Brakerski-Vaikuntanathan (BV) scheme [57], and the NTRU-based scheme [21] offer varying trade-offs between efficiency and security. While the GPV scheme boasts powerful bootstrapping capabilities, it can be computationally expensive. The BV scheme takes a different approach to noise management, making it less suited for real-time demands. The NTRU-based scheme, while secure via lattice problems, may not match the BFV scheme noise management efficiency.

A. BFV SCHEME

We will introduce the BFV scheme, a RLWE-based cryptosystem. Formally, a public key FHE scheme Π is characterised by a set of parameters that define the cryptographic system behaviour and a tuple of Probabilistic Polynomial-Time (PPT) algorithms $(\mathcal{K}, \mathcal{E}, \mathcal{D}, \mathcal{A}, \mathcal{M})$. Below is a high-level description based on BFV principles. The security parameter λ denotes the level of security FHE offers against various attacks and can be regarded as the computational cost required to breach the scheme, with a higher security parameter typically providing stronger encryption but at the cost of increased computational overhead. It is noteworthy that with 2^{λ} operations an attack would succeed with probability 1 (e.g., 128-bit or 256-bit security implies $\lambda = 128$ or $\lambda = 256$, respectively). The plaintext modulus p defines the size of the plaintext space and determines the precision of computations. Smaller values of p provide finer precision but a limited plaintext space, while larger values of p offer a broader range of values with reduced precision. The ciphertext modulus q specifies the size of the space in which encrypted values reside. A larger q can accommodate a larger noise budget, which allows for more homomorphic operations before the noise grows too large and decryption fails. However, this advantage comes with increased ciphertext size and greater computational overhead. Conversely, a smaller q leads to smaller ciphertexts and faster operations but restricts the number of allowable homomorphic operations due to a reduced noise budget. It also narrows the security margin, which can make the encryption scheme more susceptible to attacks. Polynomial degree n is a λ-dependent parameter, i.e., $n = n(\lambda)$, and it defines the degree of the polynomial ring $\mathcal{R}$ over which the FHE scheme operates. Higher values of n enhance security but may lead to increased computational overhead. Since this scheme involves representing the plaintext and ciphertext spaces over two different polynomial rings, where the plaintext ring includes encodings of unencrypted messages and the ciphertext ring includes encrypted messages, the choice of ring is crucial for the security and efficiency of the encryption process.

1) BASIC NOTATION

Let $n \in \mathbb{Z}$ be a power of two (ring dimension), then $\mathcal{R}_\alpha \triangleq \mathbb{Z}_\alpha[x]/(x^n + 1)$ denotes the ring of all polynomials with degree less than n and coefficients in $(-\alpha 2, \alpha 2] \cap \mathbb{Z}$ with $\alpha \in \mathbb{N}$, and let $\mathcal{X}$ represent the error distribution over the ring $\mathcal{R}_\alpha$. Moreover, consider two integer moduli $p, q \in \mathbb{Z}$ such that $\{p > 1 \wedge p \nmid 2\}$ and $\{q > 1 \wedge q \nmid 2\}$ with $p \ll q$, then

$$\mathcal{R}_p = \mathbb{Z}_p[x]/(x^n + 1) \tag{1}$$
$$\mathcal{R}_q = \mathbb{Z}_q[x]/(x^n + 1) \tag{2}$$

denote the plaintext and ciphertext polynomial rings, respectively. Finally, let $q \leftarrow \mathcal{Q}$ be an element randomly assigned from the distribution $\mathcal{Q}$ and $\Psi = \lfloor qp \rfloor$.

a: THE KEY-GENERATION ALGORITHM $\mathcal{K}$

For a given security parameter λ along with the degree n of the cyclotomic polynomial, it provides in output a set of keys:

- the private key is typically used for decryption, should be kept private by the user, and is calculated as $\kappa_S \leftarrow \mathcal{R}_2$,
- the public key is used for encryption and is calculated as $\kappa_P = \left(\left[-\left(a \cdot \kappa_S + e\right)\right] \bmod q, a\right) \in \mathcal{R}_q \times \mathcal{R}_q$, where $a \leftarrow \mathcal{R}_q$ and $e \leftarrow \mathcal{X}$,
- the evaluation key supports efficient homomorphic computations by enabling relinearisation of ciphertexts to manage noise and size after homomorphic

multiplication. Given the decomposition base b used to express a polynomial in $\mathcal{R}_q$ in terms of d 1 polynomials in base $b \in \mathbb{Z}$, where $d = \lfloor \log_b q \rfloor$, the evaluation key κ_E is a set of $(d \ 1)$ pairs of polynomials generated for $0 \leq i \leq d$ as $\kappa_E(i) = \left(\left[b^i \kappa_S^2 - (a_i \cdot \kappa_S \ e_i)\right] \bmod q, a_i\right)$, where $a_i \leftarrow \mathcal{R}_q$ and $e_i \leftarrow \mathcal{X}$.

b: THE PUBLIC ENCRYPTION ALGORITHM $\mathcal{E}$

It uses a public key $\kappa_P = (\kappa_0, \kappa_1)$ and a message $m \in \mathcal{R}_p$ to generate a ciphertext $\Gamma = (\gamma_0, \gamma_1) \in \mathcal{R}_q \times \mathcal{R}_q$ as follows:

$$\Gamma = \left(\left[\kappa_0 \cdot u \ e_0 \ \Psi \cdot m\right] \bmod q, \left[\kappa_1 \cdot u \ e_1\right] \bmod q\right) \quad (3)$$

where $u \leftarrow \mathcal{R}_2$ and $e_0, e_1 \leftarrow \mathcal{X}$. The first element, γ_0, encompasses the concealed plaintext, while the second element, γ_1, carries supplementary data necessary for decryption. based on the secret key

c: THE DECRYPTION ALGORITHM $\mathcal{D}$

It relies on both the secret key κ_S and the auxiliary information contained in the ciphertext $\Gamma = (\gamma_0, \gamma_1)$ as follows:

$$m = \left[\left\lfloor \frac{p}{q} \cdot \left[\gamma_0 \ \gamma_1 \cdot \kappa_S\right] \bmod q \right\rceil\right] \bmod p \in \mathcal{R}_P \quad (4)$$

d: HOMOMORPHIC ADDITION AND MULTIPLICATION

The PPT algorithms $\mathcal{A}$ and $\mathcal{M}$ provide support for homomorphic addition and homomorphic multiplication over two ciphertexts $\Gamma_0 = (\gamma_0^0, \gamma_0^1)$ and $\Gamma_1 = (\gamma_1^0, \gamma_1^1)$ by making use of the following properties:

- homomorphic addition ($\mathcal{A}$)

$$\mathcal{D}\left(\mathcal{E}\left(m_0\right) \boxplus \mathcal{E}\left(m_1\right)\right) = m_0 \ m_1, \forall m_0, m_1 \in \mathcal{R}_P \quad (5)$$

- homomorphic multiplication ($\mathcal{M}$)

$$\mathcal{D}\left(\mathcal{E}\left(m_0\right) \boxtimes \mathcal{E}\left(m_1\right)\right) = m_0 \times m_1, \forall m_0, m_1 \in \mathcal{R}_P \quad (6)$$

IV. PROPOSED FRAMEWORK

In the following, we present a holistic solution geared towards both privacy preservation and security enhancement by seamlessly integrating homomorphic encryption into retina-based biometric authentication. The proposed framework is suitable to be used in systems operating in both authentication and identification modes. For the sake of clarity and without loss of generality, we will outline the implementation of this method specifically for the authentication process, which is summarised in Algorithm 1. The biometric sample (query) is initially captured and processed to extract the relevant features, which are then encoded into a polynomial and encrypted using HE to generate the corresponding template Γ^Q, as illustrated on the left side of Figure 2. Simultaneously, the Authentication Server (AS) generates a one-time nonce h using a Pseudo-Random Number Generator (PRNG), which is also encoded into a polynomial and encrypted using the client's public key κ_P^C, resulting in the encrypted nonce $\hbar$. To ensure secure transmission to the client, this data is further protected by encrypting it with a Pre-Shared Key (PSK)

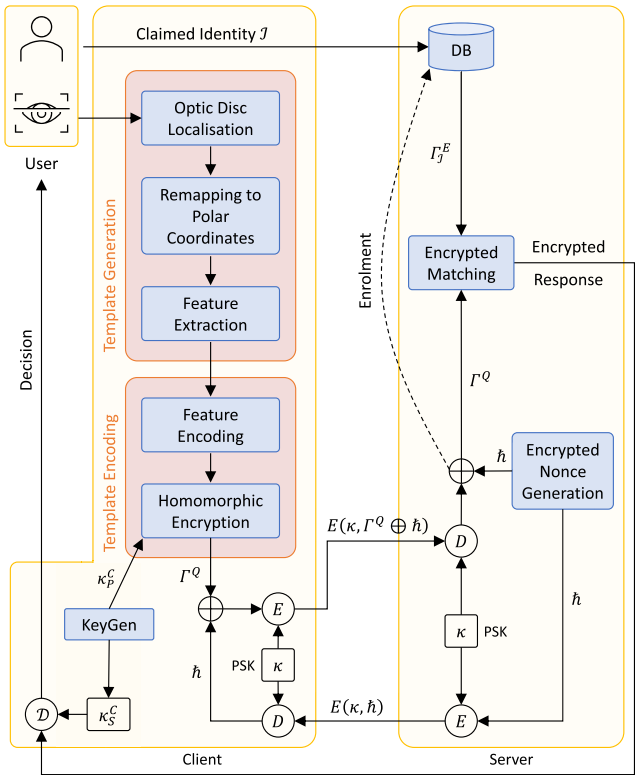


FIGURE 2. High-level view of the proposed system operating in authentication mode.

κ using the AES-128 symmetric-key encryption algorithm, chosen for its robustness against attacks [58]. The encrypted nonce $\hbar$ is then sent to the client who decrypts it and combines it with the encrypted template Γ^Q using a logical XOR operation. The client then re-encrypts the resulting data with κ and sends it back to the AS. Upon receiving the encrypted packet, the AS decrypts it with κ, applies a logical XOR operation with the nonce to recover the client's template Γ^Q, and completes the authentication process by comparing it with the stored encrypted template $\Gamma_{\mathcal{J}}^E$ corresponding to the claimed identity $\mathcal{J}$. This sequence of encryption and logical XOR operations ensures the security of the data transmission, preventing any duplication or replay of requests to the AS [59].

In this context, the XOR operation is emulated by applying homomorphic operations to encrypted data. Since the homomorphic scheme used supports only addition and multiplication, bitwise XOR is achieved on individual polynomial coefficients within this framework, as detailed in Section IV-B. This approach ensures that XOR behaviour for binary inputs is accurately replicated within the algebraic structure of the homomorphic encryption scheme.

When presented with a claimed identity $\mathcal{J}$ and a query Γ^Q, the BAS is tasked with classifying the pair $(\mathcal{J}, \Gamma^Q)$ as either 'genuine' or 'impostor'. In this regard, let $\Gamma_{\mathcal{J}}^E$ represent the encrypted template stored in the AS and associated with the identity $\mathcal{J}$. The matching score is then determined by the

Algorithm 1 Authentication Procedure

Input: Biometric sample Q and claimed identity $\mathcal{J}$
Output: Boolean authentication result

1: The client device elaborates the acquired biometric sample (query) $Q \subset \mathbb{N}^{N\times M}$ to generate $\Omega \in \{0,1\}^{n\times m}$, a boolean image containing the extracted features.
2: The client device further reduces the size of Ω to generate the RetinaCode ω, which is more suitable for HE.
3: The client device encodes the vector ω into the plaintext space polynomial ring $\mathcal{R}_p = \mathbb{Z}_p[x](x^n \ 1)$, obtaining Φ, and encrypts it using the public encryption key κ_P^C, resulting in the encrypted template $\Gamma^Q = \mathcal{E}(\kappa_P^C, \Phi)$.
4: The AS generates a one-time nonce h that is first encoded and then encrypted using the client's public key κ_P^C, resulting in the encrypted nonce $\hbar$.
5: The AS further protects the nonce $\hbar$ by encrypting it with a PSK κ using the AES-128 encryption algorithm to ensure secure transmission to the client device: $E(\kappa, \hbar)$.
6: The encrypted nonce $\hbar$ is sent to the client device, which decrypts it and combines it with the encrypted template Γ^Q using a logical XOR operation: $\Gamma^Q \oplus \hbar$.
7: The client device re-encrypts the result of the XOR operation using the PSK κ and sends the encrypted packet back to the AS, along with the claimed identity $\mathcal{J}$.
8: The AS retrieves the enrolled encrypted template $\Gamma_{\mathcal{J}}^E$ from the database (DB) for the claimed identity $\mathcal{J}$.
9: Upon decrypting the received packet using κ, the AS performs a logical XOR operation with $\hbar$ to recover the original (encrypted) template: $\left(\left(\Gamma^Q \oplus \hbar\right) \oplus \hbar\right) = \Gamma^Q$.
10: The AS computes the similarity score s in the encrypted domain between $\Gamma_{\mathcal{J}}^E$ and Γ^Q.
11: The response from the AS, which is still in the encrypted domain, is sent back to the client. The client decrypts the response using its private key κ_S^C to obtain the final result.

similarity measure between Γ^Q and $\Gamma_{\mathcal{J}}^E$:

$$\left(\mathcal{J}, \Gamma^Q\right) \in \begin{cases} \text{genuine}, & \text{if } s\left(\Gamma^Q, \Gamma_{\mathcal{J}}^E\right) \geq \xi \\ \text{impostor} & \text{otherwise} \end{cases} \tag{7}$$

where s represents a similarity function in the encrypted domain and ξ denotes a specific threshold that defines the operational criteria of the system. Typically, this authentication mode is employed to achieve positive recognition, aiming to deter multiple individuals from using the same identity [60]. Therefore, given the two encrypted templates, $\Gamma^Q, \Gamma_{\mathcal{J}}^E \in \mathcal{R}_q \times \mathcal{R}_q$, our scheme employs the Hamming Distance (HD) as the similarity measure, as outlined below. After computing this measure, the authentication server sends the result to the client. The client then decrypts the data using its unique private key κ_S^C (where 'C' denotes client-specific), to determine the final authentication outcome.

A. RETINA TEMPLATE GENERATION

The retina, located at the back of the eye, contains a complex network of blood vessels that form unique and stable patterns in each individual, moreover, these patterns remain relatively unchanged throughout a person's life. The process of retinal recognition begins with image acquisition. Specialised devices for retinal imaging, such as fundus cameras, are used to capture high-resolution images of the retina. These images are typically in colour and may include various retinal structures, including the optic disc, blood vessels, and the macula. Before vessel features can be extracted, retinal images must undergo pre-processing to enhance the quality of the images. Figure 3 illustrates the main steps involved in the processing of the raw images to generate the template, which include: (i) optic disc localisation through the Circle Hough Transform (CHT), (ii) remapping of the image from Cartesian coordinates to the polar representation, (iii) blood vessel extraction by using Laplacian of Gaussian and morphological operations, (iv) generation of the RetinaCode characterised by a reduced overall template size.

1) OPTIC DISC LOCALISATION

The optic disc is the point where the optic nerve enters the retina, and it is characterised by the absence of light-sensitive photoreceptor cells, which creates a natural scotoma. Each individual's optic disc has unique characteristics, such as the pattern of blood vessels entering and exiting the disc. The central point of the optic disc serves as a reference point for aligning and normalising retinal images [61]. The procedure starts with a pre-processing of the raw image, which consists of the selection and subsequent extraction of the green channel from the raw RGB image, as it provides the maximum contrast between exudates and the neighboring regions [62]. Hence, contrast and luminosity normalisation is applied to the extracted image I through the following pixel-wise procedure [23]:

$$I_N(i,j) = \begin{cases} \mu_n \ \rho & \text{if} \quad I(i,j) > \sigma^2 \\ \mu_n - \rho & \text{if} \quad I(i,j) \leq \sigma^2 \end{cases} \tag{8}$$

where $I(i,j)$ represents the grey-level value of the pixel (i,j), μ and σ represent the estimated mean and variance of the image I, respectively, while the term ρ is equal to

$$\rho = \sqrt{\frac{\sigma_n^2 \left(I(i,j) - \mu\right)^2}{\sigma^2}} \tag{9}$$

with $\mu_n = 128$ and $\sigma_n = 16$ representing the desired values of mean and variance, determined experimentally. After these preliminary steps, OD appears as the brightest region in the image. To eliminate the vascular structures from the pre-processed image, a morphological closing operation

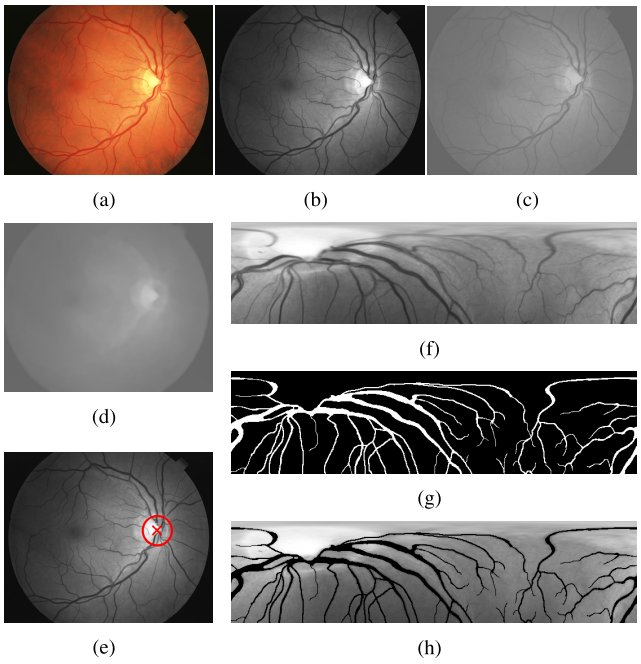


FIGURE 3. Results of the proposed pre-processing and feature extraction algorithms used to extract the binary vessel map of the retina: (a) raw retinal fundus image, (b) green channel, (c) normalised image, (d) resulting image after the application of morpohological operations, (e) optic disc localisation through the Circle Hough Transform, where the red circle and cross graphically represent the location and central point of the optic disc in the processed image, (f) remapping of the image in polar representation using the optic disc as a reference point, (g) the resulting boolean retinal vascular pattern, and (h) remapped fundus image overlapped with the extracted vascular patterns.

is performed using a disc-shaped structuring element

$$\mathcal{S} \subset \mathbb{Z}_2^{9\times 9} : \mathcal{S}_i \in \{0, 1\}\, \forall i \tag{10}$$

The Hough Transform (HT) is a widely employed method for determining the parameters of geometric shapes like lines and circles within an image. This technique operates by transforming edge points into an accumulator space and identifies peaks in that space, which correspond to potential circles. The circle parameters are then estimated from these peaks, and post-processing steps are performed to refine the results and locate the OD accurately. To detect the centre coordinates and radius of the OD region, HT considers the circle equation:

$$\left(x - x_c\right)^2 \ \left(y - y_c\right)^2 = r^2 \tag{11}$$

where the pair of values $\left(x_c, y_c\right)$ identifies the centre coordinates of the circle and r its radius. The first step aims at detecting edges in the retinal image, emphasising regions of rapid intensity change. Then, an accumulator array is set up to represent potential circles, with each element of the array corresponding to a possible circle with centre $\left(x_c, y_c\right)$ and radius r. For each edge pixel in the image, the CHT votes for potential circle parameters in the accumulator array by considering all possible circles passing through that pixel. Peaks in the accumulator array indicate potential circles, with their positions providing the centre coordinates and radius of the detected circles. Finally, to refine the results, a threshold is applied to remove weak circle candidates, and non-maximum suppression may be used to eliminate redundant or overlapping circles [63].

2) REMAPPING TO POLAR REPRESENTATION

The subject eye and head movements during the retinal scanning process may introduce slight rotations and/or translations into the acquired fundus images. Therefore, to mitigate potential verification errors stemming from image translation and/or rotation it is crucial to incorporate translation- and rotation-invariant features. To solve this problem, we use a transformation that remaps the image from Cartesian coordinates to the polar representation in order to obtain a rotation invariant image. The output consists of an image $X \in \mathbb{R}^{M\times N}$ with M points along the r axis and N points along the θ axis, using as origin of the image the centre of the given image, as shown in Figure 4. The remapping of the Region of Interest (ROI) from Cartesian coordinates $\left(x, y\right)$ to the polar coordinates $\left(r, \theta\right)$ has been modelled making use of a variant of the Daugman rubber-sheet model [64]. The traditional model, tailored for iris recognition, takes into account the iris annular region and its concentric circular structures, which are then remapped into a rectangular data block. In contrast, the proposed modified version of the model for retina images adapts the process to accommodate the different structural characteristics of the retina by mapping the entire region of interest from Cartesian to polar coordinates, thus ensuring a comprehensive representation of the retinal vasculature. This adaptation is crucial because the retina contains a branching pattern of blood vessels rather than the concentric patterns found in the iris. This procedure allows to standardise the retina data, ensuring uniformity and accuracy in recognition. The process of converting the coordinate system of a point from the Cartesian $\left(x, y\right)$ format to the polar $\left(r, \theta\right)$ format is as follows:

$$X\left(x\left(r, \theta\right), y\left(r, \theta\right)\right) \rightarrow X\left(r, \theta\right) \tag{12}$$

where

$$\begin{cases} x\left(r, \theta\right) = r\cos\left(\theta\right) \\ y\left(r, \theta\right) = r\sin\left(\theta\right) \end{cases} \tag{13}$$

The model maps each pixel within the region of interest to a corresponding position on the polar axes $\left(r, \theta\right)$, with $r \in \left[0, 1\right]$ representing the normalised radial distance and $\theta \in \left[0, 2\pi\right]$ denoting the angular rotation at that specific radius. Bilinear interpolation is used to refine pixel values at non-explicit positions within the image grid. This method enhances image quality by smoothing pixel transitions, using linear interpolation among four adjacent known points without adding new data. The resulting image after the application of the procedure is shown in Figure 3f.

3) FEATURE EXTRACTION

The suggested feature extraction process encompasses the following steps:

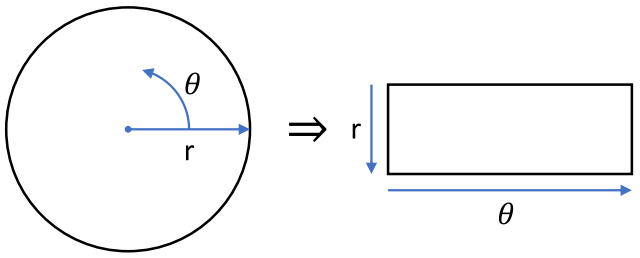


FIGURE 4. Remapping of retinal region from Cartesian coordinates to the polar representation.

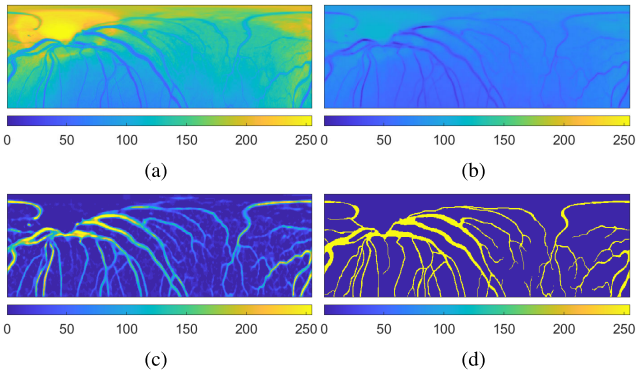


FIGURE 5. Main steps involved in the feature extraction: (a) ROI image, (b) blood vessel enhancement through contrast and sharpness adjustment, (c) highlighting line-like features and fine details using the LoG, and (d) final result after morphological operations.

1) emphasising blood vessels through contrast and sharpness enhancement,
2) employing Laplacian of Gaussian,
3) using morphological operations.

Figure 5 illustrates the results of the above-mentioned steps whose details are given below.

Blood vessel enhancement: it enhances the visibility of blood vessels, making them stand out from the surrounding background. To address uneven illumination and improve contrast, we employed a combination of the Top-Hat and Bottom-Hat transforms, designed to identify bright (or dark) objects against a backdrop with varying brightness.
The Top-Hat transform is defined as the difference between the input image X and its morphological opening by means of a cross-shaped structuring element $\mathcal{S} \in \{0,1\}^{3\times3}$:

$$\hat{X}_T = X - \left(X \circ \mathcal{S}\right) = X - \left(\left(X \ominus \mathcal{S}\right) \oplus \mathcal{S}\right) \tag{14}$$

while the Bottom-Hat transform, is defined as the difference between the morphological closing operation applied to the input image X using the structuring element $\mathcal{S}$ and the original input image:

$$\hat{X}_B = \left(X \bullet \mathcal{S}\right) - X = \left(\left(X \oplus \mathcal{S}\right) \ominus \mathcal{S}\right) - X \tag{15}$$

where the opening is achieved through the erosion of the image X with $\mathcal{S}$, followed by the dilation of the resulting image using $\mathcal{S}$. Conversely, the closing is accomplished by first dilating X with $\mathcal{S}$ and then eroding the resulting image using $\mathcal{S}$. Subsequently, to eliminate the bright objects and enhance the dark ones that correspond to blood vessels, we employ the Top-Hat and Bottom-Hat transforms as follows:

$$\hat{X} = X - \left(\hat{X}_T \;\; \hat{X}_B\right) \tag{16}$$

After that, a normalisation is required in order to standardise the mean and variance through the Equations (8) and (9), obtaining a new image $\hat{X}_N$.

Laplacian of Gaussian: it consists of a two-dimensional isotropic measure of the second spatial derivative of an image and it is frequently used to highlight line-like features and fine details. To enhance its robustness against noise, the operator is applied to an image that has been previously smoothed using a two-dimensional Gaussian operator with variance υ^2:

$$G\left(i,j\right) = \exp\left(-\frac{i^2\;j^2}{2\upsilon^2}\right) \tag{17}$$

Since convolution and differentiation are the only linear operators involved, they can be interchanged as follows:

$$\hat{X}_L(i,j) = \nabla^2\left[G * \hat{X}_N(i,j)\right] = \left[\nabla^2 G\right] * \hat{X}_N(i,j) \tag{18}$$

Hence, the Laplacian of Gaussian can be precomputed in advance so only one convolution needs to be performed at run-time on the image:

$$\nabla^2 G\left(i,j\right) = \left(\frac{i^2\;j^2 - 2\upsilon^2}{\upsilon^4}\right) G\left(i,j\right) \tag{19}$$

Morphological operations: the purpose is to refine the vascular pattern image $\hat{X}_L$ by removing small artifacts and noise, such as sporadic bright spots on a dark background and black holes within bright structures. To achieve this objective, we employ a morphological filter consisting of an opening operation, followed by a dilation operation using the same previously defined structuring element $\mathcal{S}$:

$$\Omega = \left(\hat{X}_L \circ \mathcal{S}\right) \bullet \mathcal{S} \tag{20}$$

Ultimately, an iterative thinning transformation is applied to reduce the foreground object to a minimal connected stroke while preserving its topology. This ensures that the final vascular pattern image maintains its homotopic equivalence to the original input image [23].

B. ENCRYPTED RETINA TEMPLATE MATCHING

For a fully homomorphic encryption scheme, the computational complexity is the most important technical issue to implement HE-based retina matching, and this becomes especially critical in situations where there is a need for multiple modular arithmetic operations on encrypted data. Therefore, we propose a feature encoding procedure that facilitates easier and faster arithmetic operations on encrypted data without compromising security.

1) OPTIMISATION

The size of the resulting image Ω from the feature extraction step is 120×480, leading to a 57600-bit binary vector after matrix vectorisation $\omega = \text{vec}\left(\Omega^\top\right)$. Encrypting such a large vector directly within a single ciphertext would require a polynomial modulus $n \geq 57600$, which can cause significant computational overhead and longer ciphertext computation times, affecting system efficiency.

A simple method considered for addressing this challenge consists of segmenting the binary vector ω into smaller blocks of size d [65], with the compressed size l given by:

$$l = \left\lfloor \frac{\text{Total number of bits}}{d} \right\rfloor \tag{21}$$

While compression can reduce storage and computational needs, it complicates bitwise operations (e.g., XOR) on encrypted data, potentially compromising accuracy and efficiency. To address these issues, we implemented a more efficient approach using multi-ciphertext encryption. Instead of compressing the binary vector ω, it is encrypted across multiple ciphertexts, allowing for smaller polynomial moduli and enabling direct bitwise operations without the need for decompression.

a: MULTI-CIPHERTEXT ENCRYPTION METHOD

Segmenting the large binary vector ω into smaller parts, each treated as an independent polynomial and encrypted into separate ciphertexts, preserves encryption efficiency without requiring an excessive increase in the polynomial modulus, thus avoiding negative impacts on system performance. Additionally, distributing the encrypted data across multiple ciphertexts enables parallel processing during homomorphic operations, significantly enhancing computational efficiency while maintaining encryption integrity.

It is important to observe that as each homomorphic operation is executed, the noise within the ciphertext increases. Indeed, if the cumulative noise exceeds the noise budget, decrypting the ciphertext becomes unfeasible. Therefore, while improving system efficiency and enhancing security is of primary concern, it is equally essential to ensure the successful decryption of the ciphertext. The use of multi-ciphertext encryption effectively limits noise growth within each ciphertext, thereby enabling the system to control noise accumulation even during complex operations. By handling smaller subsets of the overall data, each ciphertext reduces per-ciphertext noise growth while still enabling efficient direct homomorphic bitwise operations. Without multi-ciphertext encryption, a larger polynomial modulus n would be required. As described in Section IV-B:

1) the polynomial modulus n, coefficient modulus q, and noise distribution $\mathcal{X}$ maintain a direct proportionality; thus, a larger value of n leads to an increased value of q and consequently higher noise levels;
2) the security level is inversely proportional to the coefficient modulus q; hence a larger coefficient modulus q results in a weaker security level.

Table 1 presents a comparative analysis of key system parameters when using multi-ciphertext encryption across different configurations. Here, the parameters n, ct, $\log_2(q)$, and λ represent the polynomial modulus, number of ciphertexts, bit-length of the coefficient modulus, and the security parameter, respectively. These configurations serve to highlight how multi-ciphertext encryption successfully optimises system performance while simultaneously ensuring that high levels of security are maintained throughout the process.

TABLE 1. Comparative analysis of key system parameters for multi-ciphertext encryption of the 57,600-bit binary vector across different configurations.

Parameters		$\log_2(q)$ based on λ		
n	ct	128-bit	192-bit	256-bit
2048	29	54	37	29
4096	15	109	75	58
8192	8	218	152	118
16384	4	438	300	237

2) FEATURE ENCODING

The encryption process transforms original plaintext elements from the plaintext ring $\mathcal{R}_p$ into ciphertext elements within the ciphertext ring $\mathcal{R}_q$, over which it is possible to perform arithmetic operations. The basic idea behind this encoding technique relies on splitting the plaintext space into residue classes using different prime numbers, and then perform computations on batches of ciphertexts corresponding to these residue classes [66]. The batching technique is commonly used as a method to optimise and speed up computations by operating on batches of encrypted data instead of using single elements. With the aid of the Chinese Remainder Theorem (CRT) we are able to pack a number n of integers modulo p into a plaintext polynomial reducing the computational time required for the multiplication operations, which can be performed on the polynomial likewise on integers. In this regard, the plaintext modulus p should be chosen as a prime number such that

$$p \equiv 1 \pmod{2n} \tag{22}$$

then, there exists a ring isomorphism [67]:

$$\mathcal{R}_p \approx \prod_{i=0}^{n-1} \mathbb{Z}_p \tag{23}$$

As a result, this CRT-based batching technique, which relies on the SIMD (Single Instruction, Multiple Data) paradigm, enables the optimisation of the encryption process ensuring more efficient handling of multiple data in parallel, allowing to perform n coefficient-wise multiplication operations at the cost of a single one in $\mathcal{R}_p$.

3) TEMPLATE PROTECTION

Before applying the process of homomorphic encryption, it is necessary to encode the vector to be encrypted — specifically, RetinaCode ω — into a polynomial representation $\Phi \in \mathcal{R}_P$. Each element in this representation corresponds to a coefficient. Using the public key denoted as κ_P^C, we proceed to encrypt the encoded polynomial Φ as follows:

$$\Gamma = \mathcal{E}\left(\kappa_P^C, \Phi\right) \tag{24}$$

Remarkably, homomorphic operations (both addition and multiplication) can be directly performed on the ciphertexts without prior decryption. For instance, when adding or multiplying two encrypted vectors, we apply the corresponding operations element-wise to their ciphertexts. We recall that the plaintext and ciphertext spaces are defined over two distinct polynomial rings denoted by $\mathcal{P} = \mathcal{R}_p$ and $\mathcal{C} = \mathcal{R}_q \times \mathcal{R}_q$, as described in Section III-A, where p and q are the plaintext and ciphertext coefficients, respectively. Note that in practice, q is usually much greater than p, consequently, the cardinality of the ciphertext space $\mathcal{C}$ is significantly larger than that of the plaintext space $\mathcal{P}$. This implies that a plaintext vector ω can be mapped to multiple valid ciphertexts in $\mathcal{C}$. In other words, due to the larger size of $\mathcal{C}$, there are numerous ways to encrypt the same plaintext ω, resulting in different ciphertexts. This property is essential in cryptographic systems, as it enhances security by increasing the complexity of decryption attacks.

Besides the parameters for plaintext and ciphertext spaces (p, q, n) introduced in Section III-A, the selected scheme involves the use of several random distributions, two of which are used during encryption and are defined as follows: (i) $\mathcal{R}_2$, which is the key distribution used to sample polynomials with integer coefficients in $\{-1, 0, 1\}$; (ii) the error distribution $\mathcal{X}$, that is defined as a discrete Gaussian distribution with a mean μ and variance σ^2 operating over the ring $\mathcal{R}$ bounded by some integer β and, according to the current version of the homomorphic encryption standard [68].

These parameters are specifically set as

$$\mu = 0 \tag{25}$$

$$\sigma = 8 \Big/ \sqrt{2\pi} \approx 3.2 \tag{26}$$

$$\beta = \lfloor 6\sigma \rceil = 19 \tag{27}$$

Therefore, to encrypt the vector ω, we first generate three small random polynomials $u \leftarrow \mathcal{R}_2$ and $e_0, e_1 \leftarrow \mathcal{X}$. Subsequently, these polynomials are used to produce a ciphertext $\Gamma = (\gamma_0, \gamma_1) \in \mathcal{R}_q \times \mathcal{R}_q$, which represents the homomorphic image of $\Phi \in \mathcal{R}_P$. After the encrypted data undergoes homomorphic computations on the remote server, the encrypted results are transmitted back to the client. Upon receipt, the client performs decryption by evaluating the ciphertext Γ on its own secret key κ_S^C and reversing the scaling factor Ψ applied during encryption.

4) ENCRYPTED DISTANCE MEASURE

The encrypted matching module of the system makes use of the Hamming distance (HD) calculation to compare binary templates within the encrypted domain. The Hamming distance, which measures the number of differing bits between two binary vectors, is mapped into the arithmetic domain to enable secure computations, thus facilitating the use of homomorphic encryption operations.

The Hamming distance $HD(A, B)$ between two binary vectors $A = \{a_0, a_1, \ldots, a_{n-1}\}$ and $B = \{b_0, b_1, \ldots, b_{n-1}\}$, both of length n, is defined as the number of positions where $a_i \neq bi$, for $i \in [0, n-1]$, through the following equation:

$$HD(A, B) = \sum_{i=0}^{n-1} (a_i \oplus b_i) \tag{28}$$

In the encrypted domain, direct bitwise XOR operations are not feasible due to the arithmetic nature of homomorphic encryption. To address this, the XOR is expressed in an arithmetic form as $a_i \oplus b_i = (a_i - b_i)^2$, allowing it to be computed homomorphically. Thus, the Hamming distance is reformulated as the sum of these squared differences between the binary vector elements:

$$HD(A, B) = \sum_{i=0}^{n-1} (a_i - b_i)^2 \tag{29}$$

Thus, considering Φ^Q and Φ^T as the polynomial encodings of the binary vectors ω^Q and ω^T, the Hamming distance between the two ciphertexts, Γ^Q and Γ^T, which represent the query and stored templates respectively, is computed homomorphically. The polynomials Φ^Q and Φ^T are expressed as sums over the ring $\mathcal{R}_p$ as follows:

$$\Phi^Q = \sum_{i=0}^{n-1} \Phi_i^Q x^i \quad \text{and} \quad \Phi^T = \sum_{i=0}^{n-1} \Phi_i^T x^i \tag{30}$$

where $\Phi_i^Q, \Phi_i^T \in \{0, 1\}$ are the binary coefficients of the polynomials Φ^Q and Φ^T. The XOR operation between Φ^Q and Φ^T is computed directly over the polynomial encodings in the ring $\mathcal{R}_p$ using the expression:

$$\Phi^Q \oplus \Phi^T = \left(\Phi^Q - \Phi^T\right)^2 = \sum_{i=0}^{n-1} \left(\left(\Phi_i^Q - \Phi_i^T\right)^2\right) x^i \tag{31}$$

This operation yields another polynomial in $\mathcal{R}_p$, representing the XOR between the two ring elements. Each coefficient of the resulting polynomial corresponds to the XOR between the coefficients of Φ^Q and Φ^T. The same XOR operation can be performed homomorphically after encryption as follows:

$$\begin{aligned} \mathcal{E}\left(\Phi^Q \oplus \Phi^T, \kappa_P\right) &= \mathcal{E}\left(\left(\Phi^Q - \Phi^T\right)^2, \kappa_P\right) \\ &= \left(\mathcal{E}(\Phi^Q, \kappa_P) \boxminus \mathcal{E}(\Phi^T, \kappa_P)\right) \\ &\boxtimes \left(\mathcal{E}(\Phi^Q, \kappa_P) \boxminus \mathcal{E}(\Phi^T, \kappa_P)\right) \end{aligned} \tag{32}$$

as outlined in Section III-A through the homomorphic subtraction and multiplication operations. To compute the final Hamming distance, the sum of all coefficients of the resulting encrypted polynomial is accumulated through cyclic rotations and additions into a single position. By leveraging the properties of automorphisms, this summation is efficiently

performed with $\log_2(n)$ shifts and $\log_2(n)$ additions. For example, in Figure 6, given a vector $V = (0, 1, 2, 3)$ with homomorphic ciphertext $\mathcal{E}(V, \kappa_P) = (v_0, v_1, v_2, v_3)$ and a ciphertext slot size of 4, only two shifts and additions are required to compute the sum, with shift steps $(\delta_1, \delta_2) = (2^0, 2^1)$.

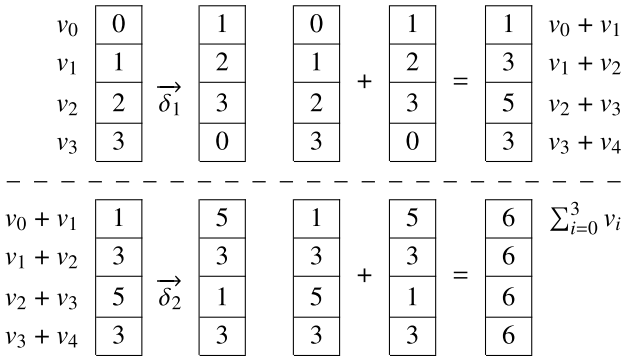


FIGURE 6. Sum of the elements in the homomorphic ciphertext slot.

C. ASSUMPTIONS AND LIMITATIONS

To assess the security of the system, we considered three possible vulnerabilities that could impact the data privacy:

1) the client side,
2) the server side,
3) the communication channel between the client and the server.

1) ASSUMPTIONS

As illustrated in Figure 2, the client side is responsible for extracting features from the retinal image and ensuring the confidentiality of the secret key. As a result, since it is crucial to ensure the security of the client device, we assume the following: (i) the client device is regarded as completely trustworthy during both the enrolment and authentication stages, serving as a robust foundation for the system integrity; (ii) the user's secret key is meticulously preserved within the local storage of the client device, maintaining its confidentiality throughout the entire enrolment or authentication process; (iii) the server employs a 'Honest-But-Curious' (HBC) security model, signifying that it honestly performs the required computations, despite a potential underlying interesting the processed data. This model assumes the server compliance with protocol, yet acknowledges the possibility of it harbouring an inclination to analyse the data for information beyond the scope of its designated role; (iv) the use of a PSK to encrypt the nonce prevents replay attacks by ensuring that each session is uniquely identifiable and secure, thereby protecting the system from the reuse of captured communications.

2) LIMITATIONS

Implementing homomorphic encryption in biometric systems introduces significant constraints. The computational overhead associated with HE necessitates intensive calculations, significantly increasing data processing time. Consequently, it becomes impractical for systems requiring real-time responses. While HE is suitable for authentication, where comparisons involve a single template, it is less appropriate for identification, which entails searching across a database of multiple entries. The integration complexity of HE into existing systems, the larger data sizes due to encryption, and the trade-off between security and system efficiency further challenge its adoption. As an emerging technology, HE for biometrics is still maturing, and future advancements may mitigate these challenges. However, for now, they remain significant hurdles to its widespread implementation [69].

V. EXPERIMENTAL APPROACH AND EVALUATION

Extensive experiments have been carried out to assess the performance and the effectiveness of the proposed solution. In this section, we provide a description of the dataset used in the experiments and the experimental setup used to evaluate the system performance considering the following criteria, as stated in the ISO/IEC 24745:2022 standard [24]: (i) irreversibility and revocability, (ii) unlinkability, (iii) renewability, and (iv) performance preservation.

A. DATABASE DESCRIPTION

Since the evaluation of an authentication system requires a diverse set of samples to ensure the system effectiveness across different individuals and conditions, the proposed authentication framework has been tested upon the Retina Identification DataBase (RIDB) [70], a globally shared retinal database available for research purposes. RIDB database includes 100 fundus images acquired from 20 subjects (five samples per individual) with a TOPCON-TRC 50 EX fundus camera having a spatial resolution of 1504×1000 with 24 bits per pixel. Prior to the image acquisition, the subject's pupils were dilated to a diameter of 4.0 mm to ensure an aperture wide enough for capturing clear images, and the images were taken with a 45-degree Field of View (FOV). Consequently, the images were saved in JPEG format, a commonly used digital image format that allows for efficient compression and storage of high-quality images. Figure 7 illustrates some image samples from the RIDB database, while Table 2 reports the database details.

B. EXPERIMENTAL SETUP AND FINAL RESULTS

1) TEST ENVIRONMENT

The experiments were conducted using a virtual machine configured with two dedicated processors and 2048 MB of RAM, hosted on an Intel Core i5-1135G7 CPU (2.42 GHz) with 8192 MB RAM, running a 64-bit Microsoft Windows 11 operating system version 22H2 (codenamed Sun Valley 2). The virtual machine runs a 64-bit Debian 12.4.0 operating system with an XFCE desktop environment. The code was implemented in Python 3.12 using the Pyfhel [71] libraries

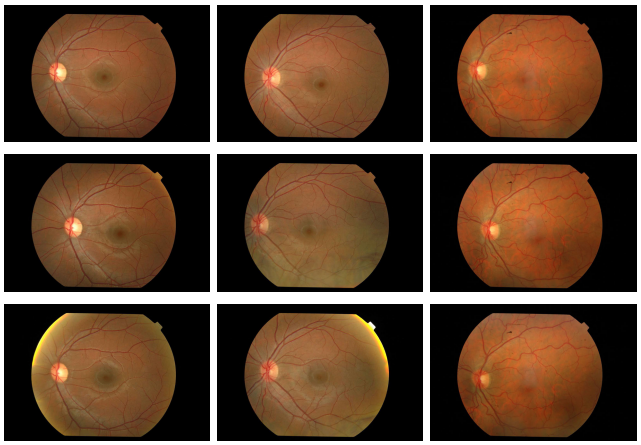


FIGURE 7. A set of fundus images from the RIDB database. In particular, each column represents theree samples belonging to a specific individual.

for HE, which were built using Python and Cython atop the Abstraction for Homomorphic Encryption Libraries (Afhel) and provide a Python wrapper for the Microsoft Simple Encrypted Arithmetic Library (SEAL) [72]. Additionally, Matlab R2023b was used for the feature extraction.

2) PERFORMANCE METRICS FOR EVALUATION

The goal is to test the following hypothesis by comparing a subject with a verified identity $\mathcal{I}^V$ and a claimed identity $\mathcal{I}^C$:

$$H_0 : \mathcal{I}^V = \mathcal{I}^C \quad \text{versus} \quad H_1 : \mathcal{I}^V \neq \mathcal{I}^C \tag{33}$$

Here, H_0 represents the null hypothesis, suggesting that the individual's claimed identity is valid (genuine or intra-class matching), while H_1 stands as the alternative hypothesis, implying that the individual's claimed identity is not valid (impostor or inter-class matching). Hence, regardless of whether the hypothesis is confirmed or not, the test is susceptible to two types of errors:

- False Acceptance Rate (FAR), which is the likelihood of accepting the null hypothesis when the input is invalid (type-I error), and
- False Rejection Rate (FRR), which is the likelihood of discarding the null hypothesis (accepting the alternative hypothesis) when the input is valid (type-II error).

FAR and FRR are closely related because an increase in one implies a decrease in the other. Mathematically, these two indices are expressed as follows:

$$\text{FAR} = \frac{1}{N}\sum_{k=1}^{N} \text{FAR}\left(\xi\right) \tag{34}$$

$$\text{FRR} = \frac{1}{N}\sum_{k=1}^{N} \text{FRR}\left(\xi\right) \tag{35}$$

where N identifies all identities being evaluated by the system, and

$$\text{FAR}\left(\xi\right) = \frac{\text{no. of FARs}}{\text{no. of impostor accesses}} \tag{36}$$

$$\text{FRR}\left(\xi\right) = \frac{\text{no. of FRRs}}{\text{no. of genuine accesses}} \tag{37}$$

On the other hand, the Genuine Acceptance Rate (GAR) is the likelihood of accepting the null hypothesis when the input is valid and may be used as an alternative to FRR:

$$\text{GAR}\left(\xi\right) = 1 - \text{FRR}\left(\xi\right) \tag{38}$$

The Receiver Operating Characteristic (ROC) curve visually illustrates the trade-off between GAR and FAR, showing how well the system performs across different discrimination thresholds. In contrast, the Detection Error Trade-off (DET) illustrates the balance between FRR and FAR as the threshold changes. The point where rejection and acceptance errors are equal is referred to as the Equal Error Rate (EER). Furthermore, the ROC and DET curves are threshold-independent, enabling the comparison of performance across different biometric systems under similar conditions [5].

TABLE 2. Retina identification database (RIDB) details.

Retina identification database	
Total no. of samples	100
No. of subjects	20
Samples for subject	5
Spatial resolution (px)	1504×1000
Dot per inch (dpi)	96
Bit depth	24

3) CRYPTOGRAPHIC PARAMETERS, EFFICIENCY, AND SCALABILITY

To estimate the total computation time, each section of the code was executed 200 times, and the mean time was considered. Table 3 provides a comparative analysis in terms of computational time varying security levels and parameters. Precisely, as described in Section IV-A, the bigger n, the more secure the scheme, but the slower the computations. Therefore, the chosen cryptographic scheme takes into account the following requirements.

1) A plaintext modulus p such that is a prime number with $p - 1$ being multiple of $2n$, taking into account that smaller p yields better performance.
   There is, however, a minimal size on p for a given polynomial modulus degree n:

$$\log_2(p)_{min} = \begin{cases} 14 & \text{if } n \leq 2^{11} \\ 16 & \text{if } n \leq 2^{12} \\ 17 & \text{otherwise} \end{cases} \tag{39}$$

   which implies that $p > 2n$, affecting the system efficiency in some cases.
2) A security level $\lambda \in \{128, 192, 256\}$ bits, which depends on both the bit-length of the total coefficient modulus q (small values ensure higher security) and the degree n

TABLE 3. Comparative computational time analysis varying security levels and parameters.

Parameters				Time w/ batching technique			
λ	n	q [bits]	p [bits | value]	$\mathcal{E}$ [s]	HD [s]	$\mathcal{D}$ [s]	Total [s]
128-bit	8192	174	17 | 114689	0.02513	0.16967	0.00139	0.19619
	16384	389	18 | 163841	0.04357	0.67491	0.00570	0.72418
192-bit	8192	114	17 | 114689	0.02048	0.13474	0.00107	0.15629
	16384	250	18 | 163841	0.03441	0.36472	0.00451	0.40364
256-bit	8192	78	17 | 114689	0.01941	0.08809	0.00096	0.10846
	16384	189	18 | 163841	0.02721	0.28571	0.00329	0.31621

of the polynomial modulus (large values ensure higher security) as follows:

$$n_{min} = \begin{cases} 2^{12} & \text{if } \lambda \in \{128, 192\} \text{ bits} \\ 2^{13} & \text{if } \lambda = 256 \text{ bits} \end{cases} \quad (40)$$

where the choice based on these values allows relinearisation used to control the growth of noise during homomorphic operations.

3) A coefficient modulus q that has to be chosen taking into consideration the upper bound for the total bit-length mined by the polynomial modulus degree n as described in [72].

Note that the choice of q has a significant impact on both the noise budget and operational efficiency of the encryption scheme. Increasing q expands the noise budget, allowing for more extensive encrypted computations, but also results in larger ciphertexts and greater computational demands. Conversely, smaller q values reduce these costs but restrict the number of operations that can be performed due to a reduced noise budget. Additionally, since the security level is inversely proportional to q, a larger coefficient modulus leads to a lower security level as well. Therefore, to limit the computational overhead while ensuring a sufficient security level and an adequate noise budget, we used the cryptographic parameters highlighted in Table 3.

Then, with the suggested configuration, the average computation times required to encode and then encrypt the vector data ω, compute the Hamming distance, and finally decrypt and decode the result are approximately 20 ms, 135 ms, and 1 ms, respectively. While the overall response time is fast enough for authentication applications, we acknowledge that the computational efficiency is indeed affected by the intensive calculations required by homomorphic encryption, which may not be conducive to real-time processing demands. In addition, for the proposed framework, scalability is a factor that can be effectively managed due to the authentication process relying on a single-sample matching approach. Then, the computational efficiency is not significantly compromised because each authentication event is an isolated instance. As such, even as the number of users grows, the system does not necessarily require more resources per authentication. The HE operations remain constant per authentication attempt, allowing the system to handle a larger load by managing the number of simultaneous authentications rather than the complexity of each. Consequently, the system performance remains stable even as scalability becomes a consideration.

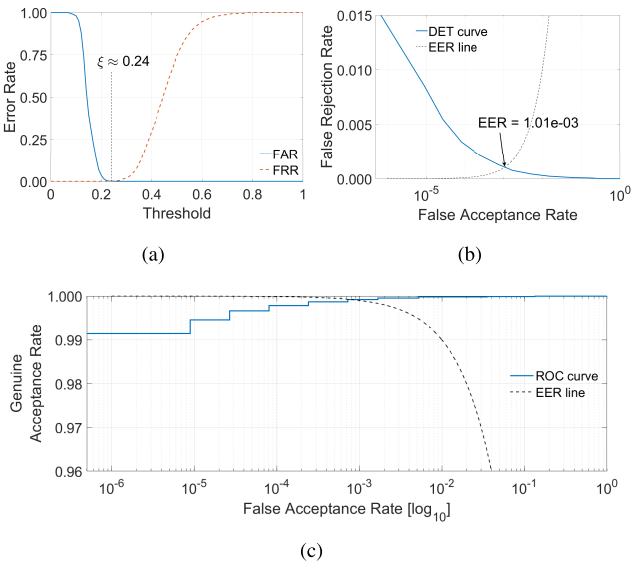


FIGURE 8. Results of the experiments performed on RIDB database: (a) False Acceptance Rate (FAR) and False Rejection Rate (FRR) curves, (b) semi-logarithmic Detection Error Trade-off (DET) curve with emphasised details, and (c) semi-logarithmic Receiver Operating Characteristic (ROC) curve.

4) PERFORMANCE ASSESSMENT

To assess the accuracy and effectiveness of the proposed solution, which relies on a single-sample approach for retina-based personal authentication, each sample in the database underwent a one-to-one matching test with each stored template. Specifically, the experimental protocol requires performing tests by considering 5 samples for each individual, amounting to a total of 100 samples. This configuration leads to a comprehensive set of $\binom{100}{2} = 4950$ tests, which

TABLE 4. Summary of the EER-based performance results derived from several published methods based on the RIDB database.

Reference	Year	Methodology		EER
		Features	Matching	
Dalal *et al.* [73]	2005	Histogram of Oriented Gradient (HOG) descriptors	Gaussian kernel SVM classifier	0.154
Fatima *et al.* [28]	2013	Gabor wavelet, multilayered thresholding, thinning	Mahalanobis distance measure	≈ 0.056
Waheed *et al.* [27] v1	2016	Gabor wavelet, supervised multilayered thresholding	Cosine similarity index matching	0.025
Waheed *et al.* [27] v2	2016	Structural information of the retinal image	Mean similarity score measure	≈ 0.286
Proposed method	2024	Laplacian of Gaussian, morphological operations, thinning	(Encrypted) Hamming distance	≈ 0.001

includes $20 \times \binom{5}{2} = 200$ genuine (intra-class) matches and the remaining tests being impostor (inter-class) matches. Figure 7 presents the experimental outcomes conducted on the RIDB database. In particular, Figure 7(a) illustrates the balance between the FRR and the FAR curves with varying thresholds corresponding to the matching score determined by the similarity measure between the templates Γ^{Q} and $\Gamma^{E}_{\mathcal{J}}$, as described in Equation (7). Figure 8(a) also provides information on the optimal point to choose as a threshold to optimise the variance between the genuine and impostor classes to assure an accurate classification of a user sample as either authentic or not. Figure 8(b) reports the detection error trade-off curve, which represents a graphical depiction of the trade-off between FAR and FRR without relying on specific operating points or thresholds. The curve is drawn by plotting the FAR and FRR values by varying the decision threshold of the classification system, thus providing a comprehensive overview of the performance of the system across different threshold values. The EER, given by the point at which the rate of false acceptances (incorrectly granting access to an unauthorised user) is equal to the rate of false rejections (incorrectly denying access to an authorised user), amounts to $1.0125 \cdot 10^{-3}$. These result indicate a highly accurate system, where the likelihood of both types of errors is extremely low. Additionally, Table 4 provides a detailed comparison of methods employing the RIDB database for performance evaluation, reporting the methodologies used for feature extraction and matching, as well as the EER outcomes. Despite the lack of any template protection in the compared methods, Table 4 enables a straightforward and unbiased comparison of EER, confirming the relevance and comparability of the results. Precisely, the tabular comparison clearly illustrates that our system achieved better performance with respect to the other works in the literature. For instance, studies by Waheed et al. [27] and Fatima et al. [28] reported an EER equal to $2.5 \cdot 10^{-2}$ and $5.57 \cdot 10^{-2}$, respectively, which are significantly higher than the results achieved by our system. Particularly, the proposed system achieved a ZeroFAR and a ZeroFRR, which represent the points on the curve where FAR becomes zero (ZeroFAR) and the point where FRR becomes zero (ZeroFRR), equal to $\mathrm{FRR}|_{\mathrm{FAR}=0} = 3.9164 \cdot 10^{-1}$ and $\mathrm{FAR}|_{\mathrm{FRR}=0} = 2.0378 \cdot 10^{-2}$. The ROC curve illustrated in Figure 8(c) is obtained by plotting GAR against FAR. In particular, the proposed system performed very well achieving a GAR at FAR $= 10^{-6}$ greater than 99%. Therefore, these results suggest that our framework is robust and reliable, making it a potentially valuable asset in very security-sensitive as well as privacy-oriented applications.

5) SECURITY ANALYSIS

Irreversibility and revocability: the irreversibility denotes the encryption characteristic whereby data, once encrypted, requires the specific decryption key to revert to its original form, ensuring security against unauthorised access. On the other hand, revocability refers to the capacity to invalidate a decryption key, thus preventing previously authorised users from accessing the data. In this scenario, homomorphically encrypted data is always protected even in the case of an attacker targeting the communication channel, since the decryption of the data without access to the private key κ_S is impractical due to the inherent complexity of the RLWE problem [55]. Thus, this aspect satisfies the standard requirement of irreversibility and, since it is not possible to retrieve the biometric data from one or more stolen encrypted biometric samples, the revocability requirement is also satisfied.

Unlinkability: the unlinkability refers to the impossibility to determine whether two or more protected biometric samples are generated from the same instance. This property is satisfied by using, for each user, a different set of keys for the encryption, thus preventing cross-matching between different databases.

Renewability: this property is assured since it is always possible to generate new protected biometric templates by using a new set of encryption keys. Therefore, in case of a security breach resulting in a database being compromised,

it is possible to invalidate previous protected templates and generate new ones using the same biometric sample.

Performance preservation: since the comparison in the encrypted domain is functionally identical (leading to the same scoring results) to its counterpart in the plaintext domain, there is no impact on biometric performance, with the benefit of ensuring user data privacy.

Apart from the above-mentioned criteria, it is also important to acknowledge that several common attacks pose significant threats to the integrity of the system. These include:

- Presentation attack: an adversary seeks to deceive the biometric sensor by presenting a fake or modified sample. However, the system inherently ensures resistance to such forgery attempts due to the complex and unique patterns of the retina, which are almost impossible to replicate convincingly.
- Replay attack: an adversary might intercept the encrypted data and replay it to the server to gain unauthorised access. To mitigate this, the system employs nonce-based authentication, rendering any replayed data obsolete and ineffective.
- Feature spoofing attack: an adversary's attempt to manipulate the biometric feature data, with the intention of matching a stored template, is mitigated by the use of feature encoding and homomorphic encryption. This approach effectively reduces the attack surface, ensuring the data remains encrypted and confidential, thereby reducing the likelihood of unauthorised access.
- Overriding final response: adversaries may seek to manipulate the AS encrypted response. However, the response remains encrypted until it reaches the client device, which uses its private key to decrypt and ascertain the final decision. This mechanism ensures that any tampered response, once decrypted, would result in an invalid outcome, thereby preserving the authenticity and integrity of the system verification process.

The proposed framework is designed to rigorously assess the system robustness against such adversarial actions, ensuring compliance with the aforementioned ISO/IEC standards and maintaining the efficacy of the retina recognition system amidst potential security breaches. Furthermore, the configuration of the FHE scheme proposed for the experimental evaluation is designed to ensure a 192-bit security level, bolstering defenses against quantum computing challenges [74]. Finally, the system verification accuracy and user data privacy are not compromised, as homomorphic encryption enables the biometric data to stay encrypted beyond the client side.

VI. CONCLUSION

This research study investigated a novel privacy-preserving framework for authentication based on homomorphic encryption and retina biometrics. Our findings highlight the effectiveness of combining these technologies to enhance data security and privacy. Indeed, using homomorphic encryption ensures that sensitive biometric data remains encrypted throughout the authentication process, addressing concerns about data exposure and unauthorised access. The adoption of retina biometrics adds an extra layer of security, leveraging unique physiological features for robust user authentication. The proposed framework has also demonstrated its robustness against a multitude of attacks, thereby fulfilling all the standard requirements described in the ISO/IEC 24745:2022. The experimental results performed on the RIDB database, showcase the ability of the system to achieve accurate authentication performance comparable to traditional methods without compromising privacy, clearly demonstrating that the proposed approach can compete with the state-of-the-art methods, achieving an EER approximately equal to 0.101%. In addition, the FAR and FRR values resulting from the experiments allow the system to be full compliant with the strict requirements of high security applications. Specifically, the value of GAR still can be considered equal to 1 for FAR levels as low as 10^{-4}, and setting the decision threshold to operate the system at ZeroFAR, the likelihood of rejecting a legitimate user is 2.038% with no possibility of accepting impostors. Moreover, the computational overhead introduced by homomorphically encrypting the biometric data, computing the Hamming distance in the encrypted domain, and then decrypting the resulting computation in order to obtain the final response, amounts to 156 ms, thus making the system suitable to be deployed for real-world applications in contexts where privacy and security are of primarily concern.

DATA AVAILABILITY

Readers can access our data on the findings by sending an e-mail to the corresponding author.

DISCLOSURES

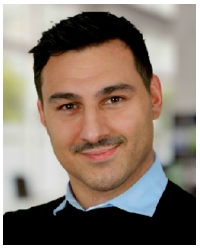
The authors have no relevant financial interests in the manuscript and no other potential conflicts of interest to disclose.

REFERENCES

[1] A. K. Jain and K. Nandakumar, "Biometric authentication: System security and user privacy," *Computer*, vol. 45, no. 11, pp. 87–92, Nov. 2012.
[2] A. K. Jain, A. Ross, and K. Nandakumar, *Introduction To Biometrics*. New York, NY, USA: Springer, 2011.
[3] D. Palma, P. L. Montessoro, G. Giordano, and F. Blanchini, "Biometric palmprint verification: A dynamical system approach," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 12, pp. 2676–2687, Dec. 2019.
[4] A. Adler and S. Schuckers, *Biometric Vulnerabilities, Overview*. Boston, MA, USA: Springer, 2009, pp. 160–168, doi: 10.1007/978-0-387-73003-5_65.
[5] D. Palma and P. L. Montessoro, "Biometric-based human recognition systems: An overview," in *Recent Advances in Biometrics*. London, U.K.: IntechOpen, 2022, ch. 2.
[6] S. M. Lajevardi, A. Arakala, S. A. Davis, and K. J. Horadam, "Retina verification system based on biometric graph matching," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3625–3635, Sep. 2013.
[7] D. Palma, P. L. Montessoro, G. Giordano, and F. Blanchini, "A dynamic algorithm for palmprint recognition," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Sep. 2015, pp. 659–662.

[8] M. K. Morampudi, M. V. N. K. Prasad, and U. S. N. Raju, "Privacy-preserving iris authentication using fully homomorphic encryption," *Multimedia Tools Appl.*, vol. 79, nos. 27–28, pp. 19215–19237, Jul. 2020.
[9] W. Wei, J. Wang, Z. Yan, and W. Ding, "EPMDroid: Efficient and privacy-preserving malware detection based on SGX through data fusion," *Inf. Fusion*, vol. 82, pp. 43–57, Jun. 2022.
[10] W. Stallings, *Cryptography and Network Security: Principles and Practice*. London, U.K.: Pearson, 2023.
[11] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Commun. ACM*, vol. 21, no. 2, pp. 120–126, 1978.
[12] B. Pulido-Gaytan, A. Tchernykh, J. M. Cortés-Mendoza, M. Babenko, G. Radchenko, A. Avetisyan, and A. Y. Drozdov, "Privacy-preserving neural networks with homomorphic encryption: C hallenges and opportunities," *Peer Peer Netw. Appl.*, vol. 14, no. 3, pp. 1666–1691, 2021.
[13] T. Elgamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 4, pp. 469–472, Jul. 1985.
[14] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proc. Int. Conf. Theory Appl. Cryptograph. Techn.* Cham, Switzerland: Springer, 1999, pp. 223–238.
[15] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in *Proc. 23rd Int. Conf. Theory Appl. Cryptol. Inf. Secur.*. Cham, Switzerland: Springer, 2017, pp. 409–437.
[16] D. Boneh, E.-J. Goh, and K. Nissim, "Evaluating 2-DNF formulas on ciphertexts," in *Proc. 2nd Theory Cryptogr. Conf. Theory Cryptogr. (TCC)*, Cambridge, MA, USA. Cham, Switzerland: Springer, 2005, pp. 325–341.
[17] C. Marcolla, V. Sucasas, M. Manzano, R. Bassoli, F. H. P. Fitzek, and N. Aaraj, "Survey on fully homomorphic encryption, theory, and applications," *Proc. IEEE*, vol. 110, no. 10, pp. 1572–1609, Oct. 2022.
[18] C. Gentry, *A Fully Homomorphic Encryption Scheme*. Stanford, CA, USA: Stanford university, 2009.
[19] Z. Brakerski, C. Gentry, and V. Vaikuntanathan, "(Leveled) fully homomorphic encryption without bootstrapping," *ACM Trans. Comput. Theory*, vol. 6, no. 3, pp. 1942–3454, Jul. 2014.
[20] J. Fan and F. Vercauteren, "Somewhat practical fully homomorphic Encryption.," *IACR Cryptol. ePrint Arch.*, vol. 2012, p. 144, Jan. 2012.
[21] J. Hoffstein, J. Pipher, and J. H. Silverman, "NTRU: A ring-based public key cryptosystem," in *Proc. Int. Algorithmic Number Theory Symp.* Cham, Switzerland: Springer, Jan. 1998, pp. 267–288.
[22] A. Aloufi, P. Hu, Y. Song, and K. Lauter, "Computing blindfolded on data homomorphically encrypted under multiple keys: A survey," *ACM Comput. Surv.*, vol. 54, no. 9, pp. 1–37, Dec. 2022.
[23] D. Palma, F. Blanchini, G. Giordano, and P. L. Montessoro, "A dynamic biometric authentication algorithm for near-infrared palm vascular patterns," *IEEE Access*, vol. 8, pp. 118978–118988, 2020.
[24] *Information Security, Cybersecurity and Privacy Protection—Biometric Information Protection*, ISO/IEC Standard 24745:2022, ISO/IEC JTC 1/SC 27, 2022, pp. 1–63.
[25] A. Uhl, "State of the art in vascular biometrics," in *Handbook of Vascular Biometrics*. Cham, Switzerland: Springer, 2020, pp. 3–61.
[26] D. Palma, F. Blanchini, and P. L. Montessoro, "A system-theoretic approach for image-based infectious plant disease severity estimation," *PLoS ONE*, vol. 17, no. 7, Jul. 2022, Art. no. e0272002.
[27] Z. Waheed, M. Usman Akram, A. Waheed, M. A. Khan, A. Shaukat, and M. Ishaq, "Person identification using vascular and non-vascular retinal features," *Comput. Electr. Eng.*, vol. 53, pp. 359–371, Jul. 2016.
[28] J. Fatima, A. M. Syed, and M. U. Akram, "A secure personal identification system based on human retina," in *Proc. IEEE Symp. Ind. Electron. Appl.*, Sep. 2013, pp. 90–95.
[29] E. Emary, H. M. Zawbaa, A. E. Hassanien, G. Schaefer, and A. T. Azar, "Retinal blood vessel segmentation using bee colony optimisation and pattern search," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 1001–1006.
[30] F. Sadikoglu and S. Uzelaltinbulat, "Biometric retina identification based on neural network," *Proc. Comput. Sci.*, vol. 102, pp. 26–33, Jun. 2016.
[31] S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, and G. Yang, "Hierarchical retinal blood vessel segmentation based on feature and ensemble learning," *Neurocomputing*, vol. 149, pp. 708–717, Feb. 2015.
[32] F. Jiu, K. Noronha, and D. Jayaswal, "Biometric identification through detection of retinal vasculature," in *Proc. IEEE 1st Int. Conf. Power Electron., Intell. Control Energy Syst. (ICPEICES)*, Jul. 2016, pp. 1–5.
[33] K. BahadarKhan, A. A. Khaliq, and M. Shahid, "A morphological Hessian based approach for retinal blood vessels segmentation and denoising using region based Otsu thresholding," *PLoS ONE*, vol. 11, no. 7, Jul. 2016, Art. no. e0158996.
[34] E. Imani, M. Javidi, and H.-R. Pourreza, "Improvement of retinal blood vessel detection using morphological component analysis," *Comput. Methods Programs Biomed.*, vol. 118, no. 3, pp. 263–279, Mar. 2015.
[35] M. Lim, A. B. J. Teoh, and J. Kim, "Biometric feature-type transformation: Making templates compatible for secret protection," *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 77–87, Sep. 2015.
[36] J. Bringer and H. Chabanne, "An authentication protocol with encrypted biometric data," in *Progress in Cryptology—AFRICACRYPT 2008*. Berlin, Germany: Springer, 2008, pp. 109–124.
[37] A. B. J. Teoh, A. Goh, and D. C. L. Ngo, "Random multispace quantization as an analytic mechanism for BioHashing of biometric and random identity inputs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 1892–1901, Dec. 2006.
[38] N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle, "Generating cancelable fingerprint templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 561–572, Apr. 2007.
[39] S. Kanade, D. Petrovska-Delacretaz, and B. Dorizzi, "Cancelable iris biometrics and using error correcting codes to reduce variability in biometric data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 120–127.
[40] A. Nagar and A. K. Jain, "On the security of non-invertible fingerprint template transforms," in *Proc. 1st IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2009, pp. 81–85.
[41] A. Juels and M. Sudan, "A fuzzy vault scheme," *Designs, Codes Cryptography*, vol. 38, no. 2, pp. 237–257, Feb. 2006.
[42] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *Proc. 6th ACM Conf. Comput. Commun. Secur.*, Nov. 1999, pp. 28–36.
[43] Y. Dodis, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," in *Proc. Int. Conf. Theory Appl. Cryptograph. Techn.*, Interlaken, Switzerland. Cham, Switzerland: Springer, 2004, pp. 523–540.
[44] M. J. Atallah, K. B. Frikken, M. T. Goodrich, and R. Tamassia, "Secure biometric authentication for weak computational devices," in *Proc. 9th Int. Conf. Financial Cryptogr. Data Secur.*, Roseau, Dominica. Cham, Switzerland: Springer, Jan. 2005, pp. 357–371.
[45] J. Bringer, H. Chabanne, M. Izabachene, D. Pointcheval, Q. Tang, and S. Zimmer, "An application of the goldwasser-micali cryptosystem to biometric authentication," in *Proc. 12th Australas. Conf. Inf. Secur. Privacy (ACISP)*, Townsville, QLD, Australia. Cham, Switzerland: Springer, 2007, pp. 96–106.
[46] W. Du and M. J. Atallah, "Secure multi-party computation problems and their applications: A review and open problems," in *Proc. Workshop New Secur. Paradigms (NSPW)*, 2001, pp. 13–22.
[47] B. Schoenmakers and P. Tuyls, "Efficient binary conversion for Paillier encrypted values," in *Proc. 24th Annu. Int. Conf. Theory Appl. Cryptograph. Techn.*, St. Petersburg, Russia. Cham, Switzerland: Springer, 2006, pp. 522–537.
[48] N. Dowlin, R. Gilad-Bachrach, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing, "CryptoNets: Applying neural networks to encrypted data with high throughput and accuracy," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2016, pp. 201–210.
[49] A. Juels and T. Ristenpart, "Honey encryption: Security beyond the brute-force bound," in *Proc. 33rd Annu. Int. Conf. Theory Appl. Cryptograph. Techn.*, Copenhagen, Denmark. Cham, Switzerland: Springer, 2014, pp. 293–310.
[50] J. Bringer, H. Chabanne, and A. Patey, "Privacy-preserving biometric identification using secure multiparty computation: An overview and recent trends," *IEEE Signal Process. Mag.*, vol. 30, no. 2, pp. 42–52, Mar. 2013.
[51] A. Abidin and A. Mitrokotsa, "Security aspects of privacy-preserving biometric authentication based on ideal lattices and ring-LWE," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2014, pp. 60–65.
[52] M. Barni, T. Bianchi, D. Catalano, M. Di Raimondo, R. D. Labati, P. Failla, D. Fiore, R. Lazzeretti, V. Piuri, A. Piva, and F. Scotti, "A privacy-compliant fingerprint recognition system based on homomorphic encryption and fingercode templates," in *Proc. 4th IEEE Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2010, pp. 1–7.

[53] H. Kikuchi, K. Nagai, W. Ogata, and M. Nishigaki, "Privacy-preserving similarity evaluation and application to remote biometrics authentication," in *Proc. 5th Int. Conf. Model. Decis. Artif. Intell. (MDAI)*, Sabadell, Spain. Cham, Switzerland: Springer, 2008, pp. 3–14.
[54] Z. Brakerski, "Fully homomorphic encryption without modulus switching from classical GapSVP," in *Proc. Annu. Cryptol. Conf.* Cham, Switzerland: Springer, Jan. 2012, pp. 868–886.
[55] V. Lyubashevsky, C. Peikert, and O. Regev, "On ideal lattices and learning with errors over rings," in *Proc. 29th Annu. Int. Conf. Theory Appl. Cryptograph. Techn.*, French Riviera, France. Cham, Switzerland: Springer, Jan. 2010, pp. 1–23.
[56] C. Gentry, C. Peikert, and V. Vaikuntanathan, "Trapdoors for hard lattices and new cryptographic constructions," in *Proc. 40th Annu. ACM Symp. Theory Comput.*, May 2008, pp. 197–206.
[57] Z. Brakerski and V. Vaikuntanathan, "Efficient fully homomorphic encryption from (standard) LWE," *SIAM J. Control Optim.*, vol. 43, no. 2, pp. 831–871, 2014.
[58] F. Weissbaum and T. Lugrin, "Symmetric cryptography," in *Trends in Data Protection and Encryption Technologies*. Springer, 2023, ch. 2, pp. 7–10.
[59] E. Hammer-Lahav, *The OAuth 1.0 Protocol*, document RFC 5849, Apr. 2010. [Online]. Available: https://www.rfc-editor.org/info/rfc5849
[60] J. L. Wayman, "Fundamentals of biometric authentication technologies," *Int. J. Image Graph.*, vol. 1, no. 1, pp. 93–113, Jan. 2001.
[61] M. Foracchia, E. Grisan, and A. Ruggeri, "Detection of optic disc in retinal images by means of a geometrical model of vessel structure," *IEEE Trans. Med. Imag.*, vol. 23, no. 10, pp. 1189–1195, Oct. 2004.
[62] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "Blood vessel segmentation methodologies in retinal images—A survey," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 407–433, 2012.
[63] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Commun. ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.
[64] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1148–1161, Nov. 1993.
[65] X. Song, Z. Chen, and D. Sun, "Iris ciphertext authentication system based on fully homomorphic encryption," *J. Inf. Process. Syst.*, vol. 16, no. 3, pp. 599–611, Jun. 2020.
[66] H. Chen, K. Laine, and R. Player, "Simple encrypted arithmetic library-seal v2. 1," in *Proc. Int. Conf. Financial Cryptogr. Data Secur.*, Sliema, Malta. Cham, Switzerland: Springer, 2017, pp. 3–18.
[67] N. Dowlin, R. Gilad-Bachrach, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing, "Manual for using homomorphic encryption for bioinformatics," *Proc. IEEE*, vol. 105, no. 3, pp. 552–567, Mar. 2017.
[68] M. Albrecht, M. Chase, H. Chen, J. Ding, S. Goldwasser, S. Gorbunov, S. Halevi, J. Hoffstein, K. Laine, K. Lauter, S. Lokam, D. Micciancio, D. Moody, T. Morrison, A. Sahai, and V. Vaikuntanathan, "Homomorphic encryption standard," Cryptol. ePrint Archive, Tech. Paper 2019/939, 2019. [Online]. Available: https://eprint.iacr.org/2019/939
[69] W. Yang, S. Wang, H. Cui, Z. Tang, and Y. Li, "A review of homomorphic encryption for privacy-preserving biometrics," *Sensors*, vol. 23, no. 7, p. 3566, Mar. 2023.
[70] M. U. Akram, A. Abdul Salam, S. G. Khawaja, S. G. H. Naqvi, and S. A. Khan, "RIDB: A dataset of fundus images for retina based person identification," *Data Brief*, vol. 33, Dec. 2020, Art. no. 106433.
[71] A. Ibarrondo and A. Viand, "Pyfhel: Python for homomorphic encryption libraries," in *Proc. 9th Workshop Encrypted Comput. Appl. Homomorphic Cryptography*, Nov. 2021, pp. 11–16.
[72] (Jan. 2023). *Microsoft SEAL (release 4.1)*. [Online]. Available: https://github.com/Microsoft/SEAL
[73] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR05)*, vol. 1, Jul. 2005, pp. 886–893.
[74] G. Alagic, J. Alperin-Sheriff, D. Apon, D. Cooper, Q. Dang, J. Kelsey, Y.-K. Liu, C. Miller, D. Moody, R. Peralta, A. Y. Robinson, and D. C. Smith-Tone, "Status report on the second round of the nist post-quantum cryptography standardization process," Dept. U.S. Dept. Commerce, NIST, Gaithersburg, MD, USA, Tech. Rep. 8309, 2020, vol. 2.



DAVID PALMA received the M.Sc. degree (magna cum laude) in electronic engineering and the Ph.D. and Doctor Europaeus degrees (Hons.) in industrial and information engineering from the University of Udine, Italy, in 2017 and 2021, respectively. He is currently a Professor in information security with the University of Udine, where he was a Research Associate with the Distributed and Dynamical Systems Research Group. From 2018 to 2019, he was a Doctoral Researcher with the Imperial College London, specializing in defense and security with the Control and Power Research Group. In 2023, he was awarded the University of Udine's Ph.D. Award and nominated for the EUROSIM Ph.D. Award for his doctoral research. His primary research interests include multidimensional signal processing, pattern recognition, and dynamical systems theory, from both theoretical and practical perspectives, with applications ranging from biometrics, surveillance, and cryptography to networking and cyber-physical security for critical infrastructures protection.

PIER LUCA MONTESSORO (Member, IEEE) was born in Turin, Italy, in 1961. He received the Dr.Eng. degree in electronic engineering from the Polytechnic of Turin, Italy, in 1986. He is currently a Full Professor in computer science with the University of Udine, Italy. Previously, he has been with the Italian National Council for Scientific Research, Italy, and a Scientific Consultant with Digital Equipment Corporation (later Compaq), MA, USA, in the field of simulation for VLSI design. His research interests include CAD systems for digital circuits design and on multimedia systems for e-learning, are currently focused on computer networks, ICT security, and pervasive computing, in particular distributed controls and algorithms for agents-based systems. He has been the Chair and the Organizer of the WCC 2013 Workshop "International Workshop on Cloud Convergence: Challenges for Future Infrastructures and Services," hosted in the IEEE ICC Conference and the Chair of the 30th Edition of Didamatica Conference, held in Udine.