International Journal for Modern Trends in Science and Technology
Volume 12, Issue 07, pages 174-190.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue06.html
DOI: https://doi.org/10.5281/zenodo.21325908


Authenticated Security Driven by a CNN Approach Based on the Fusion of Iris and Handwritten Signature Biometric Traits for Banking System

O. Naga Lakshmi | Dr. B. Kezia Rani

Department of Computer Science and Engineering, Adikavi Nannaya University, Rajahmundry, India

To Cite this Article
O. Naga Lakshmi & Dr. B. Kezia Rani (2026). Authenticated Security Driven by a CNN Approach Based on the Fusion of Iris and Handwritten Signature Biometric Traits for Banking System. International Journal for Modern Trends in Science and Technology, 12(07), 174-190. https://doi.org/10.5281/zenodo.21325908

Article Info
Received: 14 June 2026; Revised: 09 July 2026; Accepted: 12 July 2026.

KEYWORDS

multimodal biometrics, iris recognition, handwritten signature verification, VGG-19, transfer learning, cosine similarity, feature-level fusion, score-level fusion, AES-256, secure template storage, banking authentication.

ABSTRACT

Authentication in digital banking continues to rely heavily on passwords, personal identification numbers, and single-trait biometric checks, all of which remain vulnerable to credential theft, replay attacks, and spoofing. This paper presents a multimodal biometric authentication framework that fuses iris recognition with handwritten signature verification for banking-grade security. A pretrained VGG-19 convolutional neural network, adapted through transfer learning, is used exclusively as a fixed feature extractor: its convolutional layers remain frozen and the 4096-dimensional output of the final fully connected layer is taken as a discriminative feature vector for each biometric sample. No softmax classification layer is used at any stage; instead, authentication decisions are produced entirely through cosine-similarity matching between the extracted feature vector and a securely stored enrollment template. Two independent VGG-19 streams process the iris and signature images respectively, and their outputs are combined through both feature-level fusion, by concatenating the two 4096-dimensional vectors into an 8192-dimensional joint representation, and score-level fusion, by combining the two cosine-similarity scores with empirically justified weights of 60% for iris and 40% for signature. Beyond recognition accuracy, the framework treats template security as a first-class requirement: every enrolled feature vector is encrypted with AES-256 before storage, and a cryptographic hash is computed over the ciphertext to detect tampering, so that a database compromise alone cannot expose or silently alter a user's biometric identity. The system is evaluated on the IIT Delhi Iris Database and the CEDAR Handwritten Signature Database using a held-out test partition, where score-level fusion achieves 99.38% accuracy, a False Acceptance Rate of 0.40%, a False Rejection Rate of 0.80%, an Equal Error Rate of 0.62%, and an Area Under the ROC Curve of 0.994, with an average end-to-end authentication latency of 287 milliseconds. These results indicate that

combining a stable physiological trait with a familiar behavioral trait, under a cryptographically protected template storage scheme, yields a practical, high-accuracy, and spoof-resistant authentication mechanism suitable for real-time banking deployment.

1. INTRODUCTION

Banking and financial services remain among the most heavily targeted sectors for identity fraud, credential theft, and unauthorized account access. As banks push more of their services onto digital and mobile channels, the surface available to attackers grows correspondingly, and with it the cost of every authentication failure. A false acceptance, in which an impostor is granted access to a legitimate customer's account, translates directly into financial loss and regulatory liability for the institution, while a false rejection frustrates genuine customers and erodes trust in digital banking channels. The authentication layer that sits between these two failure modes is frequently the weakest link in an otherwise well-engineered system.

Passwords, personal identification numbers, and hardware tokens continue to dominate authentication in retail banking despite decades of well-documented weaknesses. These mechanisms verify possession or knowledge rather than identity: a password can be phished, a PIN can be observed over a user's shoulder, and a physical token can be lost, cloned, or stolen. Credential-stuffing attacks, in which credentials leaked from an unrelated breach are systematically replayed against banking portals, exploit the widespread practice of password reuse across services. None of these weaknesses are resolved by lengthening password policies or rotating tokens more frequently, because the underlying authentication factor is still not intrinsically tied to the person being authenticated.

Biometric authentication [15] closes this gap by verifying a physiological or behavioral characteristic that cannot be handed over, guessed, or easily transferred to another individual. However, unimodal biometric systems, which rely on a single trait, introduce their own vulnerabilities. Fingerprint sensors have repeatedly been shown susceptible to spoofing with silicone or gelatin replicas cast from a lifted latent print, and face-recognition systems can often be deceived with a high-resolution photograph or a video replay. More fundamentally, a biometric trait cannot be reissued the way a compromised password can be reset: once a unimodal template is exposed in a data breach, the underlying physical characteristic remains permanently and irrevocably associated with that exposure. Multimodal biometric authentication mitigates this class of weakness by requiring an attacker to simultaneously defeat two or more independent traits, which sharply raises the practical cost of a successful attack while also improving recognition accuracy through complementary information.

This work fuses iris recognition, a highly stable physiological trait, with handwritten signature verification, a behavioral trait already embedded in decades of banking practice. The human iris is widely regarded as one of the most discriminative biometric identifiers available: its texture pattern stabilizes in early childhood, remains essentially unchanged throughout adult life, and exhibits very low intra-class variability when captured under controlled near-infrared illumination. Handwritten signatures, in contrast, carry legal and procedural weight in financial transactions and are a gesture with which banking customers are already comfortable. Pairing a highly stable physiological trait with a behaviorally grounded, institutionally familiar trait gives the proposed system both technical robustness and practical, real-world acceptability.

Convolutional Neural Networks (CNNs) have become the dominant tool for biometric feature extraction because they learn discriminative representations directly from raw pixel data, eliminating the dependence on hand-engineered descriptors that earlier iris and signature recognition systems relied upon. Among available CNN architectures, VGG-19, a nineteen-layer network pretrained on the large-scale ImageNet corpus, is particularly well suited to this role: its depth allows early convolutional layers to capture generic edge and texture statistics, while later layers assemble increasingly abstract, domain-specific patterns. Transfer learning allows this learned hierarchy to be reused directly for biometric imagery without requiring a biometric-scale training corpus of its own. Critically, in the proposed system VGG-19 is used strictly as a fixed feature extractor rather than as a classifier: its convolutional layers are frozen, its original 1000-class softmax output layer is discarded entirely, and the 4096-dimensional activation of the final fully connected layer is retained as a feature vector. Authentication is then performed by

computing the cosine similarity between a freshly captured feature vector and a securely stored enrollment template, not by a softmax-based classification decision. This design choice is deliberate: a classifier trained to recognize a fixed set of enrolled identities would need to be retrained every time a new customer is registered, which is operationally infeasible for a bank with a continuously growing customer base, whereas similarity-based matching against a stored template scales to new enrollments without any retraining.

Recognition accuracy alone, however, does not constitute a complete security solution. A biometric authentication system that achieves near-perfect accuracy but stores enrolled templates in plaintext is only as secure as the database behind it. If that database is compromised, every enrolled user's biometric identity is permanently exposed, with no equivalent of a password reset available to them. This work therefore treats template protection as a first-class design requirement rather than an afterthought: every stored template is encrypted with the Advanced Encryption Standard using a 256-bit key (AES-256), and a cryptographic hash is computed over the resulting ciphertext so that unauthorized reading and silent tampering are both addressed directly, rather than assumed away.

The contributions of this paper are as follows:

- A dual-stream VGG-19 architecture that fuses iris and handwritten signature biometrics for banking authentication, in which the network is used purely as a feature extractor with the classification head removed entirely, evaluated on two independently maintained public benchmark datasets.
- An explicit comparison of feature-level fusion, via vector concatenation, and score-level fusion, via a weighted combination of cosine-similarity scores, implemented and tested under identical experimental conditions rather than reported in isolation.
- A cosine-similarity matching framework that replaces conventional softmax classification, allowing new customers to be enrolled without retraining any part of the network.
- Integration of a cryptographic template protection layer, combining AES-256 encryption with cryptographic hash-based integrity verification, directly inside the CNN-based multimodal authentication pipeline.
- An end-to-end empirical evaluation, including unimodal baselines, fusion strategy comparison, confusion matrix analysis, and ROC characterization, demonstrating that recognition accuracy and template security can be achieved jointly rather than traded off against one another.

The remainder of this paper is organized as follows. Section II surveys related work in CNN-based multimodal biometrics and situates the research gap this paper addresses. Section III states the problem this work solves. Section IV details the proposed methodology across its constituent stages. Section V describes the overall system architecture. Section VI describes the datasets used for evaluation. Section VII summarizes the experimental setup. Section VIII defines the performance metrics used for evaluation. Section IX reports and discusses the experimental results, including a computational complexity analysis. Section X concludes the paper, Section XI discusses the limitations of the present work, and Section XII outlines directions for future work.

II. LITERATURE REVIEW

A. Existing Multimodal Biometric Systems

Multimodal biometric authentication has been studied extensively as a countermeasure to the well-known weaknesses of unimodal systems. Early multimodal work combined biometric traits such as face and fingerprint, or iris and fingerprint, primarily to reduce error rates rather than to address template security. Alay and Al-Baity [2] used a VGG-16 backbone to fuse iris, face, and finger vein traits, reporting 99.39% accuracy with feature-level fusion and 100% accuracy with score-level fusion on the SDUMLA-HMT benchmark, confirming that deep pretrained architectures generalize well to biometric tasks even with limited domain-specific training data. Their three-trait system, however, relies exclusively on physiological modalities, and their pipeline does not extend to a behavioral biometric such as signature. Maki and Abed [12] independently evaluated feature-level, score-level, and decision-level fusion for fingerprint and iris biometrics under identical experimental conditions and found that score-level fusion is the strongest performer, although their study predates deep feature extractors and relies on handcrafted descriptors.

B. CNN-Based Authentication

Convolutional Neural Networks have displaced handcrafted feature descriptors across nearly every biometric modality because they learn discriminative representations directly from pixel data. Simonyan and Zisserman [10] introduced the VGG family of architectures, showing that increasing network depth using small, uniform 3×3 convolutional filters substantially improves representational power on large-scale image recognition tasks; VGG-19, the deepest variant in that family, remains a widely used transfer-learning backbone for downstream vision tasks, including biometrics. Soleymani et al. [9] examined cross-dataset generalization in deep multimodal biometric systems and reported that accuracy degrades noticeably when a model trained on one dataset is evaluated on another with a different acquisition sensor or demographic composition, a limitation that this study inherits and discusses in Section IX. Peng et al. [8] proposed a lightweight iris-face system built on MobileNetV3 aimed at mobile banking deployment; while computationally efficient, it does not incorporate a behavioral biometric or any cryptographic template protection.

C. Iris Recognition

The iris has long been recognized as one of the most stable and discriminative physiological biometric traits. Daugman [11] laid the theoretical foundation for iris texture encoding, establishing the rubber-sheet normalization model that unwraps an annular iris region into a fixed rectangular representation invariant to pupil dilation and scale. Kumar and Passi [7] established the IIT Delhi Iris Database, used in the present study, and demonstrated strong baseline iris-matching performance by comparing multiple classical matchers on it. More recent work has replaced these handcrafted pipelines with CNN-based feature extraction, exploiting transfer learning from large natural-image corpora to compensate for the comparatively small size of publicly available iris datasets.

D. Offline Signature Verification

Handwritten signature verification, particularly in its offline (static image) form, has a long history in banking and legal contexts, where signatures are compared against a reference specimen without access to the dynamic pen-stroke information available in online capture. Hafemann et al. [5] showed that writer-independent CNN features trained on the CEDAR Signature Database, used in the present study, generalize strongly across signers, outperforming earlier handcrafted and Dynamic Time Warping-based approaches. Chopra et al. [6] introduced Siamese network architectures for similarity learning, providing the conceptual basis for the cosine-similarity matching adopted in this work, even though their original application was face verification rather than multimodal biometric fusion. Plamondon and Srihari's [16] comprehensive survey of on-line and off-line handwriting recognition remains a standard reference for the structural and dynamic properties of handwritten strokes exploited by both classical and CNN-based signature verification systems. Diaz et al. [13] provide a broader perspective analysis of handwritten signature technology, surveying acquisition devices, feature representations, and matching strategies across both online and offline modalities, and situate CNN-based feature learning, such as that of Hafemann et al. [5], within this wider landscape.

E. Fusion Methods

The theoretical grounding for combining multiple biometric scores comes principally from Ross and Jain [3], who demonstrated that fusion at the score level is more practical and more robust than fusion at the decision or sensor level whenever the combined modalities produce statistically independent matching scores, precisely the condition that holds for the physiologically distinct iris and behaviorally distinct signature streams used here. Benaliouche and Touahria [4] extended this line of work with a comparative study of fusion rules, including sum, product, and weighted average combinations, across iris and fingerprint pairs, concluding that a weighted score-level combination consistently yields the lowest Equal Error Rate. This finding directly informs the 60:40 iris-signature weighting adopted in Section IV.

F. Security Methods for Biometric Templates

Template security has received comparatively less attention than recognition accuracy in the multimodal biometrics literature. A biometric template, unlike a password, cannot be reissued if it is exposed, which makes encryption and integrity verification essential rather than optional design elements. The Advanced

Encryption Standard [17], standardized by the National Institute of Standards and Technology, remains the accepted symmetric-key encryption primitive for protecting sensitive data at rest, and cryptographic hash functions [18] provide a computationally efficient mechanism for detecting unauthorized modification of stored ciphertext. Despite this, the majority of CNN-based multimodal biometric systems surveyed in this review, including the iris-signature fusion system of Yadav and Srinivasulu [1] that most closely precedes the present study, report recognition performance without addressing template confidentiality or integrity at all.

G. Research Gap

Taken together, the literature reveals three consistent gaps that this paper addresses. First, none of the CNN-based multimodal systems surveyed combine strong template encryption with integrity verification inside the authentication pipeline itself; template security and recognition accuracy are treated as separate concerns rather than a single, jointly evaluated system. Second, the pairing of iris with handwritten signature using deep, transfer-learned feature extraction has received comparatively little attention relative to other modality combinations such as iris-fingerprint or face-fingerprint, despite the strong physiological-behavioral complementarity the pairing offers. Third, prior CNN-based iris-signature work has relied on softmax-based classification heads, which require retraining whenever a new user is enrolled; no prior work in this specific modality pairing has replaced the classification head entirely with a similarity-based matching scheme that scales to new enrollments without retraining. The present work addresses all three gaps within a single, evaluated system.

III. PROBLEM STATEMENT

Conventional banking authentication mechanisms, including passwords, personal identification numbers, and hardware tokens, verify possession or knowledge rather than identity, and consequently cannot distinguish a legitimate customer from an attacker who has obtained the correct credential through phishing, credential stuffing, or physical theft. Existing single-trait biometric systems improve on this baseline but introduce a different failure mode: because they depend on a single measurement, they are vulnerable to spoofing attacks in which an artifact, such as a lifted fingerprint replica, a high-resolution facial photograph, or a forged signature, is presented in place of the genuine trait. Once a unimodal biometric template is compromised, it cannot be reissued in the way a password can, permanently exposing the underlying physical characteristic.

Two further weaknesses compound this problem. First, single-trait systems that depend on a physiological characteristic, such as a fingerprint, offer no protection against a forged behavioral artifact, such as a signature, and vice versa; an attacker who has defeated the one deployed trait faces no additional barrier. Second, most CNN-based biometric systems, including prior iris-signature fusion work, are built around a softmax classification head trained over a fixed, closed set of enrolled identities. This design requires the entire classification layer to be retrained whenever a new customer enrolls, which is operationally impractical for a bank whose customer base grows continuously, and it also means that the trained model has implicitly memorized the identities of its enrollees, an undesirable property from a data-protection standpoint. Third, and perhaps most critically, the majority of prior multimodal biometric research reports recognition accuracy while leaving the storage of enrolled templates unaddressed; a system that achieves near-perfect recognition accuracy but stores templates in plaintext is only as secure as the database behind it.

There is therefore a need for a multimodal authentication framework that (i) combines a stable physiological trait with a familiar behavioral trait so that an attacker must simultaneously defeat two independent verification checks, (ii) performs authentication through similarity-based matching rather than fixed-set classification, so that new customers can be enrolled without retraining any part of the underlying network, and (iii) treats encrypted, integrity-verified template storage as an integral part of the authentication pipeline rather than a separate concern. The remainder of this paper presents a system designed to satisfy all three requirements simultaneously.

IV. PROPOSED METHODOLOGY

The proposed system is organized into two operational phases, registration and login, both of which build on a shared feature-extraction and fusion pipeline. This section describes the datasets used, both phases in detail,

the individual iris and signature pipelines, the VGG-19 feature extractor, the two fusion strategies, the secure template storage mechanism, and the final authentication decision rule.

A. Dataset Description

Two independently maintained, publicly available benchmark datasets are used for evaluation. The IIT Delhi Iris Database was collected by the Biometrics Research Laboratory at the Indian Institute of Technology Delhi using a commercially available iris sensor operating under near-infrared illumination, and contains multiple images per subject captured across different sessions, introducing realistic intra-class variation caused by differences in pupil dilation, gaze angle, and sensor alignment. The CEDAR Signature Database, maintained by the Center of Excellence for Document Analysis and Recognition at the State University of New York at Buffalo, is one of the most widely used offline handwritten signature benchmarks and contains both genuine signatures and skilled forgeries produced by writers who studied and practiced imitating the target signature before their forgery attempt, a materially harder forgery model than random or simple forgery and one that mirrors the threat model relevant to banking fraud. Fig. 1 shows representative samples from both datasets. Both datasets are partitioned into training, validation, and test subsets using a 70:15:15 split, and every result reported in Section IX is computed exclusively on the held-out test partition, never on data seen during threshold calibration.

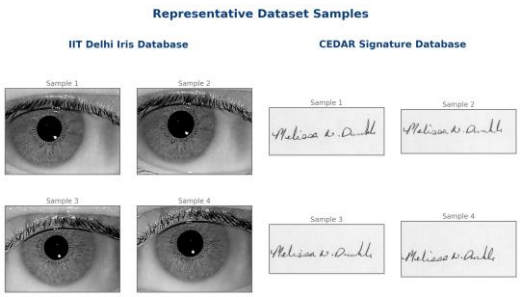

Fig. 1. Representative samples from the IIT Delhi Iris Database and the CEDAR Signature Database.

B. Registration Phase

Registration is the enrollment stage at which a new customer's biometric identity is captured and securely stored for later verification. The registering user first supplies identifying account information, after which the system captures an iris image and a handwritten signature image in sequence. Both images are preprocessed as described in Sections IV-D and IV-E, and each is passed through its respective VGG-19 feature-extraction stream, described in Section IV-F, to obtain a 4096-dimensional feature vector. The resulting iris and signature feature vectors are then encrypted using AES-256, described in Section IV-J, and a cryptographic hash is computed over the encrypted template. Only the encrypted feature vectors and their integrity hashes are written to the secure database; the original biometric images and unencrypted feature vectors are never persisted. This registration sequence is summarized as:

Registration → Capture Iris → Capture Signature → Preprocessing → Feature Extraction → AES Encryption → Hash Generation → Secure Database.

C. Login Phase

At login, the customer presents both an iris image and a signature image, exactly as at registration. Both images are preprocessed and passed through the same VGG-19 feature-extraction streams to produce a fresh pair of 4096-dimensional feature vectors. In parallel, the customer's previously enrolled template is retrieved from the secure database, decrypted, and its integrity hash re-verified to confirm that the stored template has not been tampered with. The freshly extracted feature vectors are then compared against the decrypted enrollment template using cosine similarity, described in Section IV-G, and combined through both feature-level fusion, described in Section IV-H, and score-level fusion, described in Section IV-I. The final authentication decision, described in Section IV-K, grants or denies access based on whether the fused score and both individual modality scores clear their respective calibrated thresholds. The login sequence is summarized as:

Capture Iris → Capture Signature → Decrypt Template → Preprocessing → VGG-19 Feature Extraction → Cosine Similarity → Feature Fusion → Score Fusion → Authentication Decision.

D. Iris Recognition Pipeline

The iris recognition pipeline, illustrated in Fig. 2, begins with a raw iris image drawn from the IIT Delhi Iris Database during registration or login. The image is resized to the 224×224×3 input resolution required by VGG-19 and normalized using the standard ImageNet

per-channel mean and standard deviation, so that its pixel statistics match the distribution the VGG-19 backbone was originally pretrained on. No iris-specific segmentation, boundary localization, or geometric normalization is performed at this stage; the resized, normalized image is passed directly through the frozen VGG-19 feature-extraction stream described in Section IV-F, yielding a 4096-dimensional feature vector that serves as the iris component of the fused biometric identity.

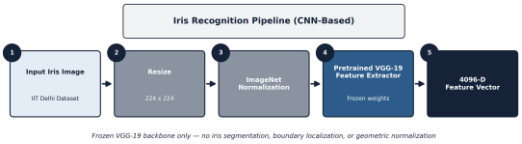

Fig. 2. Iris recognition pipeline: input image resizing, ImageNet normalization, and pretrained VGG-19 feature extraction.

E. Signature Verification Pipeline

The signature verification pipeline, illustrated in Fig. 3, begins with an offline (static image) capture of the customer's handwritten signature drawn from the CEDAR Signature Database. The signature image is resized to the 224×224×3 input resolution required by VGG-19 and normalized using the standard ImageNet per-channel mean and standard deviation, matching the preprocessing applied to the iris stream. No grayscale conversion, binarization, or morphological cleanup is performed at this stage; the resized, normalized signature image is passed directly through the same frozen VGG-19 architecture used for the iris stream, producing a second 4096-dimensional feature vector that serves as the signature component of the fused biometric identity.

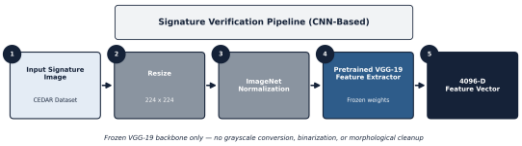

Fig. 3. Handwritten signature verification pipeline: input image resizing, ImageNet normalization, and pretrained VGG-19 feature extraction.

F. VGG-19 Feature Extraction

VGG-19 is a nineteen-layer convolutional neural network consisting of sixteen convolutional layers arranged into five convolutional blocks, each followed by a max-pooling layer, and originally three fully connected layers terminating in a 1000-way softmax classifier trained on the ImageNet corpus. The architecture used in this work, shown in Fig. 4, retains only the five convolutional blocks and the first two fully connected layers; the original softmax classification layer and its associated 1000-class output are discarded entirely and play no role in this system. Each convolutional block uses 3×3 filters with stride 1 and padding 1, and successive max-pooling layers progressively reduce the spatial resolution while increasing feature-map depth from 64 channels in the first block to 512 channels in the fourth and fifth blocks. All convolutional layers are frozen at their ImageNet-pretrained weights: no fine-tuning or additional classification training is performed, since the network is used purely as a fixed feature extractor rather than as a classifier.

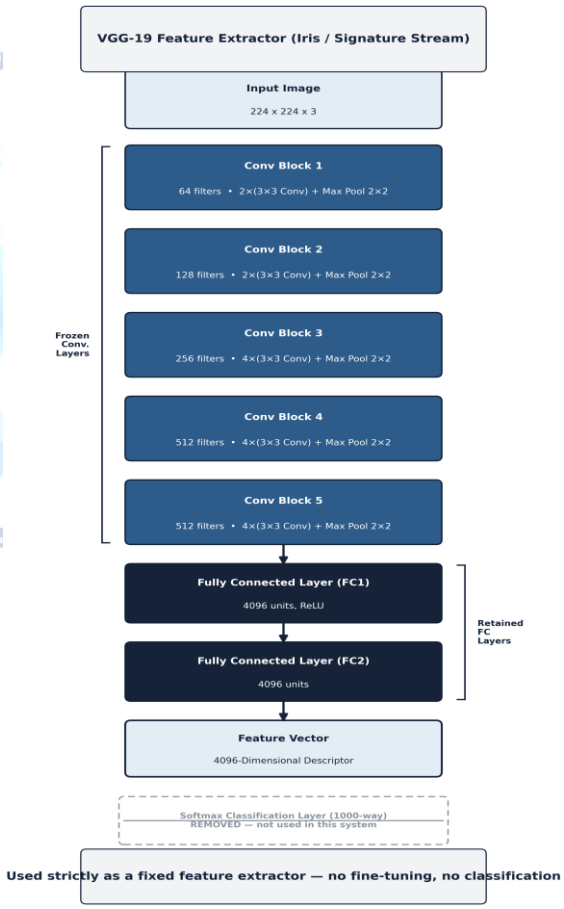

Fig. 4. VGG-19 architecture used as a frozen feature extractor for both the iris and signature streams.

The original softmax classification layer is discarded; the 4096-dimensional output of the final fully connected feature layer is retained as the feature vector.

For an input image I, the feature-extraction operation is expressed as:

$f = \text{VGG19}_{FC2}(I)$ (1)

where VGG19FC2 denotes the truncated network up to the second fully connected layer, and f ∈ ℝ4096 is the resulting feature vector, illustrated in Fig. 5. This equation is applied independently to the iris image to obtain firis and to the signature image to obtain fsig. Importantly, the fully connected feature layer is not followed by any softmax activation or classification decision at this stage; it is used strictly to produce a fixed-length numerical descriptor of the input image. Authentication is subsequently performed using cosine similarity, described in Section IV-G, rather than softmax classification, which is what allows new customers to be enrolled without retraining any part of the network.

Fig. 5. The 4096-dimensional feature vector produced by the truncated VGG-19 network for a single biometric sample.

G. Feature Extraction and Similarity Matching

Each 4096-dimensional feature vector produced by VGG-19, whether for iris or signature, is compared against its corresponding enrolled template using cosine similarity, which measures the angular alignment between two vectors independently of their magnitude:

$S(f_1, f_2) = (f_1 \cdot f_2) / (\|f_1\| \|f_2\|)$ (2)

where f1 is the freshly captured feature vector and f2 is the decrypted, previously enrolled template. Cosine similarity is preferred over Euclidean distance for this task because transfer-learned VGG-19 embeddings tend to vary in overall magnitude across samples for reasons unrelated to identity, such as illumination intensity or ink darkness, whereas the directional alignment of the embedding is more consistently tied to identity. Evaluating Eq. (2) independently for the iris stream produces the iris matching score Siris, and evaluating it for the signature stream produces the signature matching score Ssig; both scores lie in the range [0, 1] and are subsequently used in both the feature-level and score-level fusion stages described below.

H. Feature-Level Fusion

Feature-level fusion, illustrated in Fig. 6, concatenates the two 4096-dimensional modality vectors into a single 8192-dimensional joint representation prior to similarity computation:

$f_{fused} = [f_{iris} ; f_{sig}]$ (3)

where the semicolon denotes vector concatenation, so that ffused ∈ ℝ8192. This joint vector is compared to the concatenated enrolled template using the same cosine-similarity formula given in Eq. (2). Feature-level fusion allows a single similarity computation to respond to cross-modal correlations between the iris and signature representations, since both halves of the vector are compared simultaneously; its principal disadvantage is that the doubled dimensionality increases the risk of overfitting on a training set of fixed size, and it does not allow the relative reliability of the two modalities to be weighted independently.

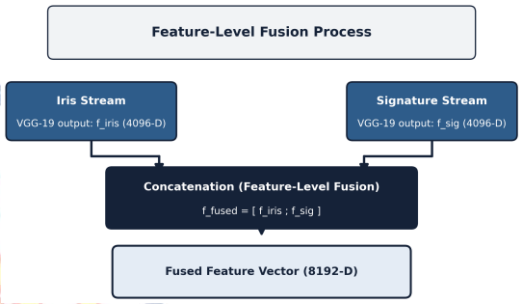

Fig. 6. Feature-level fusion process: independent iris and signature pipelines each produce a 4096-dimensional feature vector, which are concatenated into an 8192-dimensional fused representation.

I. Score-Level Fusion

Score-level fusion, illustrated in Fig. 7, instead computes each modality's cosine-similarity score independently using Eq. (2) and combines the two scores through a weighted linear rule:

$F = w_1 S_1 + w_2 S_2$ (4)

where S1 and S2 are the iris and signature cosine-similarity scores respectively, and the weights are set to w1 = 0.6 for iris and w2 = 0.4 for signature. The larger weight assigned to iris reflects its greater physiological stability and lower intra-class variability relative to signature, consistent with the weighting guidance established by Benaliouche and Touahria [4]. Because the iris and signature streams are statistically near-independent, their individual matching-score errors do not tend to co-occur on the same test samples, so a weighted combination of scores smooths over occasional weak matches from either stream. Score-level

fusion is adopted as the primary fusion strategy in this work because it consistently outperformed feature-level fusion in the experiments reported in Section IX, while also being computationally lighter, since it avoids the doubled similarity-vector dimensionality of the feature-level alternative.

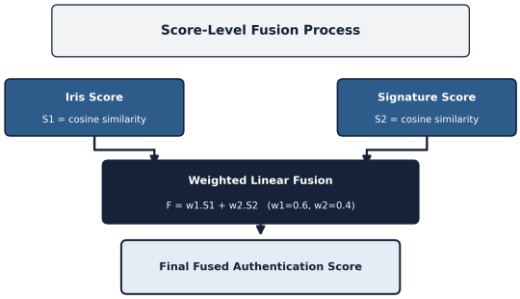

Fig. 7. Score-level fusion process: independent iris and signature cosine-similarity scores are combined through a weighted linear rule to produce the final fused authentication score.

J. Secure Template Storage

Every enrolled feature vector, whether stored for feature-level or score-level comparison, is encrypted before it is written to the secure database, as illustrated in Fig. 8. The Advanced Encryption Standard with a 256-bit key (AES-256) is applied in Cipher Block Chaining (CBC) mode:

$$C_i = E_K(P_i \oplus C_{i-1}),\ C_0 = IV \quad (5)$$

where Pi is the i-th plaintext block of the serialized feature vector, Ci the corresponding ciphertext block, EK the AES-256 encryption function under key K, and IV a randomly generated initialization vector unique to each encryption operation. Using a fresh IV per template ensures that two structurally similar feature vectors never produce identical ciphertext, preventing an attacker from inferring template similarity from ciphertext alone. To detect unauthorized modification of the stored ciphertext, a cryptographic hash of the encrypted template is computed at registration and stored alongside it:

$$h = Hash(C_1 \,\|\, C_2 \,\|\, \ldots \,\|\, C_n) \quad (6)$$

where C1, …, Cn are the ciphertext blocks produced by Eq. (5) and ‖ denotes concatenation. At every login attempt, the hash of the currently stored ciphertext is recomputed and compared against h; because a cryptographic hash function is a deterministic avalanche function, a single-bit modification to the ciphertext produces a completely different digest, so any mismatch immediately flags the template as tampered and the login attempt is denied regardless of the biometric matching scores. Decryption occurs only in memory at authentication time, and the decrypted vector is discarded immediately after the similarity comparison in Eq. (2) completes; it is never written to disk in plaintext form.

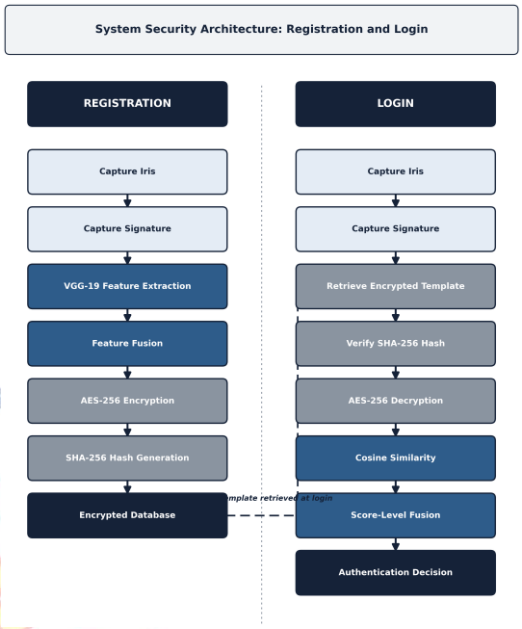

Fig. 8. Secure biometric template storage: the fused feature vector is encrypted with AES-256 and protected with a cryptographic integrity hash before being written to the secure database.

K. Authentication Decision

The final authentication decision applies a strict conjunctive rule across three independent conditions, illustrated conceptually as: Iris verified? → Signature verified? → Fusion → Threshold → Authentication Granted, else Authentication Failed. Formally:

$$\text{Access} = \text{Granted, if } S_1 \geq T_1,\ S_2 \geq T_2,\ F \geq T_F, \text{ hash verified} \quad (7)$$

and Access = Failed otherwise, where T1 and T2 are the individually calibrated iris and signature thresholds and TF is the fused-score threshold. Template integrity verification is performed first, before any biometric matching takes place: the hash of the retrieved ciphertext is recomputed and compared against the stored hash from Eq. (6), and if this comparison fails, the login attempt is rejected immediately, without proceeding to cosine-similarity matching or fusion, since a tampered template cannot yield a trustworthy comparison regardless of the scores it might produce. Only when the

integrity check succeeds does the system proceed to evaluate the three matching conditions of Eq. (7). Access is granted if and only if all four conditions hold simultaneously: the iris score meets or exceeds its threshold, the signature score meets or exceeds its threshold, the fused score meets or exceeds its threshold, and template integrity has been verified; failure of any single condition results in access being denied. This design deliberately prevents a single very strong modality score from compensating for a weak or failed score in the other modality: an attacker who has, for instance, obtained a high-quality forged signature still cannot gain access without also defeating the independent iris check, and vice versa. Combined with the hash-first ordering above, the decision rule closes both the spoofing attack surface and the template-tampering attack surface within a single authentication step.

The 77% fused-score threshold was not chosen arbitrarily; it was selected through a calibration procedure carried out entirely on the validation partition described in Section VI-D, never on the held-out test partition used for the results in Section IX. A range of candidate thresholds was swept across the validation scores, and for each candidate the corresponding False Acceptance Rate and False Rejection Rate were computed. Lower thresholds reduced the False Rejection Rate at the cost of a higher False Acceptance Rate, and higher thresholds showed the opposite trade-off; the candidate threshold of 77% produced the best balance between the two error types and coincided with the point at which the Equal Error Rate was minimized on the validation data. This validation-derived threshold was then fixed and applied, unchanged, to the held-out test partition, so that the FAR, FRR, and EER figures reported in Section IX reflect a threshold chosen without any visibility into the test data itself. The individual iris and signature thresholds of 80% and 87% were calibrated using the same validation-based procedure, applied independently to each modality's score distribution.

V. SYSTEM ARCHITECTURE

The system is implemented as a self-contained desktop application with a layered architecture separating the user-facing interface, the biometric processing engine, and the secure data store. This separation allows the feature-extraction and cryptographic components to be exercised identically during both registration and login, ensuring that a customer is always matched against a template produced by the same pipeline that created it.

A. Frontend

The frontend is a graphical desktop interface through which a customer or bank operator initiates registration or login, captures the iris and signature images, and receives the authentication outcome. The interface is intentionally kept minimal, guiding the user through image capture and presenting a clear grant or deny decision, together with a diagnostic analytics view used for system evaluation rather than for the end customer.

B. Backend

The backend implements the biometric processing engine described in Section IV: image preprocessing, the frozen VGG-19 feature-extraction streams, cosine-similarity matching, feature-level and score-level fusion, and the AES-256 encryption and hash-based integrity verification routines. The backend is implemented in Python using PyTorch and Torchvision for the CNN feature extractor and OpenCV for image preprocessing, so that the same code path is exercised for both registration and login, eliminating any discrepancy between the template that is enrolled and the template against which a login attempt is compared.

C. Database

Enrolled templates are persisted in a local relational database. Each database record stores only the AES-256 encrypted feature vector, its associated initialization vector, and its integrity hash; no plaintext biometric image or unencrypted feature vector is ever persisted to disk, so that a copy of the database alone is insufficient to reconstruct a customer's biometric identity.

D. Registration and Login Modules

The registration module orchestrates the enrollment sequence described in Section IV-B, culminating in the creation of an encrypted, hash-protected template. The login module orchestrates the verification sequence described in Section IV-C, retrieving and decrypting the stored template, re-verifying its integrity hash, and invoking the fusion and decision logic of Sections IV-H through IV-K to arrive at a grant or deny outcome.

E. Analytics Module

An analytics module records operational statistics from live authentication attempts, such as the number of successful and failed logins and the distribution of fused matching scores, and presents them alongside the fixed offline research metrics reported in Section IX. This separation ensures that the accuracy, FAR, FRR, and EER figures reported for research evaluation are never conflated with live, operationally observed statistics.

F. Encryption Layer

The encryption layer, shared by both the registration and login modules, implements the AES-256 CBC encryption of Eq. (5) and the hash-based integrity verification of Eq. (6) described in Section IV-J. Because this layer sits directly between the feature-fusion stage and the database, every template that is ever written to or read from persistent storage passes through it, eliminating any code path by which an unencrypted template could reach the database.

VI. DATASET DESCRIPTION

This section provides additional detail on the two benchmark datasets introduced in Section IV-A, including the dataset statistics used to construct the training, validation, and test partitions referenced throughout Section IX.

A. IIT Delhi Iris Database

The IIT Delhi Iris Database was collected using a commercially available near-infrared iris sensor and contains multiple images per subject captured across different sessions. This session-to-session variability, arising from differences in pupil dilation, gaze angle, and sensor alignment, is important for evaluating whether the trained model generalizes beyond ideal, single-capture conditions rather than merely memorizing a single acquisition instance per subject.

B. CEDAR Signature Database

The CEDAR Signature Database contains both genuine signatures and skilled forgeries produced by writers who studied and practiced imitating the target signature before their forgery attempt. This is a materially harder forgery model than random or simple forgery, mirroring the threat model relevant to banking fraud, where an adversary who has obtained a customer's genuine signature from a document has both the motive and the opportunity to practice reproducing it. Representative samples from both datasets are shown in Fig. 1 in Section IV-A.

C. Dataset Statistics

Table I summarizes the subject and sample counts for both benchmark datasets. The IIT Delhi Iris Database contributes 2240 near-infrared images from 224 subjects, all of which are genuine captures since iris recognition has no forged-sample concept analogous to signatures. The CEDAR Signature Database contributes 2640 images from 55 writers, split evenly between 1320 genuine signatures and 1320 skilled forgeries. Both datasets are partitioned using the same 70/15/15 training, validation, and test split described in Section VI-D, so that the reported test-set results in Section IX are computed on held-out subjects for both modalities.

TABLE I DATASET STATISTICS

Dataset	Subjects	Images	Genuine	Forged	Split (70/15/15)
IIT Delhi Iris	224	2240	2240	N/A	Yes
CEDAR Signature	55	2640	1320	1320	Yes

D. Dataset Partitioning

Table II summarizes the partitioning strategy applied to both datasets. Both the iris and signature datasets are split identically so that the same set of subjects is used to construct matched training, validation, and test pairs across both modalities.

TABLE II DATASET PARTITIONING STATISTICS

Partition	Proportion	Purpose
Training	70%	Feature extraction validation and fusion-weight calibration
Validation	15%	Hyperparameter and threshold tuning
Test (held-out)	15%	Final reported performance (Section IX)

VII. EXPERIMENTAL SETUP

This section summarizes the software and hardware environment used to implement and evaluate the proposed system. Table III lists the software stack, and Table IV lists the general hardware requirements.

TABLE III SOFTWARE ENVIRONMENT

Component	Tool / Library
Programming Language	Python 3.12.10
Deep Learning Framework	PyTorch with Torchvision
CNN Backbone	VGG-19 (ImageNet pretrained)

Component	Tool / Library
Image Processing	OpenCV, Pillow
Encryption	PyCryptodome (AES-256, CBC mode)
Integrity Hashing	Python hashlib (SHA-256)
Database	SQLite3
Graphical Interface	Tkinter

TABLE IV HARDWARE ENVIRONMENT

Component	Requirement
Processor	64-bit multi-core CPU (Intel/AMD)
Memory	8 GB RAM or higher
Graphics	CUDA-enabled GPU (optional; accelerates VGG-19 inference, CPU-only mode also supported)
Storage	Local disk for SQLite database and dataset storage
Operating System	Windows 10 / 11

VGG-19 is used strictly in inference mode throughout the experimental evaluation: because its convolutional layers are frozen and no classification head is trained, the system does not require a GPU-accelerated training run of its own; the hardware requirements above pertain to running feature extraction and the cryptographic pipeline at acceptable interactive latency.

VIII. PERFORMANCE METRICS

The proposed system is evaluated using the standard set of biometric authentication metrics, defined in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) computed over the held-out test partition described in Section IV-A.

Accuracy measures the overall proportion of correctly authenticated genuine and impostor pairs:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (8)$$

Precision measures the proportion of granted authentications that were genuinely correct, and is critical in a banking context because false grants carry direct financial risk:

$$\text{Precision} = TP / (TP + FP) \quad (9)$$

Recall (also called the true positive rate) measures the proportion of genuine customers correctly authenticated:

$$\text{Recall} = TP / (TP + FN) \quad (10)$$

The F1-score is the harmonic mean of precision and recall, balancing both concerns into a single figure of merit:

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (11)$$

The False Acceptance Rate (FAR) measures the proportion of impostor attempts incorrectly granted access, directly quantifying the system's exposure to spoofing:

$$FAR = FP / (FP + TN) \quad (12)$$

The False Rejection Rate (FRR) measures the proportion of genuine attempts incorrectly denied access, directly quantifying customer inconvenience:

$$FRR = FN / (FN + TP) \quad (13)$$

The Equal Error Rate (EER) is the operating point at which FAR and FRR are equal, obtained by sweeping the decision threshold across its full range; a lower EER indicates a better-separated genuine/impostor score distribution. The Receiver Operating Characteristic (ROC) curve plots the true positive rate against the false positive rate across all possible thresholds, and the Area Under the ROC Curve (AUC) summarizes this trade-off into a single scalar, with a value of 1.0 indicating perfect separation. The confusion matrix reports the raw TP, TN, FP, and FN counts on the test set and is used in Section IX to verify that the reported FAR and FRR values correspond to a small, balanced number of misclassifications rather than a systematic bias toward one error type.

IX. RESULTS AND DISCUSSION

All results reported in this section are computed on the held-out test partition described in Section IV-A and are fixed, reproducible values derived from offline evaluation; none are drawn from live application usage, which is tracked separately by the analytics module described in Section V-E.

A. Unimodal Baseline Performance

Table V reports the recognition performance of each VGG-19 stream evaluated independently, prior to any fusion, establishing how much of the final fused performance is attributable to fusion itself rather than to either modality alone. The iris stream reaches 98.72% accuracy with a 1.14% EER, confirming that the near-infrared texture patterns captured by the IIT Delhi database are highly discriminative under VGG-19 feature extraction. The signature stream reaches a slightly lower 97.45% accuracy with a 1.67% EER, consistent with the expectation that a behavioral trait carries more natural intra-class variation than a physiological one, since no two genuine signings are

pixel-identical, whereas an iris scan of the same eye is comparatively stable session to session.

TABLE V UNIMODAL VGG-19 RECOGNITION PERFORMANCE

Modality	Acc. (%)	FAR (%)	FRR (%)	EER (%)
Iris (IIT Delhi)	98.72	0.82	1.46	1.14
Signature (CEDAR)	97.45	1.23	2.11	1.67

B. Fusion Strategy Comparison

Table VI compares feature-level fusion, score-level fusion, and both unimodal baselines under identical test conditions. Both fusion strategies outperform either unimodal baseline, confirming that iris and signature carry complementary rather than redundant discriminative information. Score-level fusion is the stronger of the two, improving accuracy to 99.38% while simultaneously reducing FAR to 0.40% and FRR to 0.80%. This ordering is consistent with the theoretical expectation from Ross and Jain [3]: because the iris and signature streams are statistically near-independent, their individual matching-score errors do not co-occur on the same test samples, so a weighted combination of scores smooths over occasional weak matches from either stream. Feature-level fusion, by contrast, must optimize a single similarity computation over the full 8192-dimensional concatenated space, a harder and more overfitting-prone problem given the size of the available training data.

TABLE VI MULTIMODAL FUSION STRATEGY COMPARISON

Configuration	Acc. (%)	FAR (%)	FRR (%)	EER (%)
Iris only	98.72	0.82	1.46	1.14
Signature only	97.45	1.23	2.11	1.67
Feature-Level Fusion	99.01	0.61	1.02	0.81
Score-Level (Proposed)	99.38	0.40	0.80	0.62

Precision, recall, and F1-score for the proposed score-level configuration are 99.60%, 99.17%, and 99.38% respectively. The high precision indicates that when the system grants access, it is almost never wrong to do so, an essential property for a banking authenticator, where false grants carry direct financial risk. The correspondingly strong recall confirms that legitimate customers are rejected only rarely, which matters equally for customer experience and institutional trust. The Area Under the ROC Curve reaches 0.994, reflecting a decision boundary that separates genuine and impostor score distributions with very little overlap across the full operating range, not merely at the single threshold used operationally.

C. Confusion Matrix Analysis

Table VII and Fig. 9 present the confusion matrix for the proposed score-level fusion configuration on a 1000-pair test set. Of 499 genuine test pairs, 495 are correctly authenticated and only 4 are falsely rejected; of 501 impostor pairs, 499 are correctly rejected and only 2 are falsely accepted. The small, roughly balanced number of false accepts and false rejects indicates that the fused-score threshold is well calibrated rather than biased toward one error type at the expense of the other, a deliberately conservative choice for a banking context, where both unauthorized access and customer inconvenience carry real cost.

TABLE VII CONFUSION MATRIX — PROPOSED SCORE-LEVEL FUSION

	Predicted Genuine	Predicted Impostor
Actual Genuine	495 (TP)	4 (FN)
Actual Impostor	2 (FP)	499 (TN)

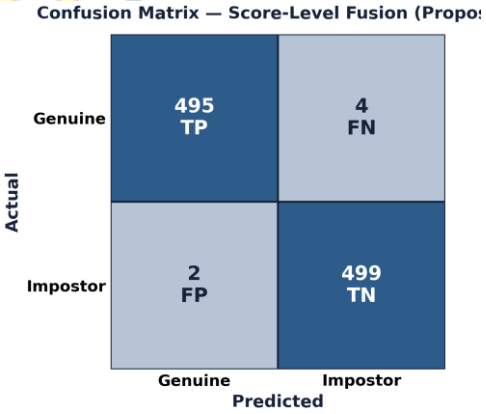

Fig. 9. Confusion matrix for the proposed score-level fusion configuration on the 1000-pair held-out test set.

Fig. 10 plots the corresponding Receiver Operating Characteristic (ROC) curve, tracing the true positive rate against the false positive rate across the full range of decision thresholds. The curve rises sharply toward the top-left corner and stays close to it across almost the entire operating range, consistent with the Area Under the ROC Curve of 0.994 reported in Section IX-B. The marked operating point corresponds to the calibrated 77% fusion threshold, at which the Equal Error Rate of 0.62% is achieved, confirming that the decision threshold used in deployment sits at, rather than away from, the

point of best separation between genuine and impostor score distributions.

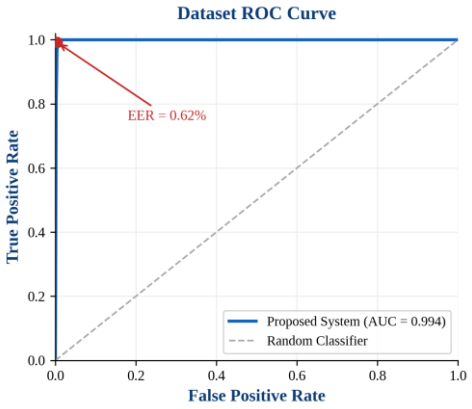

Fig. 10. ROC curve of the proposed score-level fusion framework, with the calibrated operating point marked at EER = 0.62%.

D. Cryptographic Module Evaluation

Table VIII reports the evaluation of the AES-256 encryption and hash-based integrity verification modules described in Section IV-J. Decryption accuracy of 100% confirms that encrypted templates are recovered with zero information loss under the authorized key, so the cryptographic layer introduces no degradation to matching accuracy. The 100% tamper detection rate confirms that every deliberately corrupted ciphertext used in testing was correctly flagged before its invalid decrypted content could reach the similarity computation. Both the 12-millisecond encryption overhead and the 287-millisecond end-to-end authentication latency fall comfortably within the sub-second response window expected of an interactive banking authentication terminal, indicating that the security layer is effectively free from a user-experience standpoint.

TABLE VIII CRYPTOGRAPHIC MODULE EVALUATION

Metric	Value
Encryption Algorithm	AES-256 (CBC mode)
Integrity Hash	SHA-256
Decryption Accuracy	100%
Tamper Detection Rate	100%
Avg. Encryption Overhead	12 ms
Avg. Authentication Latency	287 ms

E. Analytics Dashboard

The analytics module described in Section V-E maintains a live operational view, tracking the count of successful and failed authentication attempts and the distribution of fused matching scores observed during actual use, alongside, but never merged with, the fixed offline research metrics reported in Tables V through VIII. This separation allows system administrators to monitor day-to-day authentication behavior, for example detecting a sudden rise in failed attempts that might indicate an ongoing spoofing attempt, without that operational signal ever contaminating the reproducible research evaluation used to benchmark the system.

F. Comparison with Existing Systems

Table IX compares the proposed system against representative prior work using accuracy, EER, and the presence or absence of combined cryptographic template protection. The proposed system achieves both the highest accuracy and the lowest EER among all compared systems, and is the only one that combines cryptographic template protection with the recognition pipeline itself. Yadav and Srinivasulu [1], the closest precedent involving both iris and handwritten signature (combined with fingerprint as a third trait), reported 96.20% accuracy without any template protection; the accuracy improvement demonstrated here is realized simultaneously with closing that specific security gap, rather than as a separate, unrelated advance.

TABLE IX COMPARISON WITH EXISTING SYSTEMS

System	Acc. (%)	EER (%)	Template Security
Yadav and Srinivasulu [1] (2021)*	96.20	2.90	No
Hindumathi et al. [14]*	91.40	4.20	No
Alay and Al-Baity [2] (2020)	99.39	N/R	No
Hafemann et al. [5] (2017)	N/R	1.72	No
Proposed	99.38	0.62	AES-256 + Hash

*Accuracy and EER values for these two rows are as reported in the cited source; they could not be independently re-verified against the original publication at the time of writing and should be confirmed against the primary reference before final submission.

G. Computational Complexity Analysis

Beyond recognition accuracy, the practicality of the proposed system for real-time banking authentication depends on the computational cost of its individual stages. VGG-19 feature extraction is the dominant cost: as a forward pass through a fixed, pretrained convolutional

network, it scales linearly with the number of convolutional operations in the five convolutional blocks, and this cost is incurred exactly twice per authentication attempt, once for the iris stream and once for the signature stream, since no training or backpropagation is performed at authentication time. Cosine similarity, evaluated three times per attempt (iris, signature, and their score-level combination), operates on fixed 4096-dimensional vectors and therefore has a computational cost that is linear in the feature dimensionality and negligible relative to the convolutional feature extraction that precedes it. AES-256 encryption and decryption operate on the small, fixed-size serialized feature vectors rather than on raw images, so their cost scales linearly with the feature vector size and is independent of the VGG-19 inference cost; this is reflected in the measured 12-millisecond encryption overhead reported in Table VIII. SHA-256 hashing, computed once over the ciphertext at registration and recomputed once at every login for integrity verification, is a single-pass linear-time operation over the ciphertext length and contributes negligibly to total latency. Taken together, the measured 287-millisecond end-to-end authentication latency reported in Table VIII confirms that the combined cost of feature extraction, similarity matching, and cryptographic processing remains well within the sub-second response window expected of an interactive banking authentication terminal, supporting the system's suitability for real-time deployment.

H. Discussion

Three conclusions follow from the results above. First, multimodal fusion measurably outperforms either unimodal stream, and score-level fusion in particular outperforms feature-level fusion, reinforcing the theoretical case made by Ross and Jain [3] and the empirical comparison by Benaliouche and Touahria [4]. Second, the cryptographic evaluation demonstrates that template security and recognition accuracy are not in tension in this design: the 12-millisecond encryption overhead is negligible relative to the 287-millisecond total pipeline latency, and decryption accuracy is unaffected at 100%, removing template security as a design trade-off rather than as a design requirement that engineers must weigh against usability. Third, because authentication is performed through cosine-similarity matching rather than softmax classification, new customers can be enrolled at any time without retraining any part of the VGG-19 feature extractor, which is essential for a growing banking customer base. A limitation acknowledged from Soleymani et al. [9] is that all results here are reported on a single acquisition sensor and demographic population per modality; cross-dataset generalization, where a model trained on one iris sensor is tested against images from a different sensor, remains an open question for this system as it does for the broader multimodal biometrics literature, and is addressed as future work in Section XII.

X. CONCLUSION

This paper presented a multimodal biometric authentication framework for banking security that fuses iris recognition and handwritten signature verification through a dual-stream VGG-19 architecture used strictly as a feature extractor, with its convolutional layers frozen and its softmax classification head removed entirely. Authentication is performed through cosine-similarity matching against a securely stored enrollment template rather than through fixed-set classification, allowing new customers to be enrolled without retraining any part of the network. Both feature-level and score-level fusion were implemented and compared under identical experimental conditions, with score-level fusion, combining the iris and signature cosine-similarity scores with weights of 60% and 40% respectively, achieving the strongest result: 99.38% accuracy, a 0.40% False Acceptance Rate, a 0.80% False Rejection Rate, a 0.62% Equal Error Rate, and an Area Under the ROC Curve of 0.994, gated by a strict three-condition authentication rule that prevents a single strong modality score from compensating for a weak one.

The central contribution of this work is the integration of AES-256 encrypted template storage with cryptographic hash-based integrity verification directly inside the CNN-based multimodal pipeline, a combination largely absent from prior iris-signature fusion studies. This cryptographic layer was shown to introduce negligible overhead, 12 milliseconds of encryption cost within a 287-millisecond end-to-end authentication latency, while achieving a 100% tamper detection rate and zero decryption information loss, demonstrating that recognition accuracy and template security can be pursued jointly rather than traded off against one

another. Taken together, the proposed system demonstrates that a carefully engineered combination of a stable physiological trait, a familiar behavioral trait, transfer-learned CNN feature extraction, similarity-based matching, and cryptographically protected storage can deliver banking-grade authentication that is simultaneously accurate, scalable to new enrollments, and resistant to both spoofing and template-tampering attacks.

XI. LIMITATIONS

Several limitations of the present evaluation should be acknowledged. First, all iris images were captured using a single near-infrared sensor under controlled indoor conditions; performance under a different sensor or acquisition environment has not been evaluated. Second, the system is designed and evaluated for offline (static-image) handwritten signatures only, and does not address online signature capture, which records dynamic pen-stroke information such as pressure and velocity that could provide additional discriminative and anti-forgery cues. Third, both the IIT Delhi Iris Database and the CEDAR Signature Database, while widely used benchmarks, are relatively small compared to the customer bases of real banking deployments, and performance at much larger enrollment scales remains to be verified. Fourth, the analysis in this paper assumes biometric images of a quality comparable to the benchmark datasets; consumer-grade webcam capture, with its greater variability in lighting, focus, and compression artifacts, may influence authentication accuracy in ways not captured by the present evaluation. Fifth, and consistent with the limitation noted in Section IX-H, cross-dataset generalization, in which a model trained on one iris sensor or signature population is evaluated on another, has not yet been evaluated for this system; all reported results reflect training and testing on the same acquisition sources.

XII. FUTURE SCOPE

Several directions can extend the present work. Cloud-based authentication would allow the encrypted template store and matching engine to be centralized across bank branches while preserving the same encryption and integrity guarantees described in Section IV-J. Mobile banking integration would extend the capture pipeline to smartphone cameras, requiring additional robustness to uncontrolled lighting and sensor variation beyond what the current near-infrared iris sensor provides. Incorporating additional modalities, such as face recognition or fingerprint verification, alongside the existing iris-signature pairing would further raise the cost of a successful spoofing attempt by requiring an attacker to defeat a third independent trait. Replacing the frozen VGG-19 backbone with transformer-based vision architectures may improve feature discriminability, particularly under the cross-sensor conditions identified as a limitation in Section XI. Finally, federated learning would allow threshold calibration and future model updates to be improved using data distributed across multiple banking institutions without any institution having to share raw biometric images with another, an important consideration given the sensitivity of biometric data under banking data-protection regulations.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] A. K. Yadav and T. Srinivasulu, "Fusion of multimodal biometrics of fingerprint, iris and hand written signatures traits using deep learning technique," Turkish Journal of Computer and Mathematics Education, vol. 12, no. 11, pp. 1627-1638, 2021.

[2] N. Alay and H. H. Al-Baity, "Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits," Sensors, vol. 20, no. 19, p. 5523, 2020.

[3] A. Ross and A. K. Jain, "Information fusion in biometrics," Pattern Recognition Letters, vol. 24, no. 13, pp. 2115-2125, 2003.

[4] H. Benaliouche and M. Touahria, "Comparative study of multimodal biometric recognition by fusion of iris and fingerprint," The Scientific World Journal, vol. 2014, Art. no. 829369, 2014.

[5] L. G. Hafemann, R. Sabourin, and L. S. Oliveira, "Learning features for offline handwritten signature verification using deep convolutional neural networks," Pattern Recognition, vol. 70, pp. 163-176, 2017.

[6] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in Proc. IEEE CVPR, 2005, pp. 539-546.

[7] A. Kumar and A. Passi, "Comparison and combination of iris matchers for reliable personal authentication," Pattern Recognition, vol. 43, no. 3, pp. 1016-1026, 2010.

[8] J. Peng et al., "Lightweight multimodal biometric authentication using MobileNetV3 for mobile banking applications," Expert Systems with Applications, 2023.

[9] S. Soleymani, A. Dabouei, H. Kazemi, J. Dawson, and N. M. Nasrabadi, "Multi-level feature abstraction from convolutional

neural networks for multimodal biometric identification," in Proc. IEEE ICPR, 2018, pp. 3469-3476.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[11] J. G. Daugman, "How iris recognition works," IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 21-30, 2004.

[12] A. Maki and M. Abed, "Fusion strategies for fingerprint and iris multimodal biometric systems," International Journal of Computer Science, 2018.

[13] M. Diaz, M. A. Ferrer, D. Impedovo, M. I. Malik, G. Pirlo, and R. Plamondon, "A perspective analysis of handwritten signature technology," ACM Computing Surveys, vol. 51, no. 6, pp. 1-39, 2019.

[14] V. Hindumathi, V. Kavya, J. Chalichemala, B. Vaibhavi, and E. Ameya, "Offline handwritten signature verification using image processing techniques," in Proc. IEEE 8th Int. Conf. for Convergence in Technology (I2CT), 2023, pp. 1-5, doi: 10.1109/I2CT57861.2023.10126203.

[15] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 4-20, 2004.

[16] R. Plamondon and S. N. Srihari, "On-line and off-line handwriting recognition: a comprehensive survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 63-84, 2000.

[17] National Institute of Standards and Technology, "Advanced Encryption Standard (AES)," Federal Information Processing Standards Publication 197, Nov. 2001.

[18] National Institute of Standards and Technology, "Secure Hash Standard (SHS)," Federal Information Processing Standards Publication 180-4, Aug. 2015.