

A Deepfake Detection Framework for Securing Facial Biometrics Using Attention-Based Deep Learning and Blockchain

Zeeshan Ahmad¹ and Mamta Sharma²

¹Research Scholar, Department of Computer Science and Engineering, Jayoti Vidyapeeth Women's University, Jaipur

²Research Supervisor and Associate Professor, Department of Computer Science and Engineering, Jayoti Vidyapeeth Women's University, Jaipur

Received: 14th June, 2026; Revised: 20th July 2026; Accepted: 20th August, 2026; Available Online: 10th September 2026

ABSTRACT

Facial biometric authentication is widely deployed for identity verification, yet it is increasingly exposed to deepfake-based presentation and injection attacks and to large-scale breaches of centrally stored biometric templates. This paper proposes a deepfake-resilient facial biometric security framework that combines attention-based deep learning with blockchain technology. At registration, face images are pre-processed, augmented with a generative adversarial network (GAN), encoded by a convolutional neural network (CNN) with self-attention, and classified as real or fake by an artificial neural network (ANN). Only images classified as real are encrypted, bound to a unique user identifier and committed to a private blockchain through a cryptographic hash, whereas fake attempts are written to an immutable event log. At authentication, a fresh probe passes through the same pipeline and is verified by smart contracts that check the integrity of the registered commitment and the template similarity using zero-knowledge proofs (ZKPs), so that raw biometric features are never disclosed. We formalise the threat model, specify the architecture, the smart-contract interface and the security properties, and define an evaluation protocol based on the CIPLAB Real and Fake Face Detection and UTKFace datasets that covers detection accuracy, authentication error rates, adversarial robustness, demographic fairness and system overhead. The framework provides a basis for privacy-preserving, auditable and scalable facial authentication in digital identity management, financial services, healthcare and e-governance.

Keywords: deepfake detection; facial biometrics; self-attention; blockchain; zero-knowledge proofs; smart contracts; privacy-preserving authentication; presentation attack detection

How to cite this article: Zeeshan Ahmad, Mamta Sharma, A Deepfake Detection Framework for Securing Facial Biometrics Using Attention-Based Deep Learning and Blockchain *Int J Drug Deliv Technol.* 2026;16(1s): 146-154; DOI: 10.25258/ijddt.16.86s.17

Source of support: Nil.

Conflict of interest: None

1. INTRODUCTION

Face verification compares a live image or scan of a person with a reference image recorded from a trusted source and is one of the most convenient biometric modalities because it is accurate, contactless and non-intrusive. It is now embedded in mobile unlocking, remote onboarding (e-KYC), border control, healthcare access and digital-identity programmes. This popularity has made facial biometrics an attractive target. Advances in generative models, notably GANs [20], autoencoders and neural rendering, allow attackers to synthesise or manipulate faces that are difficult for both humans and machines to distinguish from genuine ones [24], [25]. A convincing deepfake presented to a camera, or injected directly into the capture pipeline, can defeat conventional recognition systems that were designed to resist only simple photo or screen replays.

A second weakness is architectural. Conventional systems keep biometric templates in centralised databases, so a single breach can expose very large numbers of irrevocable identifiers. Unlike a password, a face cannot be reset. A report cited in [1] records a breach in May 2024 that exposed fingerprints, facial images, signatures

and identifying marks of thousands of Indian law-enforcement officials, military personnel, educators and railway employees, and about 29,000 incidents of fraud linked to the Aadhaar Enabled Payment System were reported on the National Cyber Crime Reporting Portal. BioCatch [2] reported that account-takeover attacks accounted for more than half of the fraud cases among its customers in India. Stolen or synthesised biometric data feeds exactly this kind of identity fraud.

Blockchain technology [27] offers properties that complement deep-learning-based detection: decentralisation removes the single point of compromise, immutability makes silent modification of stored records detectable, and a shared ledger provides an auditable record of every registration, authentication and rejected attempt. Blockchain alone, however, does not recognise a deepfake, and deepfake detection alone does not protect the template store. Integrating the two is therefore a natural, but insufficiently explored, direction.

1.1 Gaps in current biometric security solutions

- 1) High-quality spoofing material and deepfake generators are escalating the threat to facial (and fingerprint) biometrics, and existing systems rely on

*Author for Correspondence: Zeeshan Ahmad

liveness and spoof detection that was not designed for synthetic media.

- 2) There is no standardised practice for encrypting and storing biometric data, which increases the risk of misuse or unauthorised sharing and raises surveillance and privacy concerns.
- 3) Current facial biometric systems struggle with emerging deepfake threats, which calls for more secure and robust detection and liveness mechanisms.

1.2 Objectives and contributions

The objectives of this work are (i) to develop a robust and efficient model for facial image authentication in which tampering can be traced, and (ii) to combine advanced deep-learning models with blockchain technology so that genuine data can be distinguished from deepfakes in real time. The main contributions are as follows.

- **Detection at the point of enrolment.** A two-phase architecture in which a CNN-self-attention encoder with an ANN classifier screens every image before it can become a registered template, and rejected images generate an auditable event.
- **Blockchain-anchored integrity and provenance.** Encrypted templates are bound to a user identifier and committed to a private blockchain by hash; smart contracts enforce access rules and keep an immutable log of fake-image events.
- **Privacy-preserving verification.** A ZKP-based matching step that tolerates the natural variability of biometric captures, which exact hash equality cannot, while keeping raw features undisclosed.
- **Security analysis and evaluation protocol.** A threat model, a mapping of threats to countermeasures, and a reproducible evaluation plan covering detection, authentication, robustness, fairness and system cost.

The rest of the paper is organised as follows. Section 2 reviews related work. Section 3 presents the proposed framework. Section 4 describes the experimental design. Section 5 outlines the evaluation reporting and expected outcomes, Section 6 discusses limitations and ethics, and Section 7 concludes.

2. LITERATURE REVIEW

Facial recognition traditionally analysed static geometric and texture cues, such as inter-ocular distance and jawline shape, to build a template for matching. CNNs substantially improved accuracy under changes in illumination, occlusion and expression. The same generative progress that helped recognition, however, produced GAN-, autoencoder- and rendering-based forgeries that threaten the trustworthiness of the whole pipeline [24], [25]. Existing work is reviewed below by methodological family.

2.1 Machine-learning-based approaches

Brown et al. [3] proposed a multimodal authentication system that fuses facial and fingerprint biometrics with

age and gender information using a Decision Tree classifier and stores data on a blockchain; reported accuracy was 82–86% on good-quality samples and 78.6% on low-quality samples. Gandhi et al. [4] compared Decision Tree, Random Forest, Bagging, XGBoost and ANN classifiers for deepfake video detection, with the ANN reaching the highest accuracy of 93%. Kumar et al. [5] combined PCA and LDA feature extraction with an ANN classifier on an IoT-oriented hardware stack (OpenCV, ESP-32) for real-time face recognition. Singh et al. [6] compared Haar and HOG descriptors on a Raspberry Pi and concluded that Haar-based models better suit embedded devices, illustrating the trade-off between computational efficiency and recognition performance. Raju et al. [7] combined local (LBP, HOG, Gabor) and global (DWT, PCA, ICA) features with neural-network and SVM classifiers on the Essex, ORL, Indian and YALE-B datasets and reported up to 99.45% accuracy with a neural network. These methods are lightweight but depend on hand-crafted features that generalise poorly to modern synthetic faces.

2.2 CNN-based deep-learning approaches

Bird and Lotfi [9] trained a custom CNN to separate real from AI-generated images (CIFAKE), used generative augmentation to address data limitations, supported the model with Grad-CAM explanations [29], and reported 92.98% accuracy, 93.7% precision, 91.2% recall and a 92.9% F1-score. Agarwal and Ratha [10] argued that a simple fusion of complementary cues can be competitive with considerably more complex deep networks for deepfake detection. Raza et al. [11] combined Xception, NASNet, MobileNet and VGG16 through hybrid transfer learning with hyper-parameter tuning, reaching 94% accuracy and 95% precision. Abukari et al. [12] proposed a lightweight CNN-based algorithm for frame-level detection of fake multimedia on social media. Marra et al. [13] benchmarked existing architectures and pretrained networks (DenseNet, Inception, Xception) on compressed data resembling social-network images and stressed the importance of architecture selection and benchmarking. Sugandhi et al. [14] used Deep Belief Networks with Restricted Boltzmann Machines on FFHQ, 100K-Faces, DFFD and CASIA-WebFace after Kalman-filter and Fisherface pre-processing, with accuracies between 94.92% and 98.82%. Goel et al. [15] applied a dual-branch CNN to copy-move forgery detection on MICC-F2000 and reported 96% accuracy. Laila and Tamilpavai [8] combined FaceNet512 and Gaussian Mixture Models for video and voice authentication in financial transactions secured with blockchain and smart contracts.

2.3 Attention-based deep-learning approaches

Attention mechanisms [21] let a model weight the most informative regions or channels, which is valuable when manipulation artefacts are subtle and local. Safwat et al. [16] integrated a modified ResNet [28] with channel attention and a GAN, obtaining an accuracy of 0.83 and a ROC-AUC of 0.825 against VGG16 and ResNet50 baselines. Ghani et al. [17] used a privacy-preserving self-attention GAN with K-means clustering and blockchain-

based storage on CelebA, reporting precision, recall, F1-score and accuracy of 0.948, 0.938, 0.943 and 0.947. Omar et al. [18] proposed an ensemble of CNNs with self-attention for deepfake video detection and used CutMix augmentation to improve generalisation across FaceForensics++ [22] manipulation types and Celeb-DF

[23], with AUC values from 87.62% to 99.74%. Prathibha et al. [19] introduced self-attention residual blocks within a continual-learning framework (SARB-DF) and reported AUC values between 90.00% and 99.78% across Celeb-DF, Deepfakes, Face2Face, FaceSwap and NeuralTextures.

Table I. Summary of representative related work (metrics as reported by the original authors, expressed as decimal fractions, with values calculated from them).

Ref.	Category	Technique	Data	Reported result	Calculated value (decimal)
[3]	ML + blockchain	Decision Tree on face, fingerprint, age and gender	Multimodal samples	Acc. 0.82–0.86 (good), 0.786 (low quality)	Mean acc. 0.8400
[4]	ML	DT, RF, Bagging, XGBoost, ANN	Deepfake videos	ANN best, Acc. 0.93	Error 0.0700
[7]	ML	PCA/DWT/ICA + NN/SVM	Essex, ORL, Indian, YALE-B	Acc. 0.9945 (NN)	Error 0.0055
[9]	CNN	Custom CNN + Grad-CAM	CIFAKE	Acc. 0.9298, F1 0.929	F1 0.9243; Error 0.0702
[11]	CNN	Hybrid transfer learning	Deepfake images	Acc. 0.94, Prec. 0.95	Error 0.0600
[14]	CNN	Deep Belief Network + RBM	FFHQ, 100K-Faces, DFFD, CASIA-WebFace	Acc. 0.9492–0.9882	Mean acc. 0.9687
[15]	CNN	Dual-branch CNN	MICC-F2000	Acc. 0.96	Error 0.0400
[16]	Attention	ResNet + channel attention + GAN	Real/fake faces	Acc. 0.830, AUC 0.825	Error 0.1700
[17]	Attention + blockchain	Self-attention GAN + K-means	CelebA	Acc. 0.947, F1 0.943	F1 0.9430; Error 0.0530
[18]	Attention	CNN ensemble + self-attention, CutMix	FF++, Celeb-DF	AUC 0.8762–0.9974	Mean AUC 0.9368
[19]	Attention	Self-attention residual block	Celeb-DF, FF++ types	AUC 0.9000–0.9978	Mean AUC 0.9489

2.4 Challenges and opportunities

Several challenges remain. Attackers can use increasingly capable generators to mimic a victim and bypass authentication, and detectors trained on one generator often degrade on another. Collecting and storing facial data raises privacy and surveillance concerns, and datasets that are unrepresentative of the population lead to unequal error rates across demographic groups. Large models and datasets hinder real-time deployment on constrained hardware, and adversarial perturbations can subtly alter inputs to mislead recognisers.

Opportunities mirror these challenges: combining multiple biometric sources (voice, gait, iris) to reduce reliance on the face alone; privacy-preserving techniques such as encrypted processing and distributed learning; fairness-aware algorithms; explainable AI to improve transparency and regulatory acceptance; user education; and cross-disciplinary collaboration among technologists, lawyers and policymakers.

2.5 Research gap

Table I shows that most deepfake detectors report strong detection scores but treat storage and verification as out of scope, whereas the blockchain-based systems [3], [8], [17] protect storage or transactions but either use shallow classifiers or address synthetic-face generation rather than detection at enrolment. To our knowledge, few works combine (a) self-attention-based real/fake screening at enrolment, (b) hash-anchored and encrypted template provenance on a ledger, and (c) zero-knowledge verification in a single, auditable framework. This is the gap addressed here.

3. PROPOSED FRAMEWORK

3.1 Overview and threat model

The system comprises a client device and a private (permissioned) blockchain platform equipped with smart contracts, as shown in Fig. 1. It operates in two phases: **Phase I (registration/training)**, in which captured images are classified as real or fake and only real images are enrolled, and **Phase II (authentication/testing)**, in which a probe image is matched against the registered record.

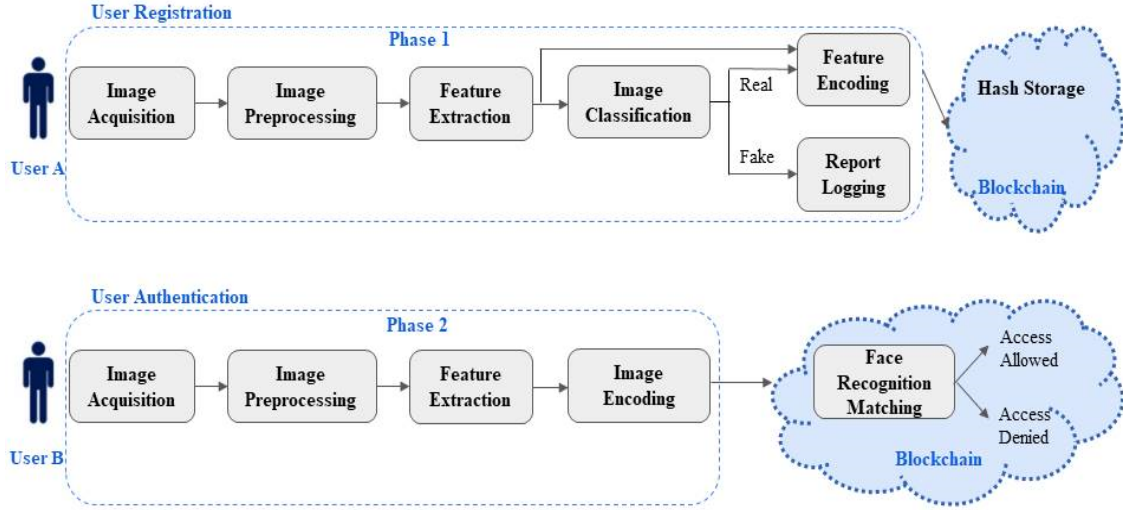


Fig. 1. Block diagram of the proposed secured facial biometric system (Phase I: registration; Phase II: authentication).

We consider the following adversary capabilities: (T1) presenting a deepfake image or video to the sensor (presentation attack); (T2) injecting synthetic media directly into the capture pipeline or the enrolment channel (injection attack); (T3) breaching the template store to steal biometric data; (T4) modifying or replacing a stored template or its metadata; (T5) replaying a previously captured authentication message; and (T6) crafting adversarial perturbations that fool the classifier. We assume that a majority of blockchain validators are honest, that cryptographic primitives are secure, and that the capture device can be authenticated by the client application.

3.2 Phase I: registration

Image acquisition. A sensor captures face images and each user receives a unique identifier (UID). The face region is detected and isolated.

Pre-processing. Noise reduction is applied to highlight distinctive facial features, and the face is cropped and resized to a standard $224 \times 224 \times 3$ representation so that downstream stages receive consistent input:

$$\tilde{x} = \text{Resize}_{224}(\text{Denoise}(\text{Crop}(I))) \quad (1)$$

Data augmentation. A GAN [20] generates additional realistic samples to enlarge the training set, and standard augmentations (rotation, flipping, brightness and contrast jitter, partial occlusion) address variation in pose, angle, illumination and occlusion. Generated images are used only for training and never for enrolment.

Feature extraction. A CNN backbone extracts local texture and structure, and self-attention layers [21] re-weight spatial positions so that the representation concentrates on regions where manipulation artefacts and identity cues are most informative. The attention operation is

$$\text{Attention}(Q, K, V) = \text{softmax}(QK^T / \sqrt{d_k}) V \quad (2)$$

where Q , K and V are the query, key and value projections of the CNN feature map and d_k is the key dimension. The encoder f_θ maps the pre-processed image to a facial feature vector:

$$F_R = f_\theta(\tilde{x}), F_R \in \mathbb{R}^d \quad (3)$$

Real/fake detection. As shown in Fig. 2, the feature vector is passed to an ANN classifier, which outputs the probability that the image is real:

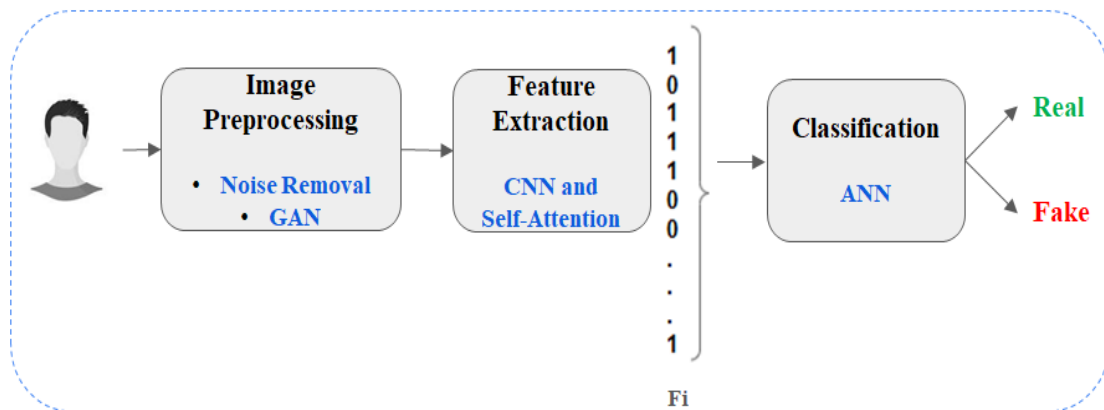


Fig. 2. Real/fake facial image detection framework.

$$p = \sigma(W_2 \cdot \text{ReLU}(W_1 F_R + b_1) + b_2), \text{ label} = \text{real if } p \geq \tau \quad (4)$$

The model is trained with binary cross-entropy over N training images with labels $y_i \in \{0, 1\}$:

$$L = -(1/N) \sum_i [y_i \log p_i + (1 - y_i) \log(1 - p_i)] \quad (5)$$

Real-image storage and fake-event logging. If the image is classified as real, the feature vector is protected with an encoding or encryption function $E(\cdot)$ to obtain $E(F_R)$. A multi-factor embedding is formed by binding $E(F_R)$ to the user's UID and the encrypted record is sent to cloud storage, while its hash commitment is written to the blockchain:

$$C_R = H(E(F_R) \parallel \text{UID} \parallel \text{nonce}) \quad (6)$$

If the image is classified as fake, no template is created and a fake-image event (UID or device identifier, timestamp and an evidence hash) is appended to the blockchain log.

3.3 Phase II: authentication

During authentication (Fig. 3), the probe image of user B is pre-processed and its features F_A are extracted, passed through the same real/fake detector, encrypted as $E(F_A)$, and bound to the claimed UID. Probes classified as fake are rejected and logged. For probes classified as real, the smart contract computes $H(E(F_A))$ and verifies that the registered commitment $H(E(F_R))$ is present on the ledger and unaltered, then evaluates the match.

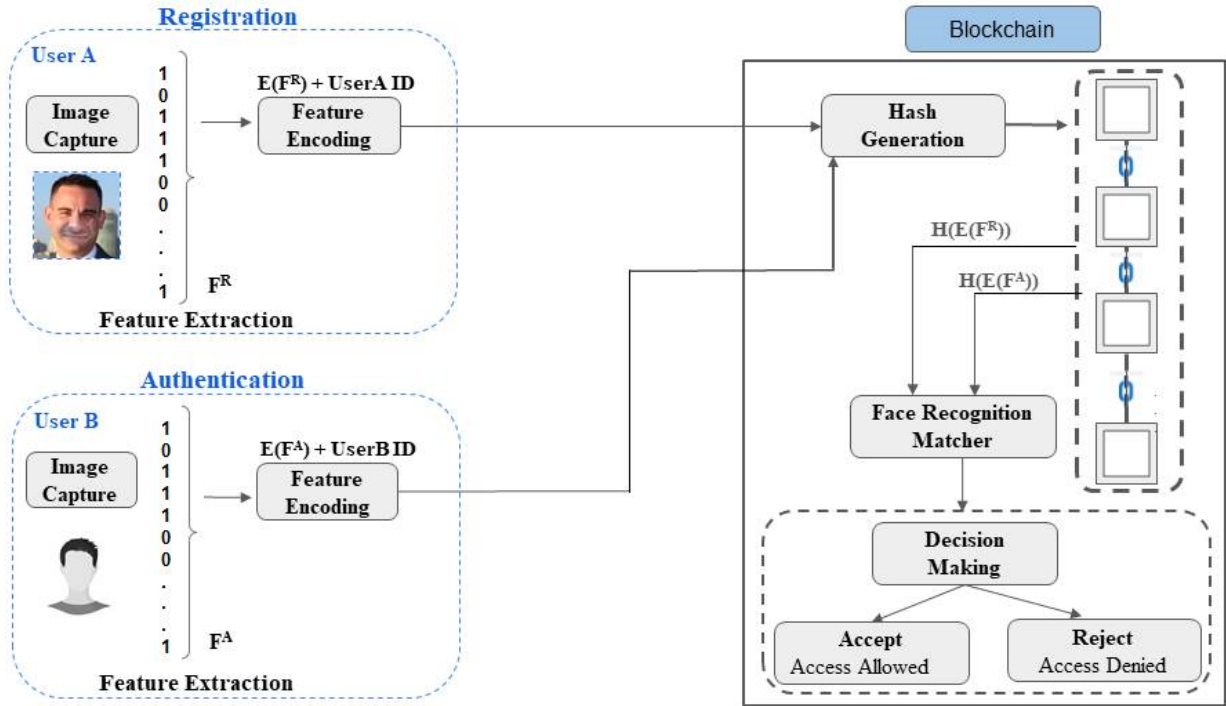


Fig. 3. Authentication process using the blockchain.

Matching in the protected domain. Two captures of the same face never produce bit-identical feature vectors, so comparing hashes for strict equality would reject genuine users. We therefore use the hash commitment as an integrity and provenance anchor, and perform the similarity decision as a threshold test on the cosine similarity between the probe and the registered features:

$$s = (F_A \cdot F_R) / (\|F_A\| \|F_R\|), \text{ accept iff } s \geq \theta \quad (7)$$

To keep F_R and F_A hidden from the verifier, the client produces a zero-knowledge proof π [26] of the statement "I know a feature vector consistent with the on-chain commitment C_R whose similarity to my fresh probe is at least θ ", and the smart contract only verifies π together with a fresh nonce to prevent replay. As an alternative for resource-constrained clients, a fuzzy extractor [33] can derive a stable key from the noisy template, so that a

conventional hash or signature check can be used. Both options are evaluated in Section 4.

Decision making. If verification succeeds the user is granted access; otherwise an integrity-failure message is returned and the attempt is logged on the ledger.

3.4 Blockchain and smart-contract layer

Blockchain stores only commitments, metadata and log entries. Encrypted templates stay in off-chain storage and are referenced from the ledger by hash, which keeps the ledger small, limits exposure of biometric material, and makes deletion of the off-chain template possible when a user withdraws consent. A private permissioned network gives predictable latency and controlled validator membership; the choice of platform (for example Hyperledger Fabric or an Ethereum-compatible private network) is an implementation decision that does not alter the protocol. Table II lists the smart-contract interface.

Table II. Smart-contract interface of the proposed framework.

Function	Caller	Purpose
registerTemplate(uid, C_R, metaHash)	Enrolment service	Stores the commitment and metadata hash of a template classified as real; rejects duplicate UIDs.
logFakeEvent(uid, evidenceHash, ts)	Enrolment / auth. service	Appends an immutable record for an image classified as fake.
verifyAuth(uid, π , nonce)	Client	Verifies the ZKP and nonce freshness against C_R and returns accept or reject.
grantAccess(uid, party, scope)	Data owner	Records explicit, scoped consent for a third party to request verification.
revoke(uid)	Data owner / admin	Marks a commitment as revoked so it can no longer be used for verification.
getAuditTrail(uid)	Auditor	Returns the ordered list of registration, authentication, rejection and consent events.

3.5 Security and privacy analysis

Table III maps the threats of Section 3.1 to the mechanisms that address them. Residual risks are discussed in Section 6.

Table III. Threats and corresponding countermeasures.

Threat	Countermeasure	Component
T1 Presentation attack with deepfake media	Real/fake classification of every probe before matching; failures logged	CNN-attention encoder + ANN
T2 Injection into capture or enrolment	Detection at enrolment; device authentication; signed capture metadata	Client app, smart contract
T3 Template-store breach	Templates encrypted, bound to UID, held off-chain; no raw features on the ledger	Encryption, off-chain store
T4 Template tampering	Hash commitment on an immutable ledger exposes any change	Blockchain, hashing
T5 Replay of authentication data	Fresh nonce bound to each proof	ZKP, smart contract
T6 Adversarial perturbation	Adversarial training and augmentation; robustness measured with FGSM/PGD	Training pipeline

4. EXPERIMENTAL DESIGN

4.1 Datasets

The **Real and Fake Face Detection** dataset published by CIPLAB [34] contains approximately two thousand face images, comprising genuine photographs and expert-manipulated faces graded by difficulty, and is used to train and evaluate the real/fake classifier. **UTKFace** [30] contains more than 20,000 face images annotated with age, gender and ethnicity, and is used to build the identity-verification experiments and to measure demographic fairness. To test cross-generator generalisation, FaceForensics++ [22] and Celeb-DF [23] are used as unseen-domain test sets without further training.

4.2 Implementation

The learning components are implemented in Python 3.8 with Keras and TensorFlow 2.0. Images are resized to $224 \times 224 \times 3$. Templates are encrypted with a standard authenticated symmetric cipher (e.g., AES-256-GCM) and committed with SHA-256. The ZKP circuit is realised with a zk-SNARK toolchain, and the smart contracts are deployed on a private permissioned test network. A dashboard and an inference API expose registration, authentication and audit functions.

4.3 Evaluation protocol

Data are split into training, validation and test sets (70/15/15) stratified by class and with no identity shared between the splits; five-fold cross-validation is used

where the data size permits. The following metrics are reported.

- **Detection:** accuracy, precision, recall, F1-score, ROC-AUC, and the presentation-attack error rates APCER and BPCER, with ACER = (APCER + BPCER)/2.
- **Authentication:** false accept rate (FAR), false reject rate (FRR) and equal error rate (EER) as a function of the threshold θ .
- **Robustness:** performance under FGSM and PGD perturbations, JPEG compression and noise, and on the unseen FaceForensics++ and Celeb-DF domains.
- **Fairness:** differences in FAR, FRR and detection recall across gender, age and ethnicity groups in UTKFace.
- **System cost:** registration and authentication latency, proof generation and verification time, ledger storage per user, and transaction throughput.

4.4 Baselines and ablations

The proposed encoder is compared with VGG16, ResNet50, MobileNet and Xception classifiers and with a CNN without attention, using the same splits. Ablations remove (i) self-attention, (ii) GAN augmentation, (iii) the real/fake gate, (iv) the ZKP step (replaced by exact hash comparison and by the fuzzy-extractor option), and (v) the blockchain layer (replaced by a conventional central

database) to quantify both the security gain and the overhead of each component.

5. RESULTS AND EXPECTED OUTCOMES

Table IV. Calculated values (decimal) derived from the reported metrics and the evaluation protocol.

No.	Quantity	Calculation	Result	Source
1	F1 of [9] from reported P and R	$2 \times 0.937 \times 0.912 / (0.937 + 0.912)$	0.9243	Reported F1 0.929
2	F1 of [17] from reported P and R	$2 \times 0.948 \times 0.938 / (0.948 + 0.938)$	0.9430	Reported F1 0.943
3	Mean accuracy of [3], good quality	$(0.82 + 0.86) / 2$	0.8400	Range 0.82–0.86
4	Accuracy drop of [3], good to low quality	$0.84 - 0.786$	0.0540	Mean vs. 0.786
5	AUC spread of [18]	$0.9974 - 0.8762$	0.1212	Range of AUC
6	AUC spread of [19]	$0.9978 - 0.9000$	0.0978	Range of AUC
7	Accuracy spread of [14]	$0.9882 - 0.9492$	0.0390	Range of accuracy
8	Input size per image	$224 \times 224 \times 3$	150528.0	Eq. (1)
9	Training images (N = 2000)	2000×0.70	1400.0	70/15/15 split
10	Validation images (N = 2000)	2000×0.15	300.0	70/15/15 split
11	Test images (N = 2000)	2000×0.15	300.0	70/15/15 split
12	SHA-256 digest size	$256 \text{ bits} / 8$	32.0 bytes	Eq. (6)
13	Cosine similarity, worked example	$F_A = (0.6, 0.8), F_R = (0.8, 0.6): 0.96 / (1.0 \times 1.0)$	0.9600	Eq. (7), illustrative

6. DISCUSSION

Practical implications. Placing the real/fake decision at enrolment prevents synthetic identities from entering the system, and placing it again before matching blocks presentation attacks. Anchoring commitments on a ledger gives regulators and auditors a verifiable history of every event without disclosing biometric content, which supports accountability obligations under data-protection regimes such as the GDPR [31] and India's Digital Personal Data Protection Act, 2023 [32].

Limitations. First, detection accuracy on one generator does not guarantee accuracy on future ones; continual learning and periodic retraining will be needed. Second, GAN-based augmentation can introduce a domain shift if the synthetic samples differ statistically from real captures. Third, proof generation on low-end mobile devices may be slow, which is why the fuzzy-extractor alternative is retained. Fourth, a ledger cannot prove that the image captured by the camera was genuine at the time of capture, so device attestation and signed capture remain necessary against injection attacks. Fifth, immutability conflicts with the right to erasure if personal data were written on-chain; the design avoids this by storing only commitments and keeping encrypted templates off-chain. Sixth, a single face modality remains vulnerable to high-quality attacks, and fusion with voice or other biometrics [3], [8] is a natural extension.

Ethics and fairness. Biometric systems must be evaluated for demographic parity before deployment. The fairness analysis on UTKFace is therefore part of the protocol, and informed, revocable user consent is enforced through the smart contract.

7. CONCLUSION AND FUTURE WORK

This paper presented a facial biometric security framework that integrates a CNN–self-attention deepfake detector with a blockchain layer using hashing, encryption, ZKPs and smart contracts. The framework screens images at enrolment and at authentication, binds encrypted templates to user identities, records integrity failures immutably, and verifies identity without exposing raw biometric features. A threat model, a security analysis and a complete evaluation protocol were provided.

Future work includes implementing and benchmarking the proof-of-concept on the datasets described, evaluating cross-generator and adversarial robustness, extending the framework to video-based liveness cues and multimodal fusion, applying federated learning to avoid centralised training data, adding explainability for audit purposes, and assessing deployment cost on mobile and edge hardware.

DECLARATIONS

Conflict of interest: The author declares no known competing interests. [Confirm before submission.]

Funding: [State funding sources or "None".]

Data availability: The datasets used are publicly available [30], [34]. Code and smart contracts will be released upon completion of the experiments. [Update as applicable.]

REFERENCES

- [1] "Biometric data: Why is it problematic?" Mondaq. [Online]. Available: <https://www.mondaq.com/india/data-protection/1508150/biometric-data-why-is-it-problematic>
- [2] BioCatch, "BioCatch releases 2024 digital banking fraud trends report in India," press release. [Online]. Available: <https://www.biocatch.com/press->

release/biocatch-releases-2024-digital-banking-fraud-trends-report-in-india

- [3] R. Brown, G. Bendiab, S. Shiaeles and B. Ghita, "A novel multimodal biometric authentication system using machine learning and blockchain," in *Selected Papers from the 12th International Networking Conference (INC 2020)*, Lecture Notes in Networks and Systems, vol. 180. Cham, Switzerland: Springer, 2021. doi: 10.1007/978-3-030-64758-2_3.
- [4] K. Gandhi, P. Kulkarni, T. Shah, P. Chaudhari, M. Narvekar and K. Ghag, "A multimodal framework for deepfake detection," arXiv:2410.03487, 2024.
- [5] A. Kumar, S. Kalumbi, M. K. Gupta and P. RaoMV, "Smart face recognition using IoT and machine learning," *Int. J. Res. Publ. Rev. (IJRPR)*, vol. 2582, p. 7421. [Complete volume, issue, pages and year.]
- [6] S. Singh, R. Ramya, V. Sushma, S. R. Roshini and R. Pavithra, "Facial recognition using machine learning algorithms on Raspberry Pi," in *Proc. 4th Int. Conf. Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT)*, Dec. 2019, pp. 197–202.
- [7] K. Raju, B. Chinna Rao, K. Saikumar and N. Lakshman Pratap, "An optimal hybrid solution to local and global facial recognition through machine learning," in *A Fusion of Artificial Intelligence and Internet of Things for Emerging Cyber Systems*, 2022, pp. 203–226.
- [8] P. N. J. Laila and D. G. Tamilpavai, "FinBTech: Blockchain-based video and voice authentication system for enhanced security in financial transactions utilizing FaceNet512 and Gaussian Mixture Models," arXiv:2310.18668, 2023.
- [9] J. J. Bird and A. Lotfi, "CIFAKE: Image classification and explainable identification of AI-generated synthetic images," *IEEE Access*, vol. 12, pp. 15642–15650, 2024. doi: 10.1109/ACCESS.2024.3356122.
- [10] A. Agarwal and N. Ratha, "Deepfake Catcher: Can a simple fusion be effective and outperform complex DNNs?," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, 2024, pp. 3791–3801. doi: 10.1109/CVPRW63382.2024.00383.
- [11] A. Raza, K. Munir and M. Almutairi, "A novel deep learning approach for deepfake image detection," *Appl. Sci.*, vol. 12, no. 19, Art. 9820, 2022. doi: 10.3390/app12199820.
- [12] A. M. Abukari, J. B. Madavarapu and E. K. Bankas, "A lightweight algorithm for detecting fake multimedia contents on social media," *Earthline J. Math. Sci.*, vol. 14, no. 1, pp. 119–132, 2023. doi: 10.34198/ejms.14124.119132.
- [13] F. Marra, D. Gragnaniello, D. Cozzolino and L. Verdoliva, "Detection of GAN-generated fake images over social networks," in *Proc. IEEE Conf. Multimedia Information Processing and Retrieval (MIPR)*, 2018.
- [14] S. Sugandhi, M. U. A. Ayobkhan, K. K. V. N. Bacanin, V. K. H. Štěpán and T. Pavel, "Deep learning model for deep fake face recognition and detection," *PeerJ Comput. Sci.*, vol. 8, Art. e881, Feb. 2022. doi: 10.7717/peerj-cs.881.
- [15] N. Goel, S. Kaur and R. Bala, "Dual branch convolutional neural network for copy move forgery detection," *IET Image Process.*, vol. 15, no. 3, pp. 656–665, 2021.
- [16] S. Safwat, A. Mahmoud, I. Eldesouky Fattoh and F. Ali, "Hybrid deep learning model based on GAN and RESNET for detecting fake faces," *IEEE Access*, vol. 12, pp. 86391–86402, 2024. doi: 10.1109/ACCESS.2024.3416910.
- [17] M. A. N. Ul Ghani, K. She, M. A. Rauf, M. Alajmi, Y. Y. Ghadi and A. Algarni, "Securing synthetic faces: A GAN-blockchain approach to privacy-enhanced facial recognition," *J. King Saud Univ.–Comput. Inf. Sci.*, vol. 36, no. 4, Art. 102036, 2024. doi: 10.1016/j.jksuci.2024.102036.
- [18] K. Omar, R. H. Sakr and M. F. Alrahmawy, "An ensemble of CNNs with self-attention mechanism for DeepFake video detection," *Neural Comput. Appl.*, vol. 36, pp. 2749–2765, 2024. doi: 10.1007/s00521-023-09196-3.
- [19] P. G. Prathibha, P. S. Tamizharasan, A. Panthakkan, W. Mansoor and H. Al Ahamad, "SARB-DF: A continual learning aided framework for deepfake video detection using self-attention residual block," *IEEE Access*, 2024.
- [20] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 2672–2680.
- [21] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 5998–6008.
- [22] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Nießner, "FaceForensics++: Learning to detect manipulated facial images," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, 2019.
- [23] Y. Li, X. Yang, P. Sun, H. Qi and S. Lyu, "Celeb-DF: A large-scale challenging dataset for DeepFake forensics," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [24] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales and J. Ortega-Garcia, "DeepFakes and beyond: A survey of face manipulation and fake detection," *Inf. Fusion*, vol. 64, pp. 131–148, 2020.
- [25] L. Verdoliva, "Media forensics and DeepFakes: An overview," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 5, pp. 910–932, 2020.

- [26] S. Goldwasser, S. Micali and C. Rackoff, "The knowledge complexity of interactive proof systems," *SIAM J. Comput.*, vol. 18, no. 1, pp. 186–208, 1989.
- [27] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008. [Online]. Available: <https://bitcoin.org/bitcoin.pdf>
- [28] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [29] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017, pp. 618–626.
- [30] Z. Zhang, Y. Song and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5810–5818.
- [31] Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation), Off. J. Eur. Union, L 119, 2016.
- [32] Government of India, "The Digital Personal Data Protection Act, 2023," 2023.
- [33] Y. Dodis, R. Ostrovsky, L. Reyzin and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," *SIAM J. Comput.*, vol. 38, no. 1, pp. 97–139, 2008.
- [34] CIPLAB, "Real and Fake Face Detection," Kaggle dataset. [Online]. Available: <https://www.kaggle.com/datasets/ciplab/real-and-fake-face-detection>