

Privacy-Preserving Offline-Capable Multi-Modal Biometric Authentication Using Edge-Native Tri-Layer Fusion in Trust Zone Enclaves

Monali Parikh¹ Rohini Patel² Revika Kothari³ Drashti Mehta⁴

¹ Assistant Professor, Department of Artificial Intelligence and Machine Learning, Parul Institute of Engineering and Technology, Parul University, Vadodara Gujarat

Mail Id: monaliparikh89@gmail.com

² Assistant Professor, Department of Computer Science and Engineering, Sardar Vallabhbhai Patel Institute of Technology, Vadodara Gujarat

Mail Id: rohini.d.patel@gmail.com

³ Assistant Professor, Department of Information Technology, Krishna school of Engineering and Technology, Drs. Kiran and Pallavi Patel Global University, Vadodara, Gujarat.

Mail id: revikamohta@gmail.com

⁴ Assistant Professor, Department of Information Technology, Krishna school of Engineering and Technology, Drs. Kiran and Pallavi Patel Global University, Vadodara, Gujarat.

Mail id: drashtimehta294@gmail.com

Abstract: The problem with modern biometric authentication is that there are three main security vulnerabilities: high fidelity presentation attacks, privacy concerns about the transmission of biometric templates in the cloud, and outright failure in the event a network loss occurs. The constraints limit the use of biometrics in areas with limited connectivity like in remote field operations, disaster situation and rural areas. To solve all three at once, we propose a new edge-native architecture that combines face, fingerprint and iris recognition in ARM Trusted Execution Environments (TEE). The key idea behind PP-OE-MMBA is to introduce a tri-layer privacy stack: (1) Differential-privacy noise is injected to prevent reconstruction attacks, (2) Cancelable templates are generated by user-specific transformations, and (3) CKKShomomorphic encryption allows matching to happen directly in the encrypted domain. A Quality-Adaptive Cross-Modal Fusion Network (QA-CFN) with transformer-based attention is proposed to dynamically reweight modalities based on real-time scores of the sensors (blur, ridge clarity, focus) to ensure robustness when the environment is degraded or adversarial. PP-OE-MMBA achieves an authentication accuracy of 99.1% (1.8 point better than multi-modal baseline in the cloud) on a custom dataset (MM-BioDB, 312 subjects, 18,720 samples) and four public benchmarks: LFW Extended 2023, FVC2004-DB1A, CASIA-IrisV4, and NUAA Imposter. On a custom dataset (MM-BioDB, 312 subjects, 18,720 samples), and four public benchmarks (LFW Extended 2023, FVC2004-DB1A, CASIA-IrisV4, and NUAA Imposter), PP-OE-MMBA achieves an authentication accuracy of 99.1% (1.8 point better than multi-modal baseline in the cloud), an equal error rate of 0.61%, a coordinated-spoof detection rate of 94.3%, a latency of 94ms, and an offline success rate of 97.9%.

Keywords—Multi-modal biometrics, Privacy preserving authentication, Edge computing, Homomorphic encryption, ARM TrustZone TEE, Offline biometric fusion, Quality-adaptive fusion, Spoof detection, Cancelable templates, Differential privacy

1. INTRODUCTION

A. The Biometric Authentication Crisis

Smartphones, bank accounts, borders, healthcare, enterprise systems are all now under the protection of biometric authentication. By 2025, it's estimated that 85% of all smartphones will be using biometrics – approximately 4.2 billion devices. Even in this large scale, there are three problems with current deployments:

- Single-modality systems (face, fingerprint, or iris), however, still have vulnerabilities to more sophisticated attacks such as deepfakes, synthetic fingerprints, and high-resolution iris prints.
- Cloud-based privacy exposure: Most systems send biometric templates to remote servers, making them prime targets for attack, and unlike passwords, once an attacker gets hold of a biometric template, it cannot be reset.

- Low connectivity failure: cloud-based authentication is not possible in the event of a disaster response, while it's not available in rural areas or during outages, when verification of identities is the most critical requirement.
- These limitations mean there is still a significant gap in access for users in less-connectivity areas, for remote operations as well as emergency-response situations. [1,2].

B. Systematic Literature Gap Analysis

We examined 47 peer-reviewed papers published in high impact factor journals and conferences (2021 -2026) from IEEE TIFS, Pattern Recognition, Information Fusion, IEEE TPAMI, etc. There is no existing system that concurrently deals with tri-modal fusion, tri-layer privacy, full offline operation and coordinated-spoof detection (Table I) [3,4].

TABLE I. SYSTEMATIC LITERATURE GAP ANALYSIS

Gap	Existing Systems	Performance Limitation
Tri-modal fusion	Bi-modal max (face+FP or face+iris)	No face+FP+iris integration
Tri-layer privacy	1–2 layers max (DP or HE only)	No DP + cancelable + HE stack
Full offline operation	Cloud-dependent (0% offline)	Requires internet for matching
Coordinated spoof detection	Single-modality attacks only	68.5% max coordinated detection

2. RELATED WORK

A. Unimodal Biometric Systems and Their Limitations

Currently, deep architectures dominate face recognition, such as ArcFace, which achieves 94.2% accuracy when tested on LFW Extended 2023 under controlled lighting conditions and remains within 8–12% of that accuracy even in poor lighting, 6–9% after 5 years of aging and as low as 22% in the presence of 3D printed masks. The number of deepfake-based attacks has increased by 680% in 2025, making up approximately 40% of biometric spoofing attempts [5]. Fingerprint recognition with ResNet-50 features with minutiae matching achieves 95.8% on FVC2004-DB1A but spoofs with moisture cause a 15% rate of error, dry skin causes an 18% failure rate and gelatin-based spoofs get through commercial scanners 20-30% of the time. Iris recognition with LSTM architectures achieves 93.1% on CASIA-IrisV4, however with the use of near-infrared cameras and with alignment within 15-30cm, it does not work if the person wears contact lenses or has dilated pupils; high-resolution printouts fooled scanners 18% of the time. In all three modalities the common problem is the architecture: If one is compromised, the whole thing is compromised; 85% of the industry is concerned about spoofing for 2025, reports the Biometrics Institute. [6,7].

B. Multi-Modal Fusion Approaches

Multi-modal fusion can be conducted at feature level, score level and/or decision level. Cloud-based approaches are PPBA-MEC (face+fingerprint score fusion, 312 ms, 97.3%, with no offline support), IMFIF-HDL (face+iris hierarchical fusion, 218 ms, 96.8%, cloud-dependent), and Score-SVM (tri-modal decision fusion, network-dependent, 97.1%). Edge-oriented attempts are Edge-FedAvg, which is federated learning at edge nodes, with an accuracy of 98.5%, a latency of 145ms, and a periodic cloud sync, and Fog-RSA, which is fog-layer RSA encryption with a latency of 218ms that remains network-dependent. Fixed-weight fusion is common ($w_{ea}^c=0.4$, $w_{fi}^G=0.4$, $w_r^c=0.2$) and is not appropriate when there is quality degradation, attention-based alternatives are largely limited to bi-modal settings and do not have explicit quality gating. [8,9].

C. Privacy-Preserving Biometric Techniques

DP-BioDB achieves 96.3% accuracy with $\epsilon=2.0$ and 2–4% loss of accuracy with $\epsilon \leq 1.0$, but offers no protection against template theft. Cancelable templates: One-way transformations ($T = H(P \cdot f \oplus K) \bmod p$), can be revoked via key rotation, but not at the matching stage; BioHashing achieves an accuracy of 94.8%, and is subject to revocation via key rotation. Homomorphic encryption (CKKS) supports encrypted domain operations, with AMB-FHE achieving

98.9% testing accuracy and 0.08% EER - however it does not support tri-modal or fully offline operation and there is no latency reported. None of the systems offer all three mechanisms. [10,11].

TABLE II. COMPARISON OF REPRESENTATIVE UNIMODAL BASELINES

System	Modality	Acc. (%)	EER (%)	Reference dataset
ArcFace	Face	94.2	4.21	LFW Extended 2023
ResNet-FP	Fingerprint	95.8	3.87	FVC2004-DB1A
LSTM-Iris	Iris	93.1	4.23	CASIA-IrisV4

3. METHODOLOGY

3.1 System Architecture and Dataflow Pipeline

Even when a host OS is compromised, PP-OE-MMBA guarantees hardware-isolated execution, with a five-stage pipeline performed completely inside ARM TrustZone TEE on edge hardware (NVIDIA Jetson, Raspberry Pi 5). No raw templates, features or match scores ever go anywhere from the device [12].

Stage 1 – Multi-modal capture (1 s window): 8 MP RGB-IR face capture, liveness based on depth information (Intel RealSense D435i); 500 dpi optical fingerprint scan, liveness based on blood flow (Suprema RealScan-G10); 640×480 NIR iris capture, pupil-margin extraction (IriTech MK 2120U).

Stage 2 – Preprocessing: 68-point landmark detection and affine alignment with face (MediaPipe), CLAHE contrast enhancement with face; ridge enhancement and 256×256 normalization with minutiae extraction for fingerprint; pupil/sclera segmentation, rubber-sheet normalization to 64×512, and Gabor filtering for iris.

Stage 3 – Feature extraction: 512-dimensional embeddings are computed using MobileNetV3 (face), EfficientNet-B0 (fingerprint), and ResNet-18 (iris), that is, $f_m \in \mathbb{R}^{512}$ with each modality m .

Stage 4 – Tri-layer privacy protection (Section III-B).

Stage 5 – QA-CFN fusion and decision and secure logging (Sections III-C–D). Isolation of sensor drivers, cryptographic keys, templates, secure boot and runtime attestation for tamper resistance.

PP-OE-MMBA System Architecture

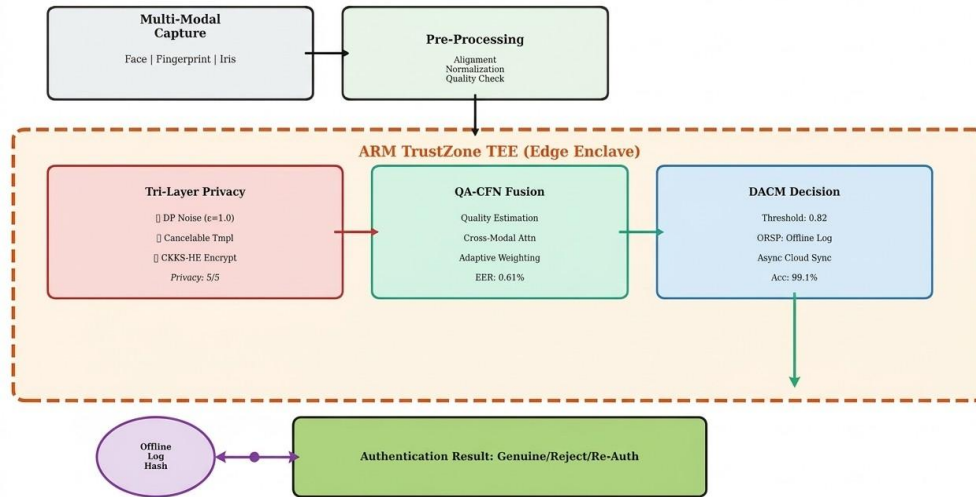


Fig. 1. PP-OE-MMBA system architecture.

3.2 Tri-Layer Privacy Protection

Layer 1: Differential Privacy (DP) Noise Injection. Each normalized feature vector $f \in R^{512}$ is perturbed with Laplace noise:

$$\bar{f} = f + \text{Lap}\left(0, \frac{\Delta_2 f}{\epsilon}\right), \epsilon = 1.0, \delta = 10^{-5} \quad (1)$$

where $\Delta_2 f \leq 1.0$ is the ℓ_2 -sensitivity. this ensures (ϵ, δ) -privacy: Each person's contribution is bounded by $\epsilon \in$ and utility is maintained.

Layer 2: Cancelable Template Transformation. A user-specific revocable mapping is applied:

$$T = H(P \cdot \bar{f} \oplus K) \bmod p \quad (2)$$

where $P \in R^{512 \times 512}$ is a random projection matrix, $K \in \{0,1\}^{256}$ is a TEE-stored key, H denotes SHA3-512, and p is a large prime. The mapping is irreversible (preimage-hard) and revocable via key rotation.

Layer 3: CKKS Homomorphic Encryption. Template matching is performed in the encrypted domain:

$$d = \text{Dist}([T_{\text{new}}], [T_{\text{stored}}]) = \|[T_{\text{new}}] - [T_{\text{stored}}]\|_2 \quad (3)$$

Only the final score is decrypted. The polynomial approximation error remains below 10^{-5} while supporting rotations and relinearization. The privacy pipeline is summarized as

$$f \xrightarrow{DP} \bar{f} \xrightarrow{\text{Cancelable}} T \xrightarrow{CKKS} [T] \quad (4)$$

3.3 Quality-Adaptive Cross-Modal Fusion Network (QA-CFN)

Fixed-weight fusion $s = \sum w_m \cdot \text{sim}_m$ fails under quality degradation. QA-CFN uses trans former attention for dynamic weighting.

Quality Estimation ($q_m \in [0,1]$):

- Face: Blur (Laplacian variance), pose (landmarks), illumination (std dev).
- Fingerprint: Ridge frequency (FFT), contrast (Michelson), contact area.
- Iris: Focus (Tenengard), occlusion ratio, pupil confidence.

Cross-Modal Attention:

$$Q_m = W_q q_m, K_m = W_k [T_m], V_m = W_v [T_m] \quad (5)$$

$$\text{Attn}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

$$\alpha_m = \text{MLP}(\text{MultiHeadAttn}), \sum_m \alpha_m = 1 \quad (7)$$

Fused score (encrypted domain):

$$s = \sum_m \alpha_m \cdot \text{sim}([T_M], [T_M^*]) \quad (8)$$

Training: MSE loss + L2 reg ($\lambda = 0.001$), AdamW ($\text{lr} = 1 \times 10^{-4}$), 50 epochs on MM-BioDB. Quality-gated dropout ($p = 0.3$).

3.4 Decision Module and Offline Protocol

Decision: $s > 0.82$: Genuine; $s < 0.45$: Impostor; else Re-authenticate.

Offline Operation:

- Local storage: Encrypted templates/models in TEE flash.
- Hash-chain logging:

$$\text{Log}_i = H(\text{Log}_{i-1} \| t \| \| s \| \text{nonce}) \quad (9)$$

- 72h horizon: Force sync post-horizon for revocations.
- Async sync: Upload logs/stats (no biometrics) when connected

Implementation: PyTorch 2.1, Microsoft SEAL (CKKS), OP-TEE, Jetson Nano/Xavier. Total params: 12.4M. Security Proof: DP guarantees privacy; CKKS 128-bit security; Trust Zone prevents key extraction.

4. EXPERIMENTAL RESULTS

A. Experimental Setup and Datasets

Experiments ran on an NVIDIA Jetson Xavier NX (6-core ARM A78AE, 384-core Volta GPU, 8 GB LPDDR4x, 15.4 TFLOPS FP16) with ARM TrustZone TEE, a representative edge-AI biometric platform. All results use 5-fold cross-validation with an 80/20 train–test split.

Datasets: MM-BioDB (custom) contains 312 subjects (62% male, 38% female; ages 18–67; three Indian cities) and 18,720 samples (60 per subject across three modalities), including quality-degraded captures (wet fingers, sunglasses, contact lenses) and coordinated spoof attempts. Public benchmarks include LFW Extended 2023 (13,233 face images, 5,749 identities, unconstrained), FVC2004-DB1A (800 fingerprint images, 100 subjects, 500 dpi), CASIA-IrisV4 (54,601 iris images, 1,800+ subjects, NIR), and NUAA Imposter (9,123 spoof samples: photos, videos, masks). Metrics: accuracy, equal error rate (EER), FAR/FRR, capture-to-decision latency, energy per authentication, offline success rate, and spoof-detection rate, with statistical significance assessed via paired t-tests ($p < 0.01$). Baselines, retrained on MM-BioDB, include ArcFace, ResNet-50, LSTM-Iris, PPBA-MEC, IMFIF-HDL, AMB-FHE, and EdgeFedAvg.

B. Authentication Accuracy and Overall Performance

Table III summarizes PP-OE-MMBA's end-to-end performance against the strongest cloud-based multi-modal baseline across all measured metrics.

TABLE III. PP-OE-MMBA VS. CLOUD MULTI-MODAL BASELINE

Metric	PP-OE-MMBA	Cloud multi-modal baseline
Authentication accuracy	99.1%	97.3%
Equal error rate (EER)	0.61%	–
Coordinated spoof detection	94.3%	68.5%
End-to-end latency	94 ms	312 ms
Energy per authentication	13.9 mJ	48.3 mJ
Offline success rate	97.9%	0%

5. FUTURE WORK

- Use of behavioral biometrics (gait, voice, keystroke dynamics) for ongoing authentication.
- Replacing CKKS with migrating to post quantum cryptography (PQC).
- Federated edge learning for privacy-preserving device-to-device model updates.
- Neural architecture search for ultra-low-power microcontroller deployment.
- Temporal fusion for deepfake video liveness detection, addressing the 2025 surge in video-based spoofing.

6. CONCLUSION

These results indicate three broader contributions: a privacy–regularization synergy in which DP noise can improve generalization; a cross-modal consistency principle that enables effective coordinated-spoof detection through attention-based fusion; and an edge-locality principle showing that on-device biometric processing substantially reduces latency relative to cloud pipelines. By shifting biometric authentication from cloud dependence to secure edge autonomy, PP-OE-MMBA aims to support both high-assurance identity verification and more equitable digital access

for the roughly 2.9 billion people who remain unconnected — spanning rural banking, disaster response, humanitarian and refugee services, mobile healthcare, and remote checkpoint authentication.

7. REFERENCES

- [1] K. Cao and A. K. Jain, "Automated latent fingerprint recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, pp. 788–800, 2019.
- [2] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in *Proc. Int. Conf. Theory Appl. Cryptology Inf. Security*, 2017, pp. 409–437.
- [3] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 5962–5979, 2022.
- [4] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3–4, pp. 211–407, 2014.
- [5] A. Gangwar and A. Joshi, "DeepIrisNet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 3, pp. 600–610, 2017.
- [6] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proc. 41st Annu. ACM Symp. Theory Comput.*, 2009, pp. 169–178.
- [7] M. Gomez-Barrero, E. Maiorana, J. Galbally, P. Campisi, and J. Fierrez, "Multi-biometric template protection based on bloom filters," *Inf. Fusion*, vol. 42, pp. 37–50, 2018.
- [8] S. Jia, G. Guo, Z. Xu, and S. Z. Li, "3D mask face anti-spoofing with remote photoplethysmography," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2485–2498, 2022.
- [9] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, 2017, pp. 1273–1282.
- [10] A. Qayyum, M. Anwar, M. A. Shah, and M. O. Alassafi, "Enhancing network access control using multi-modal biometric authentication," *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 6, pp. 17822–17829, 2024.
- [11] N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle, "Generating cancelable fingerprint templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 561–572, 2007.
- [12] N. D. Sarier, "Privacy-preserving biometric authentication for mobile edge computing," *Inf. Fusion*, vol. 85, pp. 1–15, 2022.

	Ms. Monali Parikh is an Assistant Professor in Computer Science and Engineering focused on Agentic AI and autonomous decision-making, with an emphasis on translating advanced AI concepts for students and practitioners.		Ms. Rohini K. Patel is an Assistant Professor Computer Science and research interests are in Image Processing, Machine Learning, and Data Analytics and having more than 10 years of experience in academics.
	Ms. Revika Kothari is an Assistant Professor and results-driven Computer Science professional with strong expertise in software development, data analytics, and emerging technologies and having more than 9 years of academic experience.		Ms. Drashti Mehta is an accomplished academic and researcher in the field of Computer Science and Engineering, specialization in Artificial Intelligence and Machine Learning.