



Research Paper

Explainable ArcFace-Based Deep Metric Learning for Multimodal Biometric Recognition and Authentication Systems

Ram Mohan Babu Srividya^{1*}, Arthanari Latha²

1. School of Computational Engineering, Takshashila University, Tamil Nadu, India.

2. School of Core Engineering, Takshashila University, Tamil Nadu, India.

Citation Srividya RMB, Latha A. Explainable ArcFace-Based Deep Metric Learning for Multimodal Biometric Recognition and Authentication Systems. *International Journal of Medical Toxicology and Forensic Medicine*. 2026; 16:E53370.

10.22037/ijmtfm.v16.53370

Article info:

Received: 06 September, 2026

Accepted: 22 September, 2026

Published: 13 October, 2026

Keywords:

Multimodal biometric recognition, EfficientNet-B0, Computer vision, Explainable artificial intelligence

ABSTRACT

Background: Biometric recognition offers a reliable solution for recognizing individuals based on unique physiological traits such as fingerprints and iris patterns. Fingerprint-based systems are widely employed because of their simplicity and distinct ridge structures, but they are generally affected by partial, worn, and noisy captures. Iris recognition provides higher accuracy and stability because iris patterns remain largely unchanged over an individual's lifetime. This study presents the Fingerprint-Iris Feature Fusion with Deep Metric Learning (FIF-DML) approach for multimodal biometric recognition and authentication.

Methods: The proposed model learns features using two parallel EfficientNet-B0 networks that independently extract deep representations from fingerprint and iris inputs. These extracted features are then fused at the feature level to form a unified representation. In addition, ArcFace-based metric learning maps the fused features into a compact embedding space, enhancing inter-class separability and intra-class compactness. In the identification stage, a query fingerprint-iris pair is converted into a 128-dimensional embedding and matched with gallery embeddings using cosine similarity to predict the identity. Finally, Grad-CAM visualizes the regions that most influence the model's final decision.

Results: We conduct extensive experimental analysis of the proposed FIF-DML model on the Multimodal Iris Fingerprint Biometric dataset, achieving 97.33% precision and demonstrating superior performance compared to existing methods.

Conclusion: The results confirm that the proposed FIF-DML model provides a reliable and effective solution for secure biometric authentication.

* Corresponding Author:

R. Srividya, PhD

School of Computational Engineering, Takshashila University, Tamil Nadu, India.

E-mail: srividya.r@takshashilauniv.ac.in



Copyright © 2026 The Author(s);

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0);

<https://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

B iometric recognition systems provide the most effective solution for current user identification and verification [1]. Biometrics determine an individual's identity through semi- or fully automated methods, based on behavioral and physiological traits such as signatures, voice, gestures, fingerprints, palmprints, faces, irises, and others. Such recognition supports both the identification and verification of a person using their unique behavioral pattern or physiological traits [2]. Among these physiological biometric modalities, fingerprint and iris biometric recognition have gained increasing attention because of their unique characteristics and high discriminative power, which support differentiating the person from others. Each biometric technology offers unique strengths and faces some challenges; also, relying on a single biometric trait is not sufficient for all verification and identification applications [3].

This study presents a concise review of previous research on biometric recognition and identifies research gaps. Hu [7] developed a strong DL-based multimodal authentication model that employs the iris and fingerprint methods for personal digital authentication. Binu [8] presented an innovative index phase fusion model for bimodal biometric verification using fingerprint and iris features integrated with improved locality-sensitive hashing for rapid and precise retrieval, while ensuring robustness and lower computational cost. Mohammed et al. [9] proposed a new multimodal biometric authentication system that combines DL methods with zero-trust framework principles to improve network access control. Kaur and Kant [10] developed an effective architecture for identifying spoofing attacks by employing biometric identification techniques, namely iris, fingerprint, and facial identification. In [11], an adaptable multi-biometric technique is developed that dynamically incorporates multiple methods, specifically iris and fingerprint. Nivedita et al. [12] developed an innovative architecture for multimodal biometric authentication using iris, fingerprint, and face images, leveraging hybrid DL methods and meta-heuristic optimization.

Despite recent progress in biometric recognition, several challenges remain in multimodal fingerprint-iris systems. Existing methodologies often rely on simple fusion strategies that cannot fully exploit the complementary interactions among biometric modalities. However, many approaches are affected by noise and sensor variations, reducing their robustness in practical scenarios. Limited use of DL methodologies also affects feature discrimination,

while the absence of XAI reduces model transparency and interpretability, thereby affecting trustworthiness. These challenges motivate the development of the proposed FIF-DML framework, which integrates multimodal feature fusion, deep metric learning, and explainable analysis for reliable and interpretable biometric authentication

Such limitations can be overcome by applying a multi-model biometric system that combines information from multiple data sources to improve accuracy and recognition, and performs better than a single-biometric system. Therefore, this study introduces a Fingerprint-Iris Feature Fusion with Deep Metric Learning (FIF-DML) approach for multimodal biometric recognition and authentication. The core contributions of this research are as follows:

- Develop a dual-stream feature learning architecture using two parallel EfficientNet-B0 networks to independently extract deep and discriminative representations from fingerprint and iris images.
- Introduce a feature-level fusion approach to combine fingerprint and iris embeddings into a unified representation, enabling richer biometric feature encoding for improved identity discrimination.
- Employ ArcFace-based deep metric learning to learn fused features into a highly discriminative embedding space, improving inter-class separability and intra-class compactness for robust authentication.
- Apply a cosine similarity-based identification and verification framework using 128-dimensional embeddings.
- Integrate Grad-CAM-based visual explainability to interpret model decisions and highlight salient regions contributing to biometric recognition, improving model transparency and trustworthiness.

Materials and Methods

Cell Line and Culture Conditions

The complete process of the proposed framework is shown in Figure 1. Initially, fingerprint images are enhanced using CLAHE, normalization, and resizing, while iris images undergo median filtering (MF), pupil detection, iris localization, mask generation, and extraction to obtain discriminative iris regions. The processed fingerprint and iris images are then passed through two parallel EfficientNet-B0 networks to learn deep feature representations. These features are fused and projected into a compact 128-dimensional

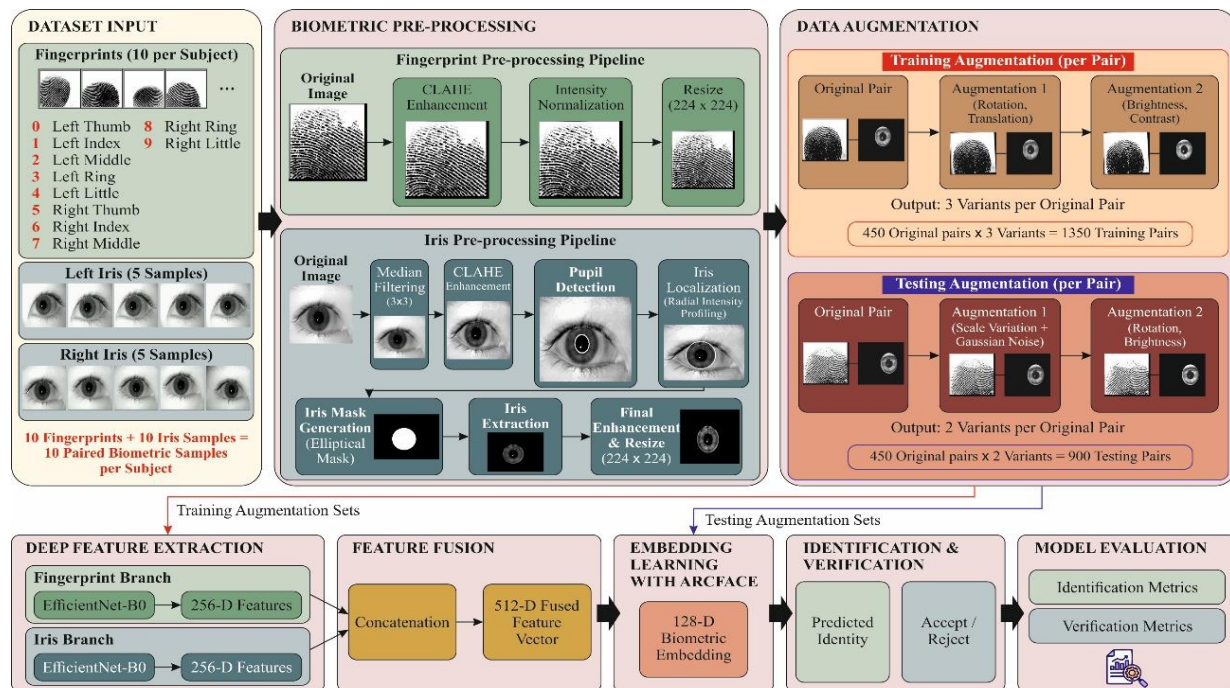

International Journal of
Medical Toxicology & Forensic Medicine

Figure 1. Complete process of the proposed framework.

embedding space using ArcFace-based metric learning. Finally, cosine similarity is used for identification and verification, allowing the system to predict user identity or make an accept/reject authentication decision based on a predefined similarity threshold.

Dataset Used

The proposed FIF-DML model was assessed using the Multimodal Iris Fingerprint Biometric dataset [13]. The dataset contains biometric data collected from 45 people and includes fingerprint and iris images for identification and verification. Figure 2 shows samples of left and right fingerprints and iris images from different individuals.

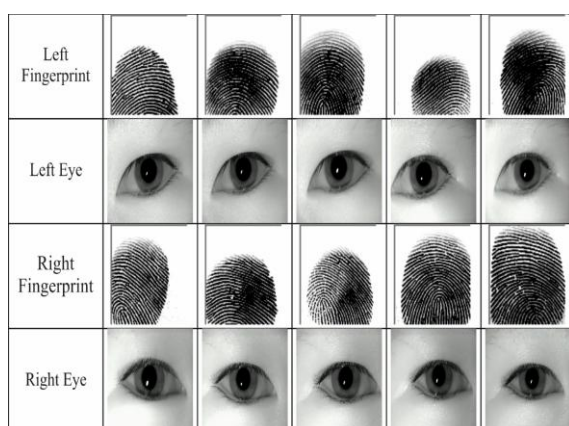

International Journal of
Medical Toxicology & Forensic Medicine

Figure 2. Samples of left and right fingerprints and iris images from different individuals.

Biometric Preprocessing and Data Augmentation

Biometric preprocessing involves enhancing fingerprint images using CLAHE, normalization, and resizing, while iris images are processed using median filtering, CLAHE, pupil detection, iris localization, masking, and resizing. The iris boundary is located from intensity changes; an elliptical mask then extracts the iris, and the image is enhanced and resized to 224x224 pixels, as shown in Figure 3. Rotation, translation, brightness, and contrast variations expand 450 original fingerprint-iris pairs to 1,350 training samples, while scale variation, Gaussian noise, and brightness variations expand 450 original fingerprint-iris pairs to 900 testing samples.

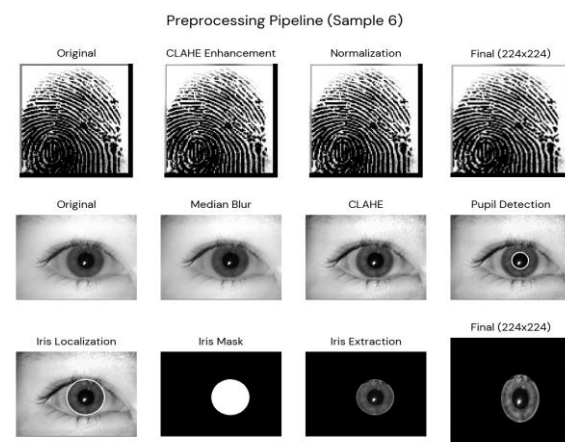

International Journal of
Medical Toxicology & Forensic Medicine

Figure 3. Preprocessing pipeline of fingerprint and iris biometric recognition.

rotation, and brightness variation expand them to 900 testing instances.

Deep Feature Extraction

Deep feature extraction is performed utilizing two parallel EfficientNet-B0 networks to independently learn fingerprint and iris representations.

Fingerprint Branch

The proposed method uses EfficientNet-B0 as the feature-extractor backbone [14]. The structure comprises MBConv blocks combined with Squeeze-and-Excitation (SE) modules. With the input image $I \in \mathbb{R}^{224 \times 224 \times 3}$, the EfficientNet-B0 backbone produces a 1280-dimensional feature embedding:

$$F_{1280} = E(I) \quad (1)$$

Then, the extracted feature vectors are fed into a fully connected (FC) layer to process a compact 256-dimensional feature embedding:

$$F_{256} = f(WF_{1280} + b) \quad (2)$$

Iris Branch

The input image is resized to $224 \times 224 \times 3$ pixels for EfficientNet-B0 feature extraction. ImageNet-pretrained EfficientNet-B0 extracts 1,280-dimensional

features, which are then passed through a fully connected layer to obtain a compact 256-dimensional representation.

Multimodal Feature Fusion

The architecture employs the feature-level fusion approach to integrate data from the biometric techniques. In the feature extraction stage, the fingerprint branch generates a 256-dimensional feature vector, while the iris branch produces a 256-dimensional feature vector.

$$F_{fused} = [F_{fp} ; F_{ir}] \quad (3)$$

The concatenation of these vectors leads to a unified representation:

$$F_{fused} \in \mathbb{R}^{512} \quad (4)$$

This fused representation correlates data from the fingerprint and iris methods in a simple and straightforward manner.

Embedding Learning with ArcFace

The embedding learning model transforms the fused feature representation into a compact, distinctive biometric embedding suitable for identity verification (Figure 4).

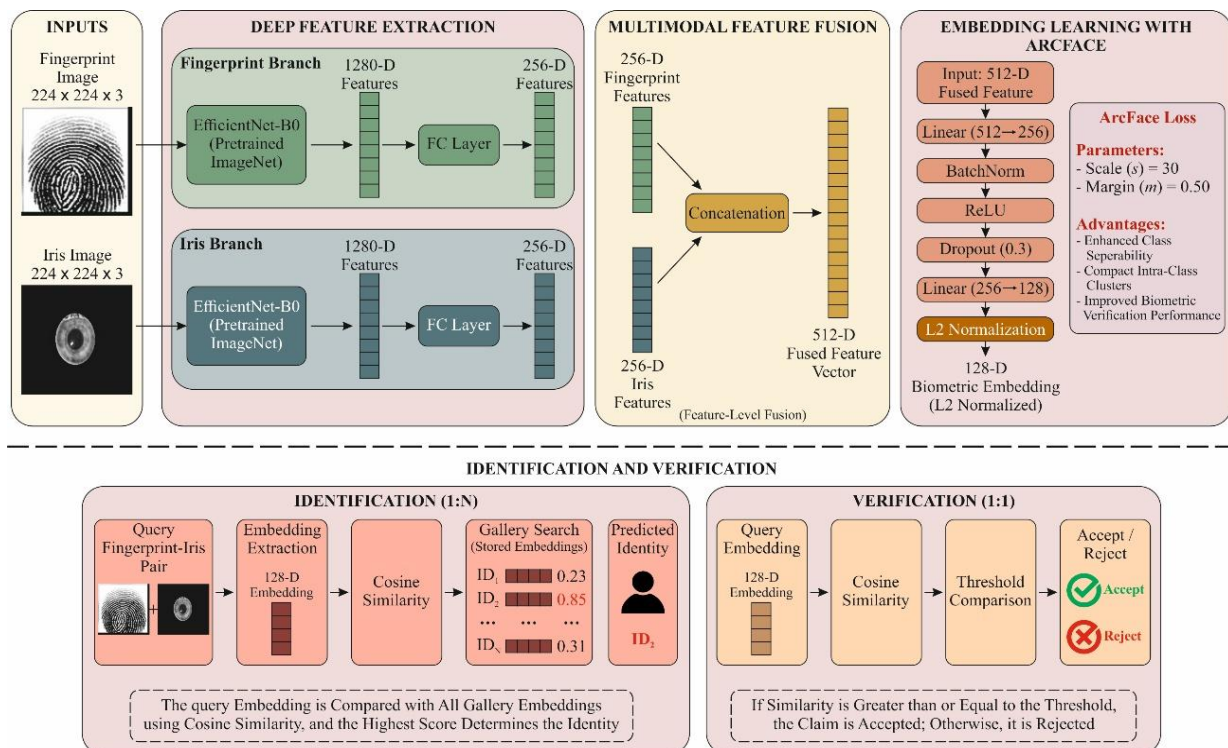


Figure 4. Fingerprint-Iris Fusion Using ArcFace.

The fused feature vector is a 512-dimensional vector. The linear layer projects the higher-dimensional fused feature vector into an intermediate representation as given below,

$$h_1 = W_1 F + b_1 \quad (5)$$

L2 normalization scales the feature vector while maintaining the direction,

$$z = \frac{h_2}{\|h_2\|_2} \quad (6)$$

ArcFace Loss improves feature discrimination by introducing an angular margin between different classes [15].

Identification and Verification

Identification matches a query biometric sample against a database to find the closest identity. Verification checks whether the claimed identity is genuine by comparing similarity scores with a threshold. During identification, the trained fusion network converts the fingerprint–iris input into a compact 128-dimensional embedding that represents the biometric features. Cosine similarity compares this embedding with the stored gallery embeddings, and the identity with the highest similarity score is selected as the final match. During verification, the query fingerprint–iris pair is converted into a 128-dimensional embedding and compared with the claimed identity template using cosine similarity. The resulting score is checked against a predefined threshold to accept or reject the claimed identity.

Explainability Assessment

Grad-CAM highlights the significant regions in an input that influence a model's prediction by generating a class-specific heatmap. It calculates the significance of each feature map by taking the gradient of the target-class score with respect to the feature map activations. This can be expressed as:

$$L_{Grad-CAM}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right) \quad (7)$$

The weighted feature maps are integrated and passed through ReLU to retain only positive influences. The resulting heatmap highlights the most relevant regions that support the model's decision for the class c .

Grad-CAM-based explainability report of the proposed model using fingerprint and iris samples. Grad-CAM visualization highlights the most influential region in both biometric patterns. Attention maps further improve interpretability by showing the region the model focuses on. This makes the system more transparent and trustworthy, ensuring reliable decision-making in the identification and verification process (Figure 5).

Results

The proposed FIF-DML model was evaluated using the Multimodal Iris Fingerprint Biometric dataset. ArcFace loss is employed with scale = 30, margin = 0.5, and dropout = 0.3. For identification, 1: N cosine-similarity matching selects the highest similarity, while verification uses 1:1 cosine similarity with an EER threshold of 0.2286.

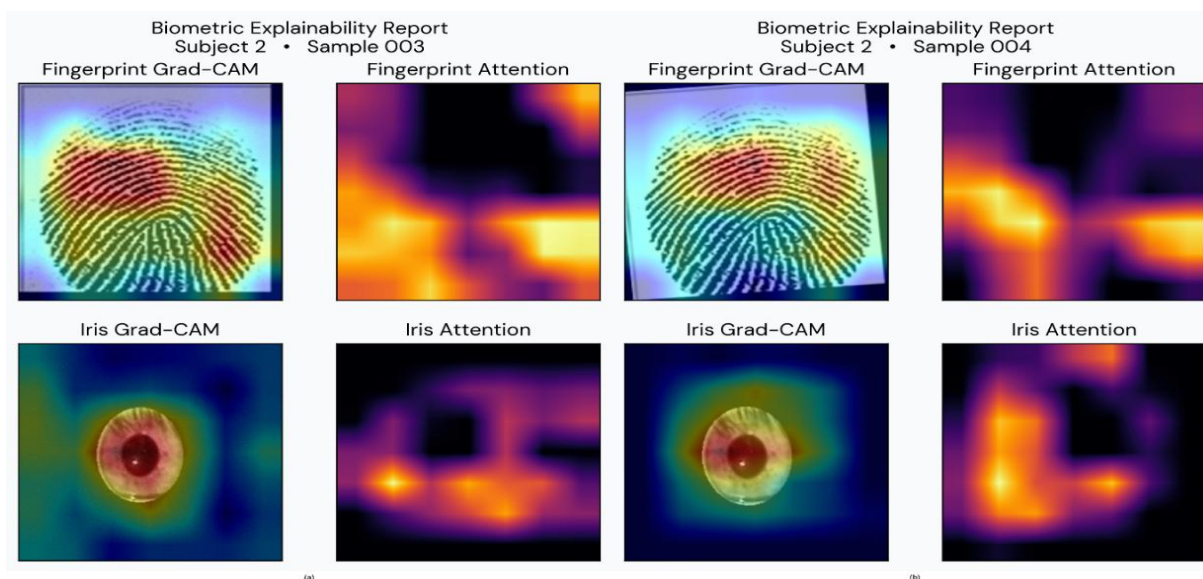


Figure 5. Grad-CAM-based explainability report of the proposed model.

Table 1 displays the key identification performance metrics of the proposed FIF-DML model. The model achieves Rank-1 and Rank-5 accuracies of 96.78% and 96.89%, respectively. The precision, recall, and F1-Score values exceed 96%, while the MCC value of 96.72% further confirms the model's reliability in the biometric identification task.

Table 1. Identification performance metrics of the model.

Metric	Values
Rank1_Accuracy	96.78
Rank5_Accuracy	96.89
Precision	97.33
Recall	96.78
F1Score	96.79
MCC	96.72

International Journal of
Medical Toxicology & Forensic Medicine

Discussion

Table 2 presents a comparative analysis of the identification metrics of the proposed FIF-DML framework against existing approaches [16-18]. Models such as VGG19, GAN, and Capsule Network achieve very moderate results across all performance metrics. The proposed model demonstrates superior identification capability by achieving higher values across all performance metrics than the existing models, with Rank1_Accuracy of 96.78%, Rank5_Accuracy of 96.89, precision of 97.33%, and less computational time of 4.23 sec. Therefore, the results show that the proposed FIF-DML model outperforms all existing models.

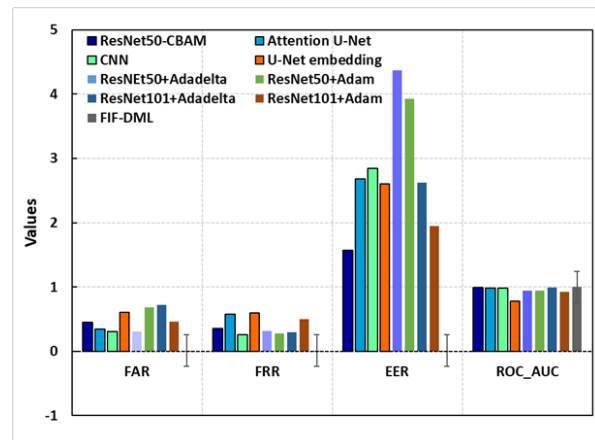
Table 2. Comparative performance analysis of identification metrics.

Model	Rank 1 Accuracy	Rank 5 Accuracy	Precision	Time (ms)
InceptionV3	86.02	94.38	94.52	8.72
CNN				
VGG19	67.53	72.09	75.21	8.23
DBN	91.50	93.29	92.65	7.04
GAN	79.00	85.57	83.09	8.30
deep 3D CNNs	94.25	93.02	94.42	7.77
PointNet++	83.00	89.40	92.72	6.13
Capsule network	74.44	76.66	81.29	8.91
FIF-DML	96.78	96.89	97.33	4.23

International Journal of
Medical Toxicology & Forensic Medicine

Figure 6 displays the comparative performance analysis of verification metrics of the proposed FIF-DML framework with existing approaches. The proposed FIF-DML achieves low FAR, FRR, and EER

values of 0.0110, along with a high ROC_AUC of 0.9983, indicating strong authentication performance. Therefore, the proposed model is effective and suitable for biometric verification and outperforms existing approaches.



International Journal of
Medical Toxicology & Forensic Medicine

Figure 6. Comparative performance analysis graph of verification metrics.

To examine the component-wise analysis of the proposed model, Table 3 provides a brief analysis. The table values indicate that including multiple components in the proposed model yields superior results.

Table 4 shows the performance of the proposed model across different folds. The values show consistent and stable performance across all folds.

Conclusion

This research work presented the FIF-DML methodology for multimodal biometric recognition and authentication by integrating fingerprint and iris features. The proposed model learns features using two parallel EfficientNet-B0 networks that independently extract deep representations from fingerprint and iris inputs. These extracted features are then fused at the feature level to form a unified representation that captures discriminative biometric characteristics from both sources. Moreover, ArcFace-based metric learning is used to learn the fused features into a compact embedding space, enhancing inter-class separability and intra-class compactness. Comprehensive experimental evaluation of the proposed FIF-DML model on the Multimodal_Iris_Fingerprint_Biometric_data achieves a precision of 97.33%, confirming its superior performance compared to existing methods. Future work can focus on developing lightweight techniques to decrease computational complexity while

Table 3. Ablation study analysis.

Model	Rank 1 Accuracy	Rank 5 Accuracy	Precision	Recall	F1 Score
EfficientNet-B0 + Fingerprint Only	88.42	91.36	89.17	87.94	88.55
EfficientNet-B0 + Iris Only	90.15	93.27	91.08	89.76	90.41
EfficientNet-B0 + Fingerprint + Iris Feature Fusion	93.24	95.18	94.02	92.83	93.42
EfficientNet-B0 + Feature Fusion + Standard Classification Loss	94.67	95.82	95.21	94.38	94.79
EfficientNet-B0 + Feature Fusion + ArcFace	95.73	96.41	96.18	95.61	95.89
FIF-DML (EfficientNet-B0 + Feature Fusion + ArcFace + 128-D Embedding)	96.78	96.89	97.33	96.72	96.99

International Journal of
Medical Toxicology & Forensic Medicine**Table 4.** Cross-validation results of proposed model.

Fold	Rank 1 Accuracy	Rank 5 Accuracy	Precision	Recall	F1Score
1	96.42	96.61	96.98	96.35	96.58
2	96.71	96.83	97.21	96.64	96.82
3	96.89	97.02	97.46	96.91	97.18
4	96.55	96.76	97.08	96.53	96.80
5	96.93	97.10	97.55	96.88	97.21
6	96.81	96.94	97.32	96.79	97.05
7	96.67	96.85	97.19	96.69	96.93
8	96.84	96.97	97.41	96.82	97.11
9	96.76	96.91	97.28	96.74	97.00
10	96.91	97.11	97.83	96.89	97.22
SD	±0.165	±0.155	±0.246	±0.178	±0.209
Mean	96.75	96.91	97.33	96.72	96.99

International Journal of
Medical Toxicology & Forensic Medicine

maintaining recognition accuracy. In addition, advanced fusion methodologies, comprising attention and transformer-based techniques, can be explored to enhance feature learning.

Ethical Approval

Not applicable.

Acknowledgment

None.

Funding

No external funding was received for this research.

Conflicts of Interest

The authors report that there are no competing interests to declare.

References

[1] Sheela SV, Radhika KR. Deep learning-driven fusion of iris biometrics for optimized multimodal authentication and informative

security. *Int J Inf Comput Secur.* 2026;29(4):397-435. [DOI: [10.1504/IJICS.2026.153343](https://doi.org/10.1504/IJICS.2026.153343)]

[2] Nurdin D, Lawi A, Toaha S. From Biometric Surfaces to Contextual Intelligence: A Comprehensive Survey of Multimodal Authentication and the Trimodal Fusion Paradigm. In: 2026 International Conference on Current Research in Artificial Intelligence and Data Science (ICCRAIDS). Vol 1. IEEE; 2026:1-6. [DOI: [10.1109/ICCRAIDS67816.2026.11519743](https://doi.org/10.1109/ICCRAIDS67816.2026.11519743)]

[3] Nurudeen BI, Sani AA, Oriyomi AM, Towobola IO. Facial Recognition System Using A Deep Learning Approach: a Review. *J Inst Res Big Data Anal Innov.* 2026;2(1):400-14.

[4] Gao C, Yang Z, Jia W, Leng L, Zhang B, Teoh ABJ. Deep Learning in Palmprint Recognition: A Comprehensive Survey. *IEEE Trans Syst Man Cybern Syst.* 2026. [DOI: [10.32388/YZ5TFY](https://doi.org/10.32388/YZ5TFY)]

[5] Ali K, Ajan A, Imtiaz MH. FPGA-Based Implementations of Biometric Recognition: A Review. *Electronics.* 2026;15(10):2145. [DOI: [10.3390/e15102145](https://doi.org/10.3390/e15102145)]

[10.3390/electronics15102145](https://doi.org/10.3390/electronics15102145)]

- [6] Kakade S, Raut U. Biometric Authentication Systems: Trends, Challenges, and Future Prospects-A Comprehensive Review. *MethodsX*. 2026;103908.
- [7] Hu B. Tiny transformers: multimodal digital biometric authentication model using tiny enhanced dual transformers with quality-gated fusion network. *PeerJ Comput Sci*. 2026;12:e3657. [DOI: [10.7717/peerj-cs.3657](https://doi.org/10.7717/peerj-cs.3657)]
- [8] Binu D. Bimodal biometric authentication based on index level fusion for combined iris and fingerprint biometrics. *Biomed Signal Process Control*. 2026;117:109556. [DOI: [10.1016/j.bspc.2026.109556](https://doi.org/10.1016/j.bspc.2026.109556)]
- [9] Mohammed A, Salama A, Shebka N, Ismail A. Enhancing network access control using a multi-modal biometric authentication framework. *Eng Technol Appl Sci Res*. 2025;15(1):20144-50. [DOI: [10.48084/etasr.9554](https://doi.org/10.48084/etasr.9554)]
- [10] Kaur P, Kant C. 2-Phase Multi-trait Biometric Authentication Model Against Spoofing Attack Using Deep Hash Model. *SN Comput Sci*. 2024;6(1):47. [DOI: [10.1007/s42979-024-03513-w](https://doi.org/10.1007/s42979-024-03513-w)]
- [11] Maghrabi LA, Altwijri M, Binyamin SS, Alallah FS, Hamed D, Ragab M. Secure biometric identification using Orca predator's algorithm with deep learning: Retinal iris image analysis. *IEEE Access*. 2024;12:18858-67. [DOI: [10.1109/ACCESS.2024.3360871](https://doi.org/10.1109/ACCESS.2024.3360871)]
- [12] Nivedita V, Joseph JA, Varghese DT, Nithyanandh S, Sundaram J, Ali G. Multi-modal biometric authentication integrating gait and Face Recognition for Mobile Security. In: 2025 International Conference on Smart & Sustainable Technology (INCSST). IEEE; 2025:1-6. [DOI: [10.1109/INCSST64791.2025.11210361](https://doi.org/10.1109/INCSST64791.2025.11210361)]
- [13] Multimodal iris-fingerprint biometric data [Dataset]. Kaggle. Accessed October 13, 2026. <https://www.kaggle.com/datasets/ninadmehendale/multimodal-iris-fingerprint-biometric-data>
- [14] Iqbal A, Jaffar MA, Jahangir R. Enhancing brain tumor multi-classification using EfficientNet B0-based intelligent diagnosis for Internet of Medical Things (IoMT) applications. *Information*. 2024;15(8):489. [DOI: [10.3390/info15080489](https://doi.org/10.3390/info15080489)]
- [15] Peng N, Chen M, Feng G. Feature extraction and recognition of Chinese mitten crab carapace based on improved MobileNetV2. *Appl Sci*. 2024;14(12):4982. [DOI: [10.3390/app14124982](https://doi.org/10.3390/app14124982)]
- [16] Alduhailan A, Kamarudin NH, Sheik Abdullah SNH, Dau A. Deep learning in biometric authentication: Challenges, recent advancements, and future trends. *J Adv Inf Technol*. 2025;16:458-77. [DOI: [10.12720/jait.16.4.458-477](https://doi.org/10.12720/jait.16.4.458-477)]
- [17] Kaur V, Bhatt DP, Tharewal S, Tiwari PK. Intelligent multimodal 3D biometric recognition using PointNet++ for robust face-ear authentication. *Sci Rep*. 2025. [DOI: [10.1038/s41598-025-27757-5](https://doi.org/10.1038/s41598-025-27757-5)]
- [18] Khatri M, Sharma A. Deep learning approach based on Iris, face, and palmprint fusion for multimodal biometric recognition system. *Int J Performability Eng*. 2023;19(6):407. [DOI: [10.23940/ijpe.23.06.p6.407416](https://doi.org/10.23940/ijpe.23.06.p6.407416)]