Secure and Advanced Score-Level Fusion Techniques for Robust Multi-modal Biometric Identification

Jothibasu Marappan[1], Karthik Murugesan[2], Malar Elanageeran[3], Usha Subramaniam[4]

[1]Department of Electronics and Communication Engineering, PSG Institute of Technology and Applied Research, Coimbatore – 641 062, Tamilnadu, India, jothibasu@psgitech.ac.in

[2,4]Department of Electrical and Electronics Engineering, Kongu Engineering College, Erode – 638 060, Tamilnadu, India , karthikprm@kongu.ac.in, usha@kongu.ac.in

[3]Department of Electrical and Electronics Engineering, PSG Institute of Technology and Applied Research, Coimbatore – 641 062, Tamilnadu, India, emr@psgitech.ac.in

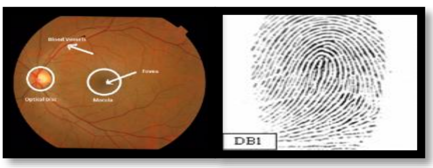
Abstract: Biometric authentication systems are widely used in security applications as they use unique physiological or behavioral traits that are difficult to to duplicate or manipulate. However, uni-modal biometric systems often suffer from noise sensitivity, poor-quality samples and vulnerability to various limitations that reduce authentication reliability. This paper presents a multi-modal biometric authentication system that integrates fingerprint and retinal biometrics to improve identification accuracy and robustness. The proposed approach combines Local Binary Pattern (LBP)-based texture features and Convolutional Neural Network (CNN)-based deep features for both modalities, enabling the extraction of complementary local and high-level discriminative information. A systematic evaluation of multiple score normalization methods and fusion strategies is carried out within a unified framework to analyze their impact on recognition performance. The normalized scores are combined using a weighted score-level fusion approach to effectively utilize information from both modalities. Experimental evaluation on standard biometric datasets shows that the proposed system achieves a recognition accuracy of 98.81% with a low error rate of 1.1235%. The results indicate improved performance compared to uni-modal systems. The study also highlights the challenges associated with multi-modal dataset availability and their impact on performance evaluation.

Keywords: Multi-modal biometric system, retina, fingerprint, score-level fusion, normalization

1 Introduction

The rapid growth of digital services and online transactions has increased the demand for secure and reliable authentication systems. The conventional methods such as passwords, PINs and smart cards are widely used, but remain vulnerable to theft, duplication and unauthorized access. As a result, biometric authentication systems have gained attention because they use physiological or behavioral characteristics that are difficult to duplicate [1]. The common biometric modalities include fingerprint, iris, face, voice and gait recognition [2]. Among these, fingerprint and retinal biometrics are considered highly reliable due to their uniqueness and permanence [3]. Fingerprint recognition is widely adopted because of its simplicity, low cost and ease of acquisition [5], whereas retinal biometrics provide higher security since retinal vascular patterns are internally protected and difficult to fake [7]. Despite these advantages, uni-modal biometric systems often suffer from noisy data, sensor variations, spoofing attacks and performance degradation under non-ideal conditions. These limitations can reduce authentication accuracy in practical applications. Multi-modal biometric systems address these issues by combining information from multiple biometric traits, thereby improving recognition performance and reliability. Several multi-modal biometric systems using different feature extraction and fusion methods have been reported in the literature [8], [9]. However, many existing methods rely on the conventional score fusion approaches without detailed analysis of weighted fusion strategies or mathematical modeling of the fusion process. In addition, limited work has explored the integration of fingerprint and retinal biometrics using deep learning based feature extraction techniques [17],[24].

To address the above limitations, this paper presents a multi-modal biometric authentication system that combines fingerprint and retinal modalities. Convolutional Neural Networks (CNNs) are used to extract fingerprint features [19], while Local Binary Patterns (LBPs) are employed for retinal texture analysis. The matching scores from both modalities are normalized and combined using a weighted score-level fusion approach to improve discrimination between genuine and impostor samples. The proposed system includes preprocessing, feature extraction, score normalization, fusion and threshold based decision making.



(a) (b)

Figure 1

(a) Retinal and (b) Fingerprint Biometric Traits

By combining fingerprint and retinal traits, the proposed system aims to improve authentication accuracy and resistance to spoofing attacks. Figure 1 (a) shows the retinal biometric trait and Figure 1 (b) shows the fingerprint biometric trait used in the proposed system.

The main contribution of this work lies in the development and systematic evaluation of a multi-modal biometric authentication framework that integrates retinal and fingerprint modalities. The proposed approach combines CNN-based deep feature extraction with LBP-based texture descriptors to capture both high-level spatial information and fine-grained local patterns. In addition, multiple score normalization techniques and fusion strategies are analyzed within a unified system to evaluate their impact on system performance. The study further investigates weighted score-level fusion to effectively combine complementary information from both modalities and improve authentication accuracy. The work also discusses the challenges associated with multi-modal dataset availability and their implications for performance evaluation.

The remainder of this paper is organized as follows. Section II reviews related works on multi-modal biometric fusion. Section III presents the proposed score-level fusion methodology. Section IV describes the experimental setup and performance evaluation. Section V presents the experimental results and discussion. Finally, Section VI concludes the paper and outlines future research directions.

2 Background of the Research

Several studies have demonstrated the multi-modal biometric systems provide improved security, reliability and spoof resistance compared to uni-modal systems. Different biometric combinations and fusion strategies have been explored to improve recognition performance and to reduce error rates.

Ross et al. [1] and Gavrilova and Monwar [2] provided foundational studies on multi-modal biometric systems and biometric fusion techniques, highlighting the advantages of combining multiple biometric modalities for improved authentication reliability and security. Chergui et al. [9] proposed a multi-modal palmprint and finger-knuckle-print framework using fusion techniques that achieved lower error rates compared to uni-modal systems. Jagadiswary and Saraswady [12] developed a multi-modal system integrating face, fingerprint, iris, and signature modalities, demonstrating improved recognition performance, although computational complexity increased with the number of modalities. Aizi and Ouslim [15] proposed score-level fusion strategies to improve biometric identification performance and reduce recognition errors. Similarly, He et al. [21] investigated score-level fusion approaches for fingerprint, face, and finger-vein biometrics using sum-rule and SVM-based fusion methods.

Retinal biometrics have also attracted considerable attention because of their uniqueness and stability. Srivastava [11] introduced a deep learning-based multi-modal framework integrating retina, finger-vein, and fingerprint modalities through score-level fusion. Mustafa et al. [13] combined iris and fingerprint biometrics using GLCM-based feature extraction and KNN classification to improve recognition accuracy. Jothibasu Marappan et al. [18] presented a retinal authentication framework using Radial Chebyshev Moments for retinal vessel segmentation and feature extraction. Recent studies have further focused on improving multi-modal fusion efficiency and recognition performance. Thangaraj et al. [7] proposed a neural-network-based retinal vessel segmentation framework for biometric analysis. Vishi and Mavroeidis [20] evaluated different score-level fusion approaches for fingerprint and finger-vein biometrics, while Elhoseny et al. [19] investigated multi-modal biometric fusion using fingerprint and iris modalities with advanced score-level fusion strategies. More recently, Mane et al. [26] proposed a deep learning-based multi-modal biometric authentication framework using fingerprint and iris modalities with score-level fusion.

Although previous studies have demonstrated the advantages of multi-modal biometric systems, the challenges such as related to computational complexity, limited dataset diversity and robust score fusion still remain. In particular, only limited research has explored retina–fingerprint fusion using deep learning based feature extraction and weighted score-level fusion techniques. To address these limitations, the present research combines CNN-based deep features and LBP texture descriptors with weighted score-level fusion to improve recognition accuracy and authentication reliability.

3 Proposed Multi-modal Biometric System

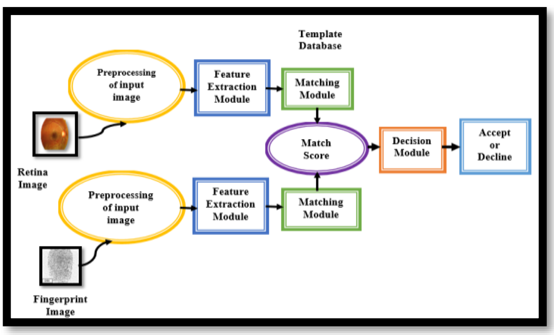


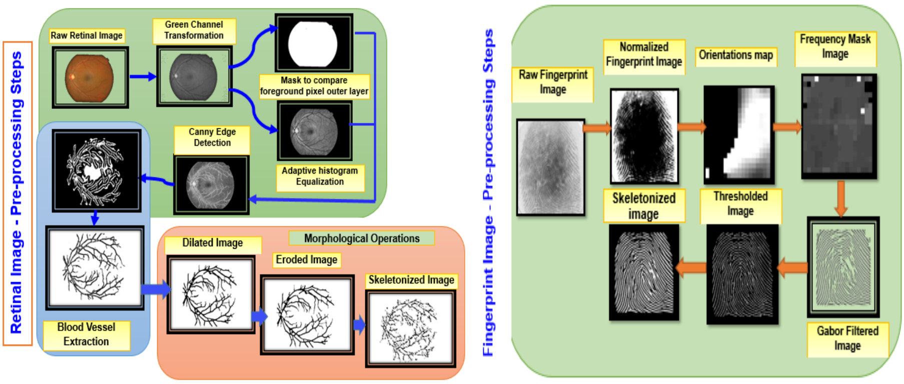
Figure 2

Block Diagram of the Proposed Multi-modal Biometric System

The proposed multi-modal biometric authentication system combines retinal and fingerprint modalities to improve authentication accuracy, reliability and resistance to spoofing attacks. Fingerprint and retinal modalities are selected because of their complementary characteristics. Fingerprint biometrics provide ease of acquisition and stable ridge structures, whereas retinal biometrics offer high security due to the uniqueness and internal protection of retinal vascular patterns. The overall system consists of pre-processing, feature extraction, score normalization, score-level fusion, and decision making. Figure 2 illustrates the block diagram of the proposed multi-modal biometric system

3.1 Preprocessing

Preprocessing is employed to improve the quality of retinal and fingerprint images before feature extraction. Since both the modalities possess different structural characteristics, separate preprocessing procedures are applied.



(a) (b)

Figure 3

(a) Preprocessing stages involved in retina (b) Preprocessing stages involved in fingerprint

Preprocessing of retinal images includes grayscale conversion, green channel extraction, contrast enhancement and vessel segmentation to improve vascular visibility. Morphological operations and matched filtering are employed to extract retinal blood vessels, followed by the extraction of Region of Interest (ROI) around the optic disc [7], [18]. Finally, Gaussian filtering is applied to suppress noise while preserving important vascular structures. Figure 3 (a) illustrates the preprocessing stages used for retinal biometrics. Preprocessing of fingerprint images involves image normalization, ridge orientation estimation and ridge enhancement using Gabor filtering. The enhanced image is then binarized and skeletonized to obtain refined ridge structures for reliable feature extraction and matching [6], [10]. Figure 3 (b) illustrates the preprocessing stages used for fingerprint biometrics.

3.2 Feature Extraction

Feature extraction is executed to obtain discriminative representations from retinal and fingerprint images. In the proposed system, LBPs and CNNs are employed to capture both local texture information and high-level spatial features.

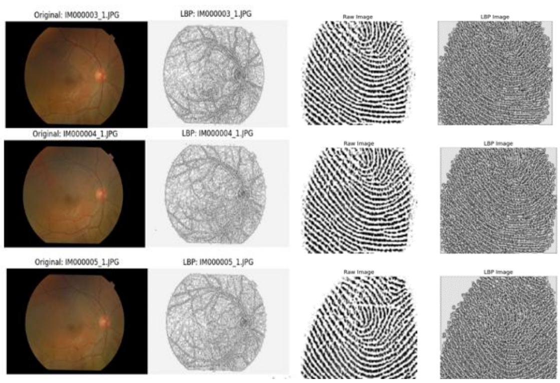
3.2.1 Local Binary Pattern (LBP) Features

LBP is used to extract local texture information from the retinal vascular structures and fingerprint ridge patterns [16]. The image is divided into local neighborhoods, and the intensity of each neighboring pixel is compared with the intensity of the center pixel. Based on this comparison, a binary pattern is generated and converted into a decimal value representing local texture characteristics. Equation (1) provides the mathematical expression for the LBP operator.

$$LBP_{(u,v)} = \sum_{p=0}^{p-1} S\left(I_p - I_c\right))2^p \tag{1}$$

where $LBP_{(u,v)}$ represents the LBP value at pixel location (u, v), I_c denotes the intensity of the center pixel, I_p represents the intensity of neighboring pixels, P is the total number of neighboring pixels and $S_{(t)}$ is the thresholding function defined as $S_{(t)} = \begin{Bmatrix} 1, & x \geq 0 \\ 0, & x < 0 \end{Bmatrix}$

The LBP histograms extracted from the retinal and fingerprint images are combined to form discriminative feature vectors. Figure 4 (a) illustrates the LBP feature maps of retinal images and Figure 4 (b) illustrates the LBP feature maps fingerprint images.



(a) (b)

Figure 4

(a) LBP features of Retinal images (b) LBP features of fingerprint images

3.2.2 CNN-based Feature Extraction

The proposed deep learning employs a CNN to extract discriminative spatial features from retinal and fingerprint biometric images [10], [14]. The CNN is designed with three convolutional blocks followed by fully connected layers for classification. The input images are resized and preprocessed before being fed into the network [17]. Each convolutional block consists of a convolution layer with 3*3 kernels, followed by batch normalization, Rectified Linear Unit (ReLU) activation and max-pooling layers. The number of filters is progressively increased as 32, 64 and 128 across the three convolutional layers to capture hierarchical feature representations ranging from low-level edges to high-level texture patterns. while Table 1 summarizes the configuration and purpose of each layer used in the network. Equation (2) provides the mathematical expression used to determine the output feature dimension in the convolution operation.

$$p = \frac{x-y+2a}{z} + 1 \tag{2}$$

where p denotes the output feature dimension, x represents the input image size, y denotes the kernel size, a represents the padding size and z denotes the stride value. The CNN system generates high-level feature representations that complement the texture-based LBP features.

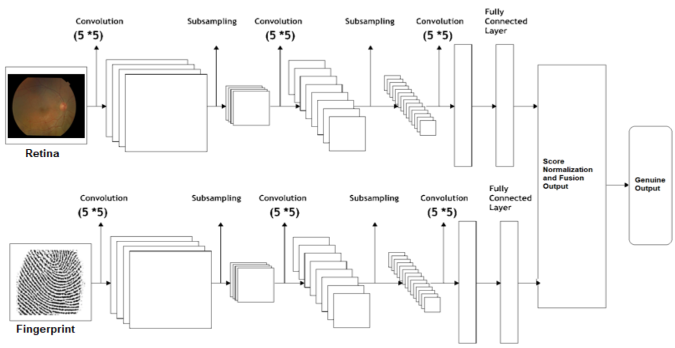


Figure 5

Proposed CNN Architecture for extracting features from Retinal and Fingerprint Images

In addition to the architectural design, the CNN training process is configured using standard deep learning practices. The input images are resized and normalized before training. The model is trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 32, which provides a balance between convergence stability and computational efficiency. The network is trained for multiple epochs (10, 15, 20, and 25) to analyze the impact of training duration on recognition performance. A dropout rate of 0.5 is applied in the fully connected layer to reduce overfitting. The dataset is divided into training and testing subsets to ensure proper evaluation of generalization performance. These parameter settings are selected based on empirical evaluation and commonly adopted configurations in the literature.

Table 1

CNN Architecture Used for Feature Extraction

Layer	Configuration	Purpose
Input Layer	Preprocessed Retina/Fingerprint Image	Input feature representation
Convolution Layer 1	3×3 kernel, 32 filters + ReLU	Low-level feature extraction
Max Pooling Layer 1	2×2 pooling	Dimensionality reduction
Convolution Layer 2	3×3 kernel, 64 filters + ReLU	Texture and edge extraction
Max Pooling Layer 2	2×2 pooling	Feature refinement
Convolution Layer 3	3×3 kernel, 128 filters + ReLU	High-level discriminative feature extraction
Flatten Layer	Feature vector conversion	Converts feature maps into 1D vectors
Fully Connected Layer	Dense layer	Classification feature learning
Dropout Layer	Dropout rate = 0.5	Reduces overfitting

3.3 Score Normalization

The matching scores generated from retinal and fingerprint modalities may exhibit different numerical ranges and statistical distributions. Therefore, score normalization is employed prior to score-level fusion to ensure score compatibility and to improve fusion performance. In this study, several normalization techniques, including Min–Max normalization, Z-score normalization, Tanh normalization, Median–MAD normalization and Double-Sigmoid normalization are employed [21],[22].

3.3.1 Min-Max (MM) Normalization

The Min-Max normalization scales the matching scores to a specified range, typically between 0 and 1, by subtracting the minimum score and dividing by the score range. Equation (3) provides the mathematical expression for Min–Max normalization.

$$MM_{xi} = \frac{x_i - min(X_i)}{max(X_i) - min(X_i)} \quad (3)$$

where x_i represents the matching score of the i^{th} sample, X_i denotes the set of matching scores, $\min(X_i)$ and $\max(X_i)$ represent the minimum and maximum scores.

3.3.2 Z-score Normalization

Z-score normalization standardizes the matching scores using the mean and standard deviation, resulting in scores with zero mean and unit variance. Equation (4) represents the Z-score normalization method.

$$Z_i = \left[\left[\frac{x_i - \mu}{\sigma}\right]\right] \tag{4}$$

where x_i denotes the matching score, μ represents the mean score, and σ denotes the standard deviation.

3.3.3 Double-Sigmoid (DS) Normalization

Double-Sigmoid normalization employs two sigmoid functions to scale the matching scores within the range of 0 to 1. This method provides flexibility in shaping the score transformation curve. Equation (5) provides the mathematical representation of Double-Sigmoid normalization.

$$x = \begin{cases} \frac{1}{1+exp(-2(\frac{s-t}{t_1}))} & s < t \\ \frac{1}{1+exp(-2(\frac{s-t}{t_2}))} & otherwise \end{cases} \tag{5}$$

where s represents the matching score, t denotes the operating threshold and t_1, t_2 are scaling parameters.

3.3.4 Median-MAD Normalization

It divides the scores by using Median Absolute Deviation (MAD) and then subtracts the median score. It is robust against outliers and useful for dealing with noisy data. Equation (6) provides the mathematical expression for Median–MAD normalization.

$$\mathrm{M_i} = \frac{\mathrm{x_i\text{-}median(X_i)}}{\mathrm{MAD(X)}} \tag{6}$$

where $\mathrm{median(X_i)}$ represents the median score and MAD(X) denotes the Median Absolute Deviation of the score set.

3.3.5 Tanh-Normalization

Tanh normalization scales the matching scores using the hyperbolic tangent function, mapping the scores into a bounded range. Equation (7) represents the Tanh normalization function.

$$T_i = \frac{1}{2}\left[tanh\left(0.01\left(\frac{x_i - \mu}{\sigma}\right)\right) + 1\right] \tag{7}$$

where x_i denotes the matching score, μ represents the mean score and σ denotes the standard deviation. Table 2 presents the computational complexity of individual normalization techniques, it is important to note that the normalization stage constitutes only a small part of the overall multi-modal biometric system. The primary computational cost is associated with feature extraction and matching processes, particularly the CNN-based representation learning. The selected normalization methods (Tanh, Double-Sigmoid and Median–MAD) introduce only moderate computational overhead and do not significantly affect the overall system efficiency. Furthermore, the score-level fusion using a weighted sum rule is computationally inexpensive, involving only simple arithmetic operations. Therefore, the proposed system maintains a balance between robustness and computational efficiency, making it suitable for practical deployment scenarios with appropriate implementation optimizations.

Table 2

Normalization techniques used in the proposed system

Score-level Normalization Methods	Computational Complexity	Robustness
Min-max	Low	No
Median-MAD	Moderate	Yes
Tanh	Moderate	Yes
Double-Sigmoid	Moderate	Yes
Z-Score	Low	Moderate

3.4 Score-Level Fusion

Multi-modal biometric systems can perform fusion at different levels, including sensor level, feature level, score-level and decision level, as illustrated in Figure 6. Among these approaches, score-level fusion is widely preferred because it effectively combines the matching scores generated by multiple biometric modalities while preserving significant discriminatory information [23],[24]. Compared to decision-level fusion, score-level fusion provides better flexibility and improved recognition performance [25].

Sum Rule: The Sum Rule combines the normalized matching scores obtained from retinal and fingerprint modalities through direct addition and is expressed as

$$F_{Su} = S_{ret} + S_{fp} \tag{8}$$

Maximum Rule: The Maximum Rule selects the highest matching score among the biometric modalities is defined as

$$F_{Max} = max\ ((S_{ret}), (S_{fp})) \tag{9}$$

Minimum Rule: It selects the minimum score among all modalities as the final matching score, assuming that the weakest modality determines the authentication decision. Equation (10) represents the Minimum Rule used for score-level fusion,

$$F_{min} = min\ ((S_{ret}), (S_{fp})) \tag{10}$$

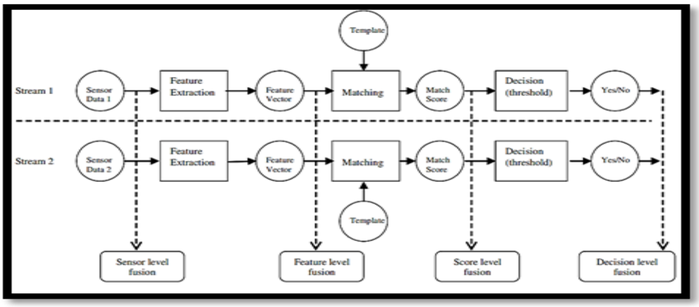


Figure 6

Fusion levels in a multi-modal biometric system

Simple Averaging Rule: The Simple Averaging Rule computes the average of the matching scores obtained from multiple biometric modalities and is expressed as

$$F_{SA} = (\frac{S_{ret}+S_{fp}}{2}) \tag{11}$$

Weighted Sum Rule: The Weighted Sum Rule assigns weights to the individual biometric modalities according to their contribution and combines the weighted scores to generate the final matching score. The fusion rule is expressed as

$$F_{ws} = W_1 \times S_{ret} + W_2 \times S_{fp} \tag{12}$$

where S_{ret} and S_{fp} represent the normalized matching scores obtained from retinal and fingerprint modalities, respectively. The weighted sum fusion method assigns weights W_1 and W_2 to the retinal and fingerprint modalities and different weight combinations are experimentally evaluated to determine their optimal contribution. Table 3 presents the recognition accuracy obtained for different weight combinations assigned to retinal and fingerprint modalities. The results show that equal weighting of both modalities (W_1 = 0.5 and W_2 = 0.5) achieves the highest recognition accuracy of 98.9%, indicating that both biometric traits contribute complementary discriminatory information with nearly equal significance.

3.5 Decision Making

The final authentication decision is made by comparing the fused matching score with a predefined threshold. If the fused score exceeds the threshold, the user will be classified as genuine; otherwise, the user will be considered an impostor. The threshold is selected experimentally to achieve an optimal balance between FAR and FRR [21].

Table 3

Weights and Recognition Accuracy for Retina and Fingerprint Biometric Modality

Retina	Fingerprint	Accuracy
0.25	0.75	96.2
0.4	0.6	97.4
0.5	0.5	98.9
0.7	0.3	98.8
0.8	0.2	96.0

4 Experimental Results and Discussion

Experiments are conducted to evaluate the effectiveness of the proposed multi-modal biometric authentication system that integrates retinal and fingerprint modalities. The performance of the proposed system is compared with individual uni-modal systems and multi-modal systems using different score-level fusion and normalization techniques. The system is evaluated using performance metrics such as Equal Error Rate (EER), Genuine Acceptance Rate (GAR), accuracy, precision, recall, specificity and Receiver Operating Characteristic (ROC) analysis.

4.1 Multi-modal Virtual Dataset

The publicly available datasets containing paired retinal and fingerprint samples from the same individuals are limited. Hence, a virtual multi-modal dataset is constructed by combining the established retinal and fingerprint databases. Retinal images are obtained from the RIDB, VARIA, DRIVE, and STARE datasets while fingerprint images are collected from the FVC 2000/2002/2004 databases [27]. The dataset created consists of 100 individuals, with five samples collected for each modality, resulting in a total of 500 images. For experimental evaluation, 230 images are used for training and 270 images are used for testing. For the retinal modality, genuine scores are generated by comparing retinal images belonging to the same individual while impostor scores are obtained from the comparisons among different individuals. This results in 100 genuine scores and 9,900 impostor scores. Similarly, the fingerprint dataset also produces 100 genuine scores and 9,900 impostor scores using intra-class and inter-class comparisons. Although the paired retinal and fingerprint samples do not originate from identical individuals, this may limit direct generalization to real-world biometric deployments. Nevertheless, such protocols are commonly adopted in multi-modal biometric research for preliminary validation when the real paired datasets are unavailable.

4.2 Evaluation Metrics

The performance of the proposed multi-modal biometric identification system is evaluated using ROC curves and EER. Genuine and impostor matching scores are generated prior to the authentication decision process. To ensure compatibility between matching scores obtained from different biometric modalities, score normalization techniques are applied before score-level fusion. The system performance is further analyzed using GAR at different operating points of FAR. ROC analysis is performed to examine the trade-off between GAR and FAR under different fusion strategies and threshold settings. In addition to ROC and EER analysis, the performance of the proposed multi-modal biometric system is evaluated using standard classification metrics, including accuracy, precision, sensitivity, specificity and F1-score, as expressed in Equations (13) - (17).

$$Accuracy = \frac{T_P+T_N}{T_P+T_N+F_P+F_N} \tag{13}$$

$$Precision = \frac{T_P}{T_P+F_P} \tag{14}$$

$$Sensitivity = \frac{T_P}{T_P+F_N} \tag{15}$$

$$Specificity = \frac{T_N}{T_N+F_P} \tag{16}$$

$$F - Measure = \frac{2\ X\ (Sensitivity\ X\ Precision)}{(Sensitivity+\ Precision)} \tag{17}$$

where T_P denotes True Positive, T_N represents True Negative, F_P indicates False Positive and F_N stands for False Negative. Classification accuracy is used to assess the overall performance. Specificity measures the effectiveness of the classifiers in identifying the negative level of data while the F-measure evaluates the balance between precision and sensitivity of the classifier. These evaluation metrics provide a comprehensive assessment of the reliability, robustness and authentication performance of the proposed multi-modal biometric system.

4.3 Performance Analysis of Fusion Methods

The performance of the proposed multi-modal biometric system is evaluated using several score-level fusion methods, including Sum Rule, Min Rule, Max Rule, Simple Average Rule and Weighted Sum Rule. The effectiveness of each fusion method is assessed using EER, GAR and FAR. Low FAR operating points such as

0.01 and 0.001 are particularly important in biometric authentication systems because even a small number of false acceptances can compromise system security. ROC curves are used to analyze the trade-off between GAR and FAR for different fusion strategies. Table 4 presents the performance comparison of different score-level fusion techniques using EER as the evaluation metric. The results demonstrate that multi-modal fusion significantly improves authentication performance compared to uni-modal retinal and fingerprint systems. Among the evaluated methods, the Weighted Sum Rule achieves the lowest EER of 1.123%, indicating superior recognition performance and improved reliability. The superior performance achieved by the weighted sum fusion method demonstrates the effectiveness of combining complementary retinal and fingerprint information within a multi-modal biometric system.

Table 4

Performance of Score-Level Fusion Techniques

Biometric Traits	Biometric System	Fusion Rule	EER (%)
Retina	Uni-modal	NA	3.592
Fingerprint		NA	4.836
Retina + Fingerprint	Multi-modal	Sum rule	1.725
		Min rule	2.320
		Max rule	2.020
		Simple Average	1.865
		Weighted Sum Rule	1.123

Figure 7 presents the ROC curves for retinal, fingerprint and multi-modal fusion methods. The multi-modal fusion methods consistently outperform the uni-modal systems across different FAR operating points, demonstrating improved robustness and recognition reliability. The experimental results further indicate that the normalization techniques such as Tanh, Double-Sigmoid and Z-score normalization improve score compatibility and enhance multi-modal fusion performance. Among the fusion methods evaluated, the Sum Rule and Weighted Sum Rule consistently produce better recognition accuracy compared to Min Rule, Max Rule and Simple Average Rule.

Table 5

Recognition accuracy for Raw input images and LBP features

Modality	Epoch 10	Epoch 15	Epoch 20	Epoch 25
Raw Retina	82.45	91.78	94.85	96.87
LBP feature map based Retina	80.90	92.00	95.50	97.90
Raw Fingerprint	82.95	86.35	87.95	89.95
LBP feature map based Fingerprint	77.35	85.20	90.80	97.35

4.4 Recognition Accuracy

The recognition performance of retinal and fingerprint modalities is evaluated using both raw biometric images and LBP-transformed feature maps. Table 5 presents the recognition accuracy at different training epochs, showing a consistent improvement with increasing epochs. For retinal images, accuracy improves from 82.45% at epoch 10 to 96.87% at epoch 25 for raw inputs, while LBP-based features achieve a higher accuracy of 97.90%. Similarly, for fingerprint images, accuracy increases from 82.95% to 89.95% for raw inputs, whereas LBP features reach 97.35%. These results highlight the effectiveness of LBP in capturing discriminative texture information in retinal vascular structures and fingerprint ridge patterns. The CNN model effectively learns both local and high-level representations from the biometric inputs. Figure 8 compares raw features, LBP feature maps, and fusion strategies. Equal-weight fusion (50%–50%) and LBP-dominant fusion (75%–25%) are evaluated, with LBP-based fusion showing improved performance. Figure 7 presents the ROC curves, further confirming the effectiveness of multi-modal fusion.

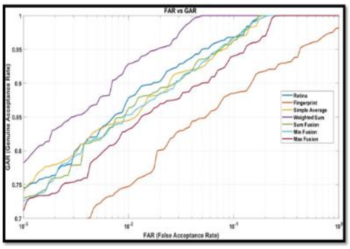


Figure 7

ROC of Fingerprint, Retina and Multi-modal Dataset

4.5 Computational Complexity Analysis

In addition to recognition performance, the computational efficiency of the proposed system is analyzed. The LBP preprocessing step is computationally lightweight, as it involves simple local neighborhood operations. The score-level fusion using the weighted sum rule introduces minimal computational overhead. The overall computational cost is primarily dominated by the CNN-based feature extraction stage, which remains consistent across both raw and LBP-based inputs. Therefore, the inclusion of LBP features and fusion does not significantly increase the computational burden of the system. This analysis indicates that the proposed approach is computationally efficient and suitable for practical deployment, including real-time or resource-constrained environments with optimized implementation.

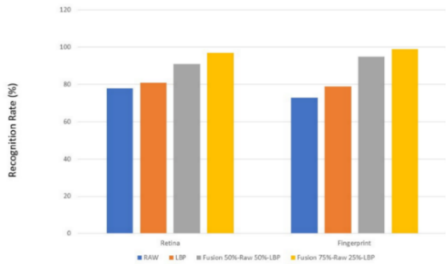
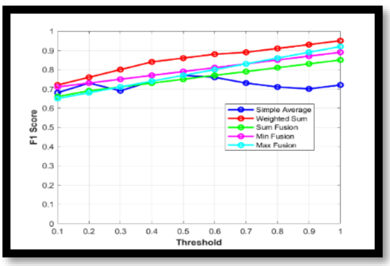


Figure 8

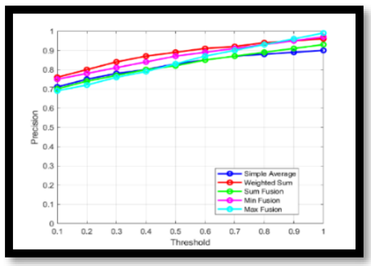
Performance Comparison of Raw Images, LBP Features and Fusion Techniques

4.6 Performance Metrics

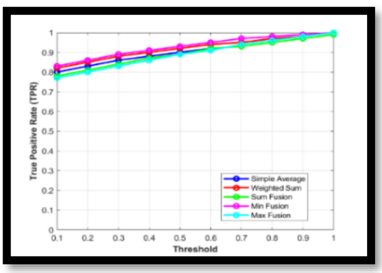
To further evaluate the effectiveness of the proposed multi-modal biometric system, additional performance metrics including Precision, Recall, True Positive Rate (TPR), False Positive Rate (FPR), F1-Score and Accuracy are analyzed for the weighted sum fusion technique.



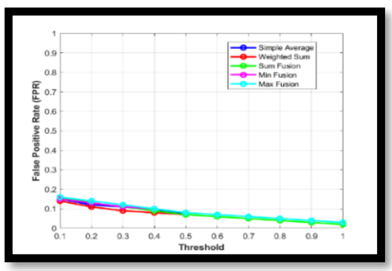
(a) F1 Score



(b) Precision



(c) TPR



(d) FPR

Figure 9

Comparison of the performance metrics for fusion models

Table 6 presents the threshold-wise performance analysis of the weighted sum fusion method. The results show consistent improvement in precision, F1-score and overall accuracy with the increasing threshold values, while the FPR decreases gradually. At the optimal threshold setting, the proposed system achieves high recognition performance with an accuracy of 99%, precision of 1.00 and F1-score of 1.00, demonstrating reliable biometric authentication capability. Figures 9 illustrates the comparative performance of different fusion models in terms of F1-score, Precision, TPR and FPR. The weighted sum fusion method consistently achieves superior performance compared to the other score-level fusion methods. The above results confirm that the proposed multi-modal biometric system effectively combines retinal and fingerprint modalities to achieve improved authentication accuracy, robustness and security.

Table 6

Performance Metrics of the Weighted Sum Fusion Method

Threshold	TPR	FPR	Precision	F1 Score	Accuracy
0.1	0.82	0.12	0.75	0.78	0.82
0.2	0.85	0.10	0.80	0.82	0.87
0.3	0.87	0.08	0.82	0.84	0.90
0.4	0.89	0.07	0.85	0.87	0.91
0.5	0.92	0.05	0.88	0.90	0.92
0.6	0.94	0.03	0.90	0.92	0.93
0.7	0.95	0.03	0.91	0.93	0.94
0.8	0.97	0.02	0.94	0.95	0.97
0.9	0.98	0.01	0.95	0.96	0.98
1.0	1.00	0.00	1.00	1.00	0.99

In addition, the performance across different threshold values shows low variation, indicating that the proposed method maintains stable behavior under varying decision conditions. This suggests that the observed improvements are consistent and not dependent on a specific threshold setting, thereby supporting the reliability of the proposed fusion approach. The results demonstrate the effectiveness of integrating retinal and fingerprint traits within a multi-modal biometric system. Based on the observations from Table 3 and Figure 7, all the score-level fusion methods significantly improve recognition performance compared to uni-modal systems. The retinal modality achieves an EER of 3.592% while the fingerprint modality produces an EER of 4.836%. Among the fusion methods evaluated, the weighted sum rule achieves the best performance with a GAR of 98.76% at FAR = 0.01 and an EER of 1.123%. An optimal performance was obtained when equal weights were assigned to retinal and fingerprint modalities (W_1=0.5, W_2=0.5), highlighting the complementary nature of both the biometric traits.

4.7 Comparative Analysis with the Existing Methods

Table 7 presents a comparative analysis of the proposed multi-modal biometric authentication system with the existing systems reported in literature. Earlier studies have employed different biometric modalities, including iris, fingerprint, face, finger-vein and retina using feature-level, decision-level and score-level fusion methods. Recent approaches have increasingly used deep learning-based feature extraction and fusion strategies to improve recognition accuracy and authentication reliability. The proposed method achieves an EER of 1.123% and an accuracy of 98.81%, outperforming several existing multi-modal biometric systems. However, direct comparison with existing methods should be interpreted carefully due to differences in datasets, experimental protocols and biometric modalities used across studies.

Table 7

Comparative Analysis of the Proposed Multi-modal Method and the existing methods

Methods	Modalities	Database	Fusion Methods	EER (%	Acc (%)
Jagadiswary et al. [12]	Fingerprint + Retina + Finger Vein	Self-collected Dataset	Feature-Level Fusion + RSA	—	95.30
Walia et al. [4]	Fingerprint + Finger Vein + Iris	IITD + PolyU + CASIA	Score-Level Fusion	1.57	96.8%
Choraś et al.	Iris + Retina	Private Database	Multimodal Authentication	—	93.20
Mustafa et al. [13]	Iris + Fingerprint	CASIA + FVC	GLCM + KNN + Decision Rules	—	95.00
Gavisidappa et al. [17]	Face + Iris + Fingerprint	Virtual Multi-modal Dataset	Handcrafted Features + Multi-SVM	—	97.09
Mane et al. (2025) [26]	Iris + Fingerprint	Multi-modal Paired Dataset	CNN + ResNet50 + Score-Level Fusion	—	96.80
Proposed method	Retina + Fingerprint	RIDB, VARIA, DRIVE and STARE and FVC2000/FVC2002/FVC2004	CNN + LBP + Weighted Score-Level Fusion	1.123	98.81

Conclusions

This paper presents a multi-modal biometric authentication system integrating retinal and fingerprint modalities using score-level fusion techniques. The proposed system combines LBP and CNN-based feature extraction to capture both texture-based and high-level discriminative features. Several normalization and fusion methods are evaluated to improve score compatibility and recognition performance. Experimental results demonstrate that the proposed multi-modal system significantly outperforms uni-modal biometric systems. Among the fusion methods evaluated, the weighted sum rule achieves the best performance, with an EER of 1.123% and an overall recognition accuracy of 98.81%. These results confirm that the integration of retinal and fingerprint traits improves recognition accuracy, robustness and authentication reliability. However, the current study is based on a virtual multi-modal dataset constructed from independent retinal and fingerprint databases due to the limited availability of real paired datasets, which may limit the ability to fully capture inter-modal relationships observed in real-world scenarios. Future work will focus on validating the proposed system using real paired multi-modal biometric datasets for more realistic performance evaluation. In addition, adaptive and data-driven fusion strategies will be explored to further enhance system flexibility and robustness under varying conditions. The system will also be extended to support efficient real-time biometric applications.

References

[1] Ross, A. A., Jain, A. K., & Nandakumar, K. (2006). Handbook of multibiometrics. Boston, MA: Springer US.

[2] Gavrilova, M. L., & Monwar, M. (Eds.). (2013). Multimodal biometrics and intelligent image processing for security systems. IGI Global.

[3] Asem, Morteza Modarresi, and Iman Sheikh Oveisi. "Biometric retinal authentication based on multi-resolution feature extraction using mahalanobis distance." Biometrics and Biostatistics International Journal 7(1): 28-46 (2018).

[4] Walia, G. S., Singh, T., Singh, K., & Verma, N. (2019). Robust multimodal biometric system based on optimal score-level fusion model. Expert Systems with Applications, 116, 364-376.

[5] Win, K. N., Li, K., Chen, J., Viger, P. F., & Li, K. (2020). Fingerprint classification and identification algorithms for criminal investigation: A survey. Future Generation Computer Systems, 110, 758-771.

[6] Moujahdi, C., Bebis, G., Ghouzali, S., & Rziza, M. (2014). Fingerprint shell: Secure representation of fingerprint template. Pattern Recognition Letters, 45, 189-196.

[7] Thangaraj, Sumathi, Vivekanandan Periyasamy, and Ravikanth Balaji. "Retinal vessel segmentation using neural network." IET Image Processing 12, no. 5 (2018): 669-678.

[8] Huo, G., Liu, Y., Zhu, X., Dong, H., & He, F. (2015). Face–iris multimodal biometric scheme based on feature level fusion. Journal of Electronic Imaging, 24(6), 063020-063020.

[9] Chergui, O., Bendjenna, H., Meraoumia, A., & Chitroub, S. (2016, March). Combining palmprint & finger-knuckle-print for user identification. IT4OD, pp. 1-5. IEEE..

[10] Praseetha, V.M., Bayezeed, Saad and Vadivel, S. "Secure Fingerprint Authentication Using Deep Learning and Minutiae Verification" Journal of Intelligent Systems, vol. 29, no. 1, 2020, pp. 1379-1387.

[11] Srivastava, R. (2020). Score-level multimodal biometric authentication of humans using retina, fingerprint, and fingervein. International Journal of Applied Evolutionary Computation (IJAEC), 11(3), 20-30.

[12] Jagadiswary, D., & Saraswady, D. (2016). Biometric authentication using fused multimodal biometric. Procedia Computer Science, 85, 109-116.

[13] Mustafa, A. S., Abdulelah, A. J., & Ahmed, A. K. (2020). Multimodal biometric system iris and fingerprint recognition based on fusion technique. International Journal of Advanced Science and Technology, 29(03), 7423-7432.

[14] Song, H. K., AlAlkeem, E., Yun, J., Kim, T. H., Yoo, H., Heo, D., ... & Yeob Yeun, C. (2020). Deep user identification model with multiple biometric data. BMC bioinformatics, 21(1), 315.

[15] Aizi, K., & Ouslim, M. (2022). Score-level fusion in multi-biometric identification based on zones of interest. Journal of King Saud University-Computer and Information Sciences, 34(1), 1498-1509.

[16] Ojala, T., Pietikainen, M., Maenpaa, T. (2002). "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." IEEE Trans Pattern Anal Machine Intell, 24(7), 971–987.

[17] Gavisiddappa, G., Mahadevappa, S., & Patil, C. (2020). Multimodal biometric authentication system using modified ReliefF feature selection and multi support vector machine. International Journal of Intelligent Engineering and Systems, 13(1), 1-12.

[18] Jothibasu Marappan, Karthik Murugesan, Malar Elangeeran, and Usha Subramanian (2023) "Human retinal biometric recognition system based on multiple feature extraction," Journal of Electronic Imaging 32(1), 013008

[19] Elhoseny, M., Elkhateb, A., Sahlol, A., & Hassanien, A. E. (2018). Multimodal biometric personal identification and verification. Advances in Soft Computing and Machine Learning in Image Processing, 249-276.

[20] Vishi, K., & Mavroeidis, V. (2018). An evaluation of score-level fusion approaches for fingerprint and finger-vein biometrics. arXiv preprint arXiv:1805.10666.

[21] He, M., Horng, S. J., Fan, P., Run, R. S., Chen, R. J., Lai, J. L., ... & Sentosa, K. O. (2010). Performance evaluation of score-level fusion in multimodal biometric systems. Pattern Recognition, 43(5), 1789-1800.

[22] Cheniti, M., Boukezzoula, N. E., & Akhtar, Z. (2018). Symmetric sum-based biometric score fusion. IET Biometrics, 7(5), 391-395.

[23] Hamd, M. H., & Rasool, R. A. (2020, November). Score-level fusion technique for Human identification. In IOP Conference Series: Materials Science and Engineering (Vol. 990, No. 1, p. 012021). IOP Publishing.

[24] Rasool, R. A. (2021). Feature-level vs. score-level fusion in the human identification system. Applied Computational Intelligence and Soft Computing, 2021, 1-10.

[25] R. Raghavendra, Bernadette Dorizzi, Ashok Rao, & Hemantha Kumar (2011). Designing Efficient Fusion Schemes for Multimodal Biometric System using Face and Palmprint, Pattern Recognition, 44(5), 1076-1088.

[26] Mane, J. S., Bhosale, S., & Urooj, S. (2025). A Deep Learning-Based Multimodal Biometric Authentication Framework Using Fingerprint and Iris with Score-Level Fusion. Ingénierie des Systèmes d'Information, 30(12).

[27] https://data.mendeley.com/datasets/tjw3zwntv6/1

https://www.isi.uu.nl/research/databases/

http://bias.csr.unibo.it/fvc2000/databases.asp

[28] El_Rahman S., Alluhaidan A. S. "Enhanced multimodal biometric recognition systems based on deep learning and traditional methods in smart environments." PLoS ONE. 2024;19(2):e0291084.