A Multimodal Biometric System for Iris and Face Traits Based on Hybrid Approaches and Score Level Fusion

Ola Najah Kadhim¹*, Mohammed Hasan Abdulameer² and Yahya Mahdi Hadi Al-Mayali³

¹Department of Computer Science, Faculty of Computer Science & Mathematics, University of Kufa, Iraq
²Department of Computer Science, Faculty of Education for Women, University of Kufa, Iraq
³Department of Computer Techniques Engineering, College of Technical Engineering, University of Alkafeel, Najaf, Iraq

Abstract. The increasing need for information security on a world scale has led to the widespread adoption of appropriate rules. Multimodal biometric systems have become an effective way to increase recognition precision, strengthen security guarantees, and reduce the drawbacks of unimodal biometric systems. These systems combine several biometric characteristics and sources by using fusion methods. Through score-level fusion, this work integrates facial and iris recognition techniques to present a multimodal biometric recognition methodology. The Histogram of Oriented Gradients (HOG) descriptor is used in the facial recognition system to extract facial characteristics, while the deep Wavelet Scattering Transform Network (WSTN) is applied in the iris recognition system to extract iris features. Then, for customized recognition classification, the feature vectors from every facial and iris recognition system are fed into a multiclass logistic regression. These systems provide scores, which are then combined via score-level fusion to maximize the efficiency of the human recognition process. The realistic multimodal database known as (MULB) is used to assess the suggested system's performance. The suggested technique exhibits improved performance across several measures, such as precision, recall, accuracy, equal error rate, false acceptance rate, and false rejection rate, as demonstrated by the experimental findings. The face and iris biometric systems have individual accuracy rates of 96.45% and 95.31% respectively. The equal error rates for the face and iris are 1.79% and 2.36% respectively. Simultaneously, the proposed multimodal biometric system attains a markedly enhanced accuracy rate of 100% and an equal error rate as little as 0.26%.

1 Introduction

The necessity of information security, which seeks to protect data from unintentional access has resulted in the common usage of traditional user authentication techniques like usernames, passwords, and keys. These conventional approaches, however, have drawbacks, including being susceptible to theft, duplication, loss, or cracking. By collecting biometric information about an individual's physical characteristics and behaviour, biometric security which relies on a pattern recognition system offers a potentially viable approach to information security [1] Many biometric modalities have been developed in the modern era, such as the face [2], palmprint [3], iris [4], and handwritten signature [5]. The real-world performance of unimodal biometric systems that rely solely on one source of information might be hampered by limitations such non-universality, lack of uniqueness, and sensitivity to noisy input. Multimodal biometric systems, on the other hand, combine information from several modalities to improve decision-making and successfully thwart spoofing efforts. These systems perform better than unimodal systems because they integrate data from several sources [6]. Multiple modalities, units, sensors, and representations can be used by multimodal biometric systems, with fusion taking place at various levels [7, 8]. The rationale behind selecting the face and iris for this study is their uniqueness, data accessibility, and long-term stability. Faces are ubiquitous and so provide a basic means of human recognition [9], but iris patterns are less variable and more stable, which enhances the dependability of iris recognition systems [10]. Pre-processing, feature extraction, matching, and decision-making stages are all involved in traditional biometric systems [11]. System performance is greatly impacted by feature extraction techniques [12]; in this work, deep learning is used to iris images while traditional methods are used for face images. The study compares the performance of unimodal and multimodal biometrics in terms of recognition, and it presents a framework that combines

*Corresponding author: ola.najah@atu.edu.iq

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
iris and face at the match score level. Originally, unimodal biometric systems were created for distinct iris and face modalities. To extract features, they used techniques such as deep Wavelet Scattering Transform Network for irises and Histogram of Oriented Gradients (HOG) for faces. After that, a multi-class logistic regression machine learning technique is used to carry out the classification. In order to create a single score vector for the ultimate decision, the mean rule approach is utilized in the proposed fusion strategy at the score level to merge match score vectors from each unimodal system. The following can be used to define the contributions made in this paper:

1. Establishing a multimodal biometric framework at the match score level that combines face and iris.
2. A comparison of the recognition abilities of unimodal biometric systems and multimodal modes.
3. The extraction of features from the iris and face using deep learning and conventional learning techniques, respectively.
4. Applying a fusion technique at the score level to enhance human recognition decision-making.

The subsequent sections of the paper are organized according to the following outline: Section 2 provides a succinct summary of pertinent studies. Section 3 outlines the approach utilized in this study. Section 4 explores a comprehensive examination and evaluation of the acquired outcomes. Section 5 functions as the concluding section.

2 Literature review

Several academic studies have investigated the concept of multimodal biometric systems, including several recognition techniques. This section provides a critical analysis of the literature that has been written about multimodal biometric systems using both traditional machine learning techniques and deep learning methodology.

A multimodal verification technique incorporating facial and iris features through feature-level fusion was presented by Bouzouina and Hamami [13]. Although the study used a variety of feature extraction techniques and the support vector machines (SVM) algorithm, obtaining an impressive accuracy of 98.8%, a thorough understanding of the method's effectiveness is limited by the lack of a comparative analysis with alternative fusion approaches.

Building on this, a multimodal convolutional neural network (CNN) architecture combining iris, face, and fingerprint features was developed by Soleymani et al. [14]. The research used weighted feature fusion and multi-abstract fusion algorithms to show how employing three different biometric kinds yields the best results. Nevertheless, a thorough assessment of the strategy is hampered by the lack of examination around the choice and rationale of fusion approaches.

Expanding the scope, Hamd and Mohammed [15] used a feature-level fusion strategy in conjunction with several feature extraction techniques to improve the accuracy of a face-iris human recognition system. Although many approaches yield a high accuracy rate, the study's comprehensive examination of the constraints and resilience of each feature extraction methodology is lacking, which restricts the applicability of the results.

With regard to decision-level fusion, Dwivedi and Dey [16] introduced an assessment model that made use of score level fusion, cancellable modalities, and Dempster-Shafer theory. The study demonstrated higher robustness when compared to standalone fusion techniques, but the suggested hybrid fusion framework's generalizability and dependability are called into question due to the lack of a critical analysis of the Dempster-Shafer theory's underlying assumptions and potential biases.

Moving on to feature extraction strategies, Ammour et al. [11] presented a novel approach that makes use of face-iris properties to create a multimodal biometric system. A comparison with established feature extraction methods is lacking, which limits the evaluation of the proposed technique's superiority, even though the study used a novel multi-resolution 2D Log-Gabor filter and singular spectrum analysis (SSA) in conjunction with wavelet transform for iris and facial features, respectively.

Using deep learning approaches, Yadav and Srinivasulu [17] introduced a multi-biometrics recognition system that uses attributes from offline signatures, fingerprints, and iris. The proposed framework's robustness and generalizability are hindered by the lack of a comprehensive investigation of possible biases, dataset dependencies, and the influence of hyperparameters, even if it achieves excellent accuracies at the feature and score levels.

In order to investigate yet another direction, Alay and Al-Baity [18] combined three convolutional neural network (CNN) models for face, iris, and finger vein identification. Even while feature-level and score-level fusion allowed for high accuracy, the study's findings' wider application was limited by its absence of a thorough examination of the generalization potential and potential weaknesses of the chosen CNN models.

3 Methodology

In this study, a multimodal biometric system that combines face and iris modalities integrated at the score level to recognize person is presented and shown Fig. 1. The details and a full description of the suggested system are provided in the following sections.
3.1 Data pre-processing

Prior to feeding input data into models, image pre-processing entails standardizing and improving it. For model efficiency, resizing guarantees constant image size, and grayscale conversion streamlines data. These procedures also extract certain ROIs, or areas of interest. A multi-task cascaded convolutional neural network (MTCNN) employs a cascade structure for face detection. The three phases (P-Net, R-Net, and O-Net) produce potential face regions, hone them, and classify faces. [19]. Fig. 2 illustrates the images after pre-processing.

To extract the Region of Interest (ROI) in the iris modality, one must first localize and then segment the data using Daugman’s approach in conjunction with an integro-differential operator and order statistics (maximum), as depicted in Eq. (1) [20].

\[
\max_{(r, x_0, y_0)} [G_\sigma(r) \ast \frac{1}{2\pi \sigma^2} e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}}] = \max_{(r, x_0, y_0)} [G_\sigma(r) \ast I(x, y)]
\] (1)

In this system, "I" denotes the iris image, and "I (x, y)" signifies the pixel intensity at coordinates (x, y). The Gaussian filter "G_\sigma(r)" is applied radially with radius "r" and scale "\sigma," centred at coordinates (x_0, y_0). "\sigma" represents the standard
deviation, assuming \((x_0, y_0)\) as the iris center. The symbol "ds" outlines the shape of the circle defined by parameters \((r, x_0, y_0)\). The integro-differential equation aims to identify circle boundaries crucial in image processing. The operator systematically scans the image at \((x_0, y_0)\), detecting the highest value among blurred partial derivatives. Blurring involves convolving with a Gaussian function with a scale parameter denoted as "\(\sigma\)." The operator identifies the maximum order of differentiation of the normalized contour integral along a circular arc "\(ds\)" with center \((x_0, y_0)\) and radius "\(r\)." This operator attends as a circular edge detector, blurring with scale "\(\sigma\)," enhancing research at smaller sizes via repetitive exploration in the parameter space \((x_0, y_0, r)\). Iris segmentation normalization involves converting the segmented iris from Cartesian \((x, y)\) coordinates to polar \((r, \theta)\) coordinates using Daugman’s Rubber Sheet model [21]. The circular iris texture is transformed into a rectangular shape via this transformation, which is explained in full in Eq. (2), Eq. (3), and Eq. (4). The pixel transformation fits the equation-specified design.

\[
I(x(r, \theta), y(r, \theta)) \rightarrow I(r, \theta) \tag{2}
\]
\[
x(r, \theta) = (1 - r)x_p(\theta) + rx_i(\theta) \tag{3}
\]
\[
y(r, \theta) = (1 - r)y_p(\theta) + ry_i(\theta) \tag{4}
\]

Where \(I(x, y)\) the iris region image, \((x, y)\) are original Cartesian coordinates, \((r, \theta)\) are the corresponding normalized polar coordinates, and \(x_p, y_p\) and \(x_i, y_i\) are pupil and iris boundary coordinates along the \(0\) direction. Fig 3 illustrates the iris characteristic processing procedures.

![Fig. 3. Iris pre-processing steps](image)

### 3.2 Features extraction

Features are retrieved from face and iris images after post-processing in this stage of the suggested approach. The extraction of features from face images is done using the Histogram of Oriented Gradient (HOG) approach, whereas the extraction of features from iris images is done using the Wavelet Scattering Transform Network (WSTN). These techniques are explained in more detail in the sections that follow.

#### 3.2.1 HOG features descriptor

As a feature descriptor, the Histogram of Oriented Gradients (HOG) describes the local look and form of an image. By focusing on local gradient information across different image areas, HOG captures structural and morphological characteristics, which makes it useful for object recognition. HOG feature extraction is commonly used in object recognition and image classification [22]. The methods involved are as follows:

1. **Image normalization:** Before being converted to a specific color space—typically grayscale—the input image is resized to a predetermined size and normalized. By doing this, you can improve the HOG feature’s resistance to variations in illumination and image size.

2. **Computation of gradients:** The gradient’s magnitude and direction are determined for every pixel in the image. Eq. (5), Eq. (6), and Eq. (7) define the gradient direction as edge orientation and the gradient magnitude as edge strength. This calculation is usually performed using Sobel filters.

\[
G_x = \frac{\partial I}{\partial x} \quad G_y = \frac{\partial I}{\partial y} \tag{5}
\]
Consider I as the image, where Gx and Gy represent the partial derivatives of the image intensity in the x and y directions, respectively.

3. Calculation of the cell histogram: The image is divided into a grid of 16-sized cells. The gradient direction and magnitude are used to calculate an oriented gradient histogram for each pixel in a cell. There are 16 bins in this histogram, which represent 16 different gradient directions.

4. Block normalization: To lessen the effect of fluctuations in illumination, cells are arranged into blocks, and the histograms of oriented gradients for the cells within a block are normalized.

5. Descriptor construction: The HOG descriptor, also called the HOG feature vector, is created by concatenating the normalized histograms of oriented gradients for each block in the image. Fig. 4 shows how the HOG feature descriptor is applied to a facial image.

\[
\text{Magnitude} = \sqrt{G_x^2 + G_y^2} \quad \text{(6)}
\]

\[
\text{Orientation} = \tan^{-1}\left(\frac{G_y}{G_x}\right) \quad \text{(7)}
\]

\[
\text{Fig. 4. Face features descriptors}
\]

3.2.2 WSTN features extraction

The scattering network is a variant of deep convolutional neural networks utilizing wavelet transforms as its filtering mechanism. This architecture enhances the network’s ability to remain unaffected by various transformations, such as translation and rotation [23]. The scattering network creates the multi-layered representation of the input image. Each layer contains local descriptors that are obtained by a series of three operations: wavelet decompositions, complex modulus computation, and local averaging.

Take an image denoted as ‘I’ into consideration. The initial scattering coefficient, representing the image average, is determined by convolving the image with a filter called the averaging filter (low-pass filter) \( \varphi_J \), as expressed in Eq. (8) [24].

\[
S_{0,J}(I) = I \ast \varphi_J \quad \text{(8)}
\]

The initial layer’s scattering coefficients can be obtained by applying wavelet transforms across different scales and orientations. Subsequently, the complex phase is removed, and the average is computed utilizing \( \varphi_J \), as illustrated in Eq. (9) [24].

\[
S_{1,J}(I) = |I \ast \psi_{j1,1}| \ast \varphi_J \quad \text{(9)}
\]

The variables \( j_1 \) and \( \lambda_1 \) denote distinct scales and orientations. The determination of the magnitude of wavelet coefficients can be compared to the utilization of non-linear pooling functions that bear resemblance to the ones present in convolutional neural networks. It’s important to highlight that eliminating the complex stage of wavelet coefficients imparts invariance to local translation, but it also leads to the loss of some high-frequency content due to the averaging effect.

To retrieve the high frequency components removed by the wavelet coefficients in the first layer, one can convolve \( |I \ast \psi_{j1,1}| \) with other set of wavelets at a scale \( j_2 < J \). Subsequently, taking the absolute value of the resultant wavelet coefficients is tracked by averaging, as outlined in Eq. (10) [24].

\[
S_{2,J}(I) = |I \ast \psi_{j1,1}| \ast |\psi_{j2,2}| \ast \varphi_J \quad \text{(10)}
\]

The coefficients of the k-th layer of the scattering network can be obtained by continuing this approach as in Eq. (11) [24]:
\[ S_{k,j}(I) = \left| I \ast \psi_{k,j} \right| \ast \ldots \ast \left| I \ast \psi_{k,j} \right| \ast \varphi_j \quad (11) \]

The final output vector ‘S’ contains information on the local spatial arrangement of texture characteristics, providing useful information for various image analysis and classification tasks. The coefficients produced at different scales and orientations make up the characteristics obtained from the scattering transform. These coefficients provide a meaningful representation for further processing inside the neural network by efficiently encapsulating information about the structure of the input image across several scales.

3.2.3 Architecture of proposed WSTN

Representing the input image as I (x, y), the construction of a scattering wavelet transform involves considering the Gaussian lowpass filter \( \phi \), modulated by the scaling factor \( J \). The mother wavelet, denoted as \( \psi \), gives rise to a set of dilated and rotated variants represented by \{ \psi_{\lambda} \}, where \( \lambda = (0, j) \). Here, \( 0 \) signifies the orientation, and \( 2^j \) corresponds to the dyadic scales. The values of \( j \) belong to the set \{1, 2, ..., J\}. The visual depiction of the computation of scattering coefficients for each layer is illustrated in Fig. 5.

3.3 Classification process

In order to predict labels for input patterns, multiclass logistic regression (MLR) is used for classification once feature vectors from the face and iris have been extracted. The MLR classifier receives each vector representing a face or iris feature and uses it to produce a similarity score. A supervised learning approach called logistic regression (LR) predicts a category’s probability using explanatory factors. When dealing with situations involving multiple classes, LR can run binary LR for each class using the One-vs-Rest (OvR) technique or a SoftMax function. [25]. This stage outputs score vectors for both face and iris traits.

3.4 Score-level fusion

In multibiometric systems, score-level fusion is frequently used. The recognition results are calculated separately for every unimodal system in this approach. These recognition scores are then combined into a single multimodal system to improve system performance as a whole, as Fig. 1 illustrates. Score vectors are produced separately for the two characteristics (face and iris) by the classification procedure. The second step creates a fused score vector by applying the mean rule, which is represented by Eq. 12, to the combination of face and iris scores. The fused scores (S) and the threshold score (S\text{th}) are used to make the final conclusion. If \( S < S_{\text{th}} \), the item is deemed to be fraudulent, but if \( S > S_{\text{th}} \), it is deemed authentic.

\[
\text{mean} = \frac{1}{N} \Sigma_j^n (S_{C_f,j} + S_{C_i,j}) \quad (12)
\]

The variables SCf\_j and SCi\_j represent the score values of the face and iris biometric sample \( j \), respectively. \( N \) denotes the total number of biometric systems used.
4 Experimental Results

With Google Colab, a platform that provides Python3 in a free environment, the suggested multimodal authentication method is put into practice. A GPU with a 12GB CPU and 13GB RAM, the Nvidia Tesla k80, is used in this manner. In addition, Google Colab is used for pre-processing the face images, and MATLAB v18.a is used for processing and region of interest extraction from the iris images. Real multimodal dataset (MULB), a unique database that the authors gathered from Al-Furat Al-Awsat Technical University personnel and students in Kufa, has been used to assess the performance of the suggested multimodal biometric face-iris system. Using different proportions, we randomly partitioned each sub-dataset of the MULB multimodal dataset into training, validation, and testing subsets (80:10:10), (60:20:20), and (70:10:20) for our study. The percentage of (70:10:20) produced the best results. As a result, the data used in this study are split up into three categories: 70% are set aside for training (2464 images), 10% are set aside for validation (352 images), and 20% are set aside for testing (704 images).

4.1 Database description

MULB in [26] is a real multimodal database that includes an individual's iris, palmprint, and facial characteristics. This study's suggested multimodal biometric recognition system only makes use of the face and iris components from MULB database. The MULB database's attributes are shown in Table 1, and examples of face and iris samples are shown in Fig. 6. The iPhone 14 Pro Max's tiny camera was used to carefully record each biometric piece of information by the authors. The MULB database, which was assembled at Al-Furat Al-Awsat Technical University in Kufa, Iraq for some staff and student will soon be freely available online for scholarly and research uses.

<table>
<thead>
<tr>
<th>Name</th>
<th>The MULB database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of biometric traits</td>
<td>3</td>
</tr>
<tr>
<td>Type of biometric traits</td>
<td>Face, palmprint, iris</td>
</tr>
<tr>
<td>Color image</td>
<td>yes</td>
</tr>
<tr>
<td>Number of unique persons</td>
<td>176</td>
</tr>
<tr>
<td>Number of images per person</td>
<td>20</td>
</tr>
<tr>
<td>Gender type</td>
<td>118 males, 58 females</td>
</tr>
<tr>
<td>ages</td>
<td>from 17 to 54</td>
</tr>
<tr>
<td>Images format</td>
<td>.jpg</td>
</tr>
<tr>
<td>Total number of images</td>
<td>20<em>3</em>176=10,560</td>
</tr>
<tr>
<td>Different Conditions</td>
<td>Face: posses, expressions, and accessories</td>
</tr>
<tr>
<td></td>
<td>Palmprint: different angles</td>
</tr>
<tr>
<td></td>
<td>Iris: lighting and angles</td>
</tr>
</tbody>
</table>
4.1 Evaluation metrics

The proposed methodology is evaluated via a variety of measures shown as the following:

1. False Acceptance Rate (FAR): pertains to the probability of a system erroneously granting access to an individual who is not registered or authorized within the system. calculated by the Eq. (13) [27].

$$FAR = \frac{F_{pos}}{F_{pos} + T_{neg}}$$  \hspace{1cm} (13)

2. False Rejection Rate (FRR) indicates the probability of a system incorrectly declining access to an individual who is not registered or authorized within the system. calculated by the Eq. (14) [27].

$$FRR = \frac{F_{neg}}{T_{pos} + F_{neg}}$$  \hspace{1cm} (14)

3. Equal Error Rate (EER): the rate at which the False Acceptance Rate (FAR) is equal to the False Rejection Rate (FRR). It can be expressed as Eq. (15) [28].

$$ERR = \frac{F_{pos} + F_{neg}}{2}$$  \hspace{1cm} (15)

4. Accuracy (Acc): it represents the ratio of correctly predicted instances to the total instances in the dataset. The accuracy of a model is calculated using the Eq. (16) [3].

$$ACC = \frac{T_{pos} + T_{neg}}{(T_{pos} + T_{neg} + F_{pos} + F_{neg})}$$  \hspace{1cm} (16)

5. Precision (Pre): the proportion of correctly predicted positive instances out of all instances predicted as positive by the model. The precision formula in Eq. 17 [3].

$$Pre = \frac{F_{pos}}{(T_{pos} + F_{pos})}$$  \hspace{1cm} (17)
6. Recall (Rec): it indicates the proportion of actual positive instances that the model successfully predicted as positive. It can be expressed as Eq. (18) [9].

\[
Rec = \frac{T_{pos}}{(T_{pos} + F_{neg})}
\]

(18)

7. F1-score: the harmonic means of precision (Pre) and recall (Rec). The harmonic mean is used to give equal weight to both precision and recall. The formula for F1-score as Eq. (19) [3].

\[
F_{1\text{score}} = 2 \cdot \frac{(Pre \cdot Rec)}{(Pre + Rec)}
\]

(19)

Where Tpos the model predicted "positive," and it was correct. Tneg the model predicted "negative," and it was correct. Fpos the model predicted "positive," but it was incorrect (it should have been "negative"). Fneg the model predicted "negative," but it was incorrect (it should have been "positive").

4.2 Face unimodal experiments

The MULB face datasets were used to generate experimental results in an effort to choose the best feature vectors and enhance performance. After pre-processing, the MTCNN method was used to recognize faces in images, as shown in Fig. 2. After that, the HOG descriptor approach was used for feature extraction in order to extract useful features and produce HOG feature vectors. Face score vectors were generated from the retrieved features by using the MLR approach for classification. With an accuracy rate of 96.45%, precision of 97.24%, recall of 96.45%, and an F1-score of 96.42%, the face model performed well. Furthermore, it was found that the FAR, FRR, and ERR values were, respectively, 0.02%, 3.55%, and 1.79%.

4.3 Iris unimodal experiments

As shown in Fig. 3, we used Daugman's approach in our tests to locate, segment, and normalize each iris image from the MULB iris database. The Wavelet Scattering Transform Network (WSTN) was employed to extract pertinent characteristics from the iris image. Then, connections between neurons in the preceding and following layers were established by applying a Rectifier Linear Unit (ReLU) activation function to a completely linked layer. This layer was followed by the generation of feature vectors that were prepared for classification using the MLR technique, producing iris score vectors. With an accuracy rate of 95.31%, precision of 96.26%, recall of 95.31%, and an F1-score of 95.11%, the iris modality performed well. Furthermore, it was found that the FAR, FRR, and ERR values were, respectively, 0.03%, 4.69%, and 2.36%.

4.4 Multimodal experiments

The arithmetic mean rule described in section 3.4 was particularly used in the score-level fusion procedure to fuse the categorization scores from the iris and face models. With the help of this method, a system with 98.91% accuracy, 98.11% precision, 97.07% recall, and 98.89% F1-score were produced. Furthermore, the FAR, FRR, and ERR values were negligible, with respective values of 0.01%, 0.15%, and 0.26%.

4.5 Discussions of results

The recognition accuracy, precision, recall, and F1-score values for the unimodal and multimodal models are summarized in Table 2 and showed in Fig 7 based on the tests that were run. The findings demonstrate that, in terms of assessment metrics, the multimodal biometric model performed better than the unimodal biometric model. This emphasizes that multimodal biometrics, in its natural state, offers a highly successful approach to enhancing a biometric system's evaluation metrics.
Table 2. The outcomes of evaluating both unimodal and multimodal biometric systems' performance.

<table>
<thead>
<tr>
<th>Biometric model</th>
<th>Acc.</th>
<th>Pre.</th>
<th>Rec.</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face</td>
<td>96.45%</td>
<td>97.24%</td>
<td>96.45%</td>
<td>96.42%</td>
</tr>
<tr>
<td>Iris</td>
<td>95.31%</td>
<td>96.26%</td>
<td>95.31%</td>
<td>95.11%</td>
</tr>
<tr>
<td>Face &amp; Iris</td>
<td>98.91%</td>
<td>98.11%</td>
<td>97.97%</td>
<td>98.89%</td>
</tr>
</tbody>
</table>

Fig. 7. Performance of unimodal and multimodal biometric systems.

Table 3 and Fig. 8 show the suggested system's error rate values when examining the findings in relation to error rates (FAR, FRR, and ERR). It was discovered that the suggested multimodal system had lower mistake rates than the unimodal biometrics system. EER It is often used to evaluate the overall performance of a biometric system. Lower ERR values indicate better overall performance.

Table 3. Error rates for both unimodal and multimodal biometric systems

<table>
<thead>
<tr>
<th>Biometrics</th>
<th>FAR</th>
<th>FRR</th>
<th>ERR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face</td>
<td>0.02%</td>
<td>3.55%</td>
<td>1.79%</td>
</tr>
<tr>
<td>Iris</td>
<td>0.03%</td>
<td>4.69%</td>
<td>2.36%</td>
</tr>
<tr>
<td>Multimodal</td>
<td>0.01%</td>
<td>0.15%</td>
<td>0.26%</td>
</tr>
</tbody>
</table>
5 Conclusion

An important accomplishment has been the successful deployment of a highly efficient and effective multimodal biometric system that includes both face and iris recognition. The Histogram of Oriented Gradients (HOG) descriptor is used in conjunction with a multiclass logistic regression classifier for face recognition, and a deep wavelet scattering transform network (WSTN) is used in conjunction with the same classifier for iris detection. A reliable multimodal biometric system has been created as a consequence of the fusion of these several biometric modalities. Strategic application of the score-level fusion approach allows for the seamless integration of iris and face characteristics, demonstrating the versatility of the system. The trials, which used an actual database, highlight how much more security the multimodal biometric system provides than unimodal biometric recognition techniques. Notably, in terms of accuracy and Equal Error Rate (EER), the multimodal system outperforms separate unimodal biometric systems. 96.45% for face recognition and 95.31% for iris recognition yield an impeccable 98.91% accuracy rate for multimodal recognition. The EER for individual unimodal biometric systems, on the other hand, is 2.36% for the iris and 1.79% for the face. On the other hand, the suggested multimodal biometric system has a better EER of 0.26%. These findings demonstrate the multimodal biometric approach's effectiveness and reliability and point to its potential for improved security and accuracy in authentication systems. In the future, as there are fewer and mostly inaccessible multi-modal datasets, we plan to implement our approach in another database when it becomes available.

References