Geomatics and Information Science of Wuhan University
武汉大学学报·信息科学版

Multimodal Biometric Identification Using Yolov7 With Integrated Face, Palmprint, Fingerprint, And Iris Features

Sheetal Shrikant Shevkari [1] , Dr. Kelapati[2] , Anita Vinayak Atiwadkar[3]

[1] Research Scholar, Department of Computer Science & Engineering, SJJTU, Rajasthan, India
[2] Research Guide, Department of Computer Science & Engineering
SJJTU, Rajasthan India
[3] MAEER'S MIT Arts Commerce and Science college Alandi Pune, Maharashtra , India

Abstract:

Feature-based image classification is a commonly used approach in human identification for multimodal biometric systems. It involves extracting distinctive features from images, such as facial features, finger print, Palm print and iris texture. This study presents a hybrid YOLOv7-based multimodal biometric recognition system that inte grates four distinct biometric traits—face, palm print, fingerprint, and iris—to enhance the accuracy, reliability, and security of identity verification. By leveraging the advanced object detection capabilities of YOLOv7 and combining features from multiple modalities, the system achieves robust recognition even under challenging conditions. Experimental results across multiple individuals demonstrate high performance in terms of accuracy, precision, specificity, and sensitivity, confirming the model's ability to minimize false acceptances and rejections. The proposed approach effectively addresses the limitations of unimodal systems by fusing complementary biometric features, making it well-suited for high-security applications such as border control, secure access, and surveillance.

Keywords: *Feature extraction; Biometrics; Iris recognition; Multimodal fusion; deep learning; Face recognition; Palmprint Recognition; Fingerprint Recognition.*

1. Introduction

In recent times, Personal authentication using biometric security plays a vital role in all over the world. Biometric means the involving of automated recognition of individuals based on the unique physical characteristics It requires at least some level of security to assure the identity. Any biometrics based authentication system is better than the traditional possession or knowledge based system. Biometric traits cannot be lost, forgotten or misplaced; hence they are easy to manage.

Unimodal biometric systems have to be contend with a variety of problems such as noisy data, intra-class variations, restricted degrees of freedom, nonuniversality, spoof attacks, and unacceptable error rates. Some of these limitations can be addressed by deploying multimodal biometric systems. In some applications, more than one biometric trait is used to attain higher security and to handle failure to enroll situations for some users. Such systems are called multimodal biometric systems [1]. Multimodal biometric systems represent two/more combined biometric recognition technologies.

A number of biometrics characteristics have been proposed till date. Several types of biometrics have been used for finding the uniqueness of individuals. These techniques are like fingerprints, palm prints, finger veins, hand veins, face veins, palm veins, face shapes, palates, iris, DNA recognition, facial expression, gait signature, voice recognition, body language, heartbeat and foot veins etc [2]. Biometrics is considered as a currently ongoing scientific research topic with many applications, regarding safety and convenience [1]

Physiological Based Traits:

This section discuss biometric traits which are based on physiological characteristics. Face is the most common and non-intrusive trait and hence, can be captured easily using cheap sensors. Most commonly geometric distances between facial key-points such as eyes, nose, mouth. Face recognition is non-intrusive but due to several challenges like pose, illumination, occlusion, aging and expression, its performance is found to be restricted [3]-[4]. The face pose varies with the viewing angle along with illumination. The facial expressions can deform the face significantly while partial occlusion may hide some facial regions. Aging is a big challenge to deal with because facial features do not vary in a specific pattern.

Palm print recognition has emerged as a highly accepted biometric system due to its easy acquisition and reliability. Palm is the inner surface of hand between wrist and fingers. The inner surface of palm contains three flexion creases, secondary creases, and ridges for each finger. Palm features are unique for every individual and have rich information that can be used for feature extraction. The palm lines and wrinkles are formed during the third and fifth month of the formation of the fetus. Variation of illumination and rotation as well as translation and handling the problem of occlusion are the major challenges [5].

Any typical fingerprint is made up ridges. The ridge ending and the ridge bifurcation are known as minutia features. These features are assumed to be unique and stable. The poor quality of fingerprints may be due to poor quality of sensor or external factors such as dirt, oil, sweat etc [6]. Lesser amount of user cooperation is required for its acquisition and can be captured through cheap sensors.

Iris is basically a donut shape annular region that is bounded by sclera and pupil. There exist discriminative textures within iris in the form of furrows, ridges, crypts. Iris posseses highly discriminative unique texture that is naturally well protected. Accurate iris segmentation in varying illumination is very challenging. The eyelid and eyelash along with motion blur and specular reflection is another issue[3]-[7].
Objectives of the paper are as follows,

i. Utilize YOLOv7's model capabilities to quickly and accurately detect and localize biometric regions (face, palm, fingerprint, iris) from input images or video streams.
ii. Develop a single detection architecture capable of identifying multiple biometric modalities simultaneously, reducing the need for separate detection systems.
iii. Improve overall biometric recognition performance by combining features from multiple modalities, leveraging the high precision of YOLOv7 for region-of-interest (ROI) extraction.
iv. Use YOLOv7's backbone and neck architecture to extract discriminative features from each detected biometric region for subsequent recognition or classification tasks.

The article is organized as, section 2 contains the related works, section 3 contains challenges of former methods, section 4 give the objectives of the proposed method, section 5 describes all methods involved in the recognition of the face palmprint, fingerprint, and iris biometric, section 6 elaborates the results obtained through the proposed research, and section 7 states the conclusion and the future works.

2. Literature survey

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.

H. Singh et.al.[8] presented a novel multimodal biometric, encryption and watermarking-based method for digital image security in which multimodal biometric features are extracted and fused using the customized deep learning model. Y. Qiao et.al.[9] propose a novel method for the precise decoupling of hand multimodal features by accurately segmenting various hand regions with diverse backgrounds based on low-level details and high-level semantic information and construct an authentication system based on this method. S. Vatchala et al.[10] introduce a novel multi-modal biometric authentication system that integrates facial, vocal, and signature data to enhance security measures. Utilizing a combination of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), uniquely incorporates dual shared layers alongside modality-specific enhancements for comprehensive feature extraction.

I. Riaz, et.al. [11] proposes a multimodal biometric system based on hand modalities, which combines the extracted information of dorsal finger crease and finger knuckle print. A. Hayat et.al.[12] proposes a multimodal authentication approach fusing fingerprint, ear, and face biometrics, in which dynamic weightage is assigned to each of the matching systems, calculated with experimental analysis using a Genetic algorithm.

Byeon, H. et.al.[13] presents a novel multimodal biometric fusion model that significantly enhances accuracy and generalization through the power of artificial intelligence. Various fusion methods, encompassing pixel-level, feature-level, and score-level fusion, are seamlessly integrated through deep neural networks. Hattab, A et.al.[14] proposed a robust multimodal biometric recognition system based on the fusion of the face and both irises modalities using YOLOv4-tiny to detect regions of interest and a new effective Deep Learning model inspired by the Xception pre-trained model to extract features.

V. D. Babu et.al.[15] proposes a hybrid multimodal biometric recognition system (HMBRS) combining complex features of iris, face, and finger vein traits using deep learning techniques to overcome challenges like environmental fluctuations and spoofing vulnerability in Biometric Recognition Systems. R. Sharma et.al. [16] analyzed the principles and restrictions of unimodal identity recognition and performance of various techniques such as Convolutional Neural Network (CNN), Support Vector Machine(SVM), Hybrid Localized model, and deep learning algorithms compared with Hybrid Localized Convolutional Neural Network (HLCNN). L. Singh et.al.[17] presents a novel multi-model biometric system such as fingerprints, face, iris, and voice, leveraging deep learning (DL) techniques. To classify data from various biometric sources, the suggested framework uses Bi-directional Long Short-Term Memory (Bi-LSTM).

Multimodal biometric recognition methods face several significant challenges despite recent advancements. Many systems involve complex fusion techniques and deep learning architectures, which result in increased computational overhead and latency, making real-time implementation difficult. Some approaches rely heavily on precise segmentation of biometric regions such as the hand, iris, or face, which can be sensitive to variations in background, lighting, or pose. Modalities like voice and signature are prone to noise and variability in unconstrained environments, affecting the robustness of authentication. Fusion strategies based on genetic algorithms or dynamic weighting add additional complexity and may not generalize well across different user populations. Systems that integrate multiple modalities often demand high-quality input from each sensor, which can be a limitation in low-resource or mobile settings. Moreover, deep neural networks used for multimodal fusion require large annotated datasets and extensive training, which may not be feasible for all applications. Scalability, overfitting, and interoperability with existing systems also remain key challenges in deploying these methods effectively in real-world scenarios.

Here are challenges of traditional multimodal biometric recognition involving face, palm print, fingerprint, and iris:

a) Fusion Complexity:

i. Feature-level fusion requires alignment of heterogeneous data.
ii. Score-level fusion must normalize different confidence measures.
iii. Decision-level fusion may reduce discriminative power compared to lower-level fusion.

b) Overfitting and Generalization

i) Risk of overfitting due to limited data or unbalanced modality representation.
ii) Poor generalization to unseen users or real-world conditions.

c) Modality Compatibility

i) Each modality has different characteristics (e.g., texture, structure).
ii) Designing a unified model architecture to effectively extract features from all modalities is challenging.

d) Preprocessing and Alignment:

i) Each modality requires specialized preprocessing (e.g., segmentation, alignment).
ii) Misalignment can lead to degraded model performance.

e) Cross-Modality Feature Representation

i) Difficulty in learning shared feature spaces that capture complementary information from all modalities.
ii) CNNs trained on one modality may not transfer well to another.

3. Methodology

In this study, a multimodal biometric system that combines face, fingerprint, palmprint 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.

In YOLOv7, internal feature extraction is a multi-stage process designed to efficiently capture and process visual information for accurate object detection. It begins with the backbone network, an enhanced version of CSPDarknet, which incorporates Cross Stage Partial (CSP) connections and convolutional layers to extract rich hierarchical features from the input image. The process starts with a Focus layer that reduces the spatial dimensions while increasing channel depth, allowing the model to preserve fine-grained details such as edges and textures—crucial for biometric traits like fingerprint ridges and iris patterns. As the data flows through the backbone, multiple layers extract features at different abstraction levels, from low-level details to high-level semantic information.

Next, the neck of the network, composed of Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) modules, fuses features from multiple scales. This fusion enables the model to maintain accuracy across objects of varying sizes—important for detecting both small and large biometric regions like irises and palms. Finally, the detection head uses the aggregated features to predict bounding boxes, objectness scores, and class probabilities. For multimodal biometric recognition, each detected region (face, palm print, fingerprint, and iris) can then be cropped and further processed using dedicated recognition models. This hierarchical and multi-scale feature extraction pipeline makes YOLOv7 highly effective for real-time, accurate multimodal biometric detection.

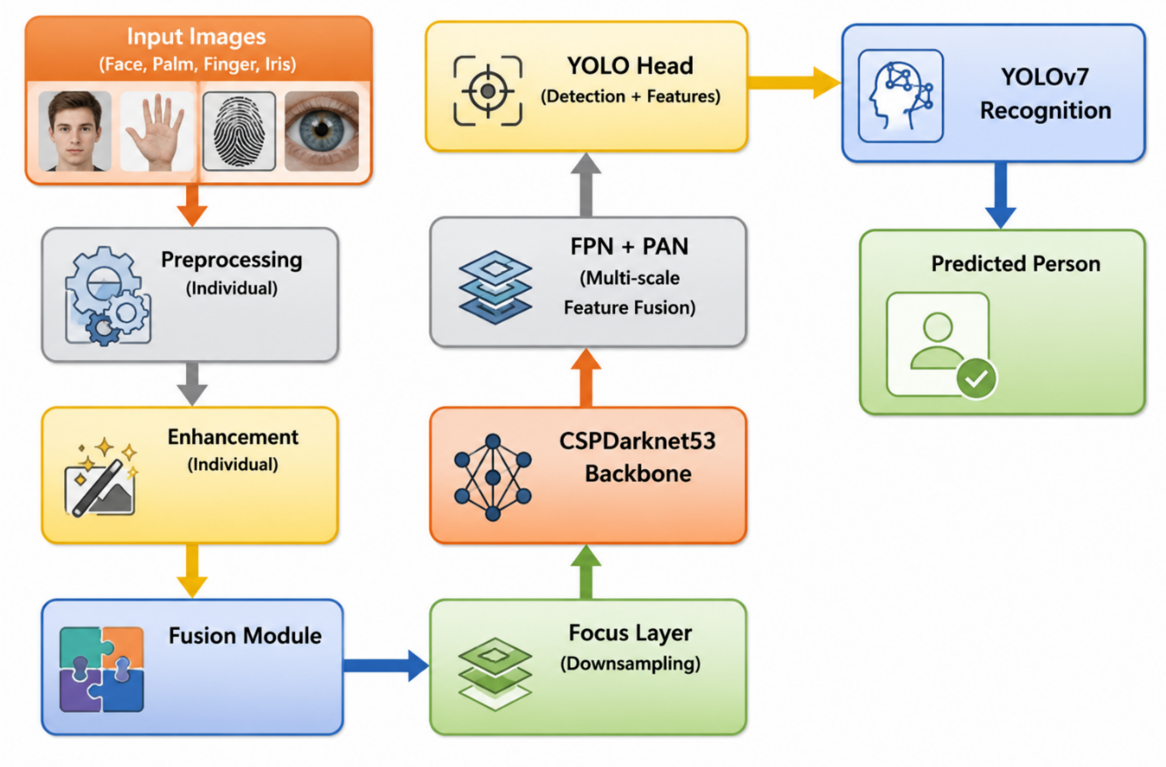

Fig. 1. Proposed YOLOv7-Based Multimodal Biometric Recognition Framework for Face, Palm, Fingerprint, and Iris Identification

Each biometric image (face, iris, palmprint, fingerprint) is processed as input:

$I \in R^{HxWxC}$ [1]

H ,W: height, width

C: number of channels (e.g., 3 for RGB, 1 for grayscale)

a) CSPDarknet53 - Backbone: Deep Feature Extraction
This is where YOLOv7 captures low-level to high-level visual features. It Captures spatial and texture features (edges, patterns, object contours).
i) CSP (Cross Stage Partial) layers: Split feature map into two parts — one is processed, one bypasses.
This reduces computation while maintaining accuracy. In this Conv + BatchNorm + SiLU (a.k.a. Swish) activations used throughout.
ii) Focus Layer (initial): Slices input into 4 and concatenates to reduce spatial size and increase channel depth.
YOLOv7 uses a CSPDarknet or ELAN module as a backbone to extract features:

$$F = f_{backbone}(I) \quad [2]$$

Where,
$f_{backbone}$ is a stack of convolution, batch norm, and activation layers:

$$F_l = Activation(BN(w_l * F_{l-1} + b_l)) \quad [3]$$

Each layer l uses convolution with learned weights w_l, bias b_l , BatchNorm and an activation function (usually SiLU or LeakyReLU .

b) Neck: Feature Aggregation
This is Used for multi-scale feature fusion (important for detecting both large and small biometric regions). It Enhances the representation of facial regions, palm lines, iris texture, and fingerprint ridges at multiple scales.
i) FPN + PAN (Feature Pyramid Network + Path Aggregation Network): Combines low-level (fine details) and high-level (semantic) features from backbone. It Merges features from different resolutions (e.g., 80×80, 40×40, 20×20). In this Lateral and up/down-sampling to ensure rich contextual representation.
Neck layers collect features from different stages:

$$F_{neck} = f_{PANet}(F) \quad [4]$$

This allows detecting small details like iris or fingerprint ridges:
Uses feature fusion:

$$F_{fused} = concat(F_{small}, \text{Upsample}(F_{large})) \quad [5]$$

c) Head: Detection Layer
This module Produces final predictions using anchor-based regression and classification. Each feature scale predicts Bounding Boxes (x, y, w, h) , Objectness score and Class probability (for biometric region categories)
YOLOv7 predicts bounding boxes and class scores:

$$\hat{y} = [x, y, w, h, c_1, \ldots, c_n] \quad [6]$$

Where:

x, y: center of box
w, h: width and height
c_i: confidence of class i (e.g., face, palm, fingerprint, iris)

This is computed per anchor box using:

$$x = \sigma(t_x) + c_x \quad [7a]$$
$$y = \sigma(t_y) + c_y \quad [7b]$$
$$w = p_w e^{t_w} \quad [7c]$$
$$h = p_h e^{t_h} \quad [7d]$$

t_x, t_y, t_w, t_h : network outputs
c_x, c_y : grid cell offset

Geomatics and Information Science of Wuhan University
武汉大学学报·信息科学版

p_w, p_h : anchor dimensions

The total loss $\mathcal{L}$ is a sum of three components:

$$\mathcal{L} = \lambda_{box} . \mathcal{L}_{box} + \lambda_{obj} . \mathcal{L}_{obj} + \lambda_{cls} . \mathcal{L}_{cls} \quad [8]$$

a. Bounding Box Loss (IoU or CIoU):

$$\mathcal{L}_{box} = 1 - \mathrm{CIoU}(\hat{B}, \mathrm{B}) \quad [9]$$

b. Objectness Loss (Binary Cross Entropy):

$$\mathcal{L}_{obj} = BCE(\hat{o}, \mathrm{o}) \quad [10]$$

c. Classification Loss (Cross Entropy):

$$\mathcal{L}_{cls} = \sum_{i=1}^{C} BCE(\hat{c_i}, c_i) \quad [11]$$

4. Experimental Results

The following figure illustrates the performance of the proposed hybrid YOLOv7-based multimodal biometric recognition system across 10 individuals. The system integrates multiple biometric modalities, likely including face, fingerprint, palm print, and iris, for enhanced identification accuracy.

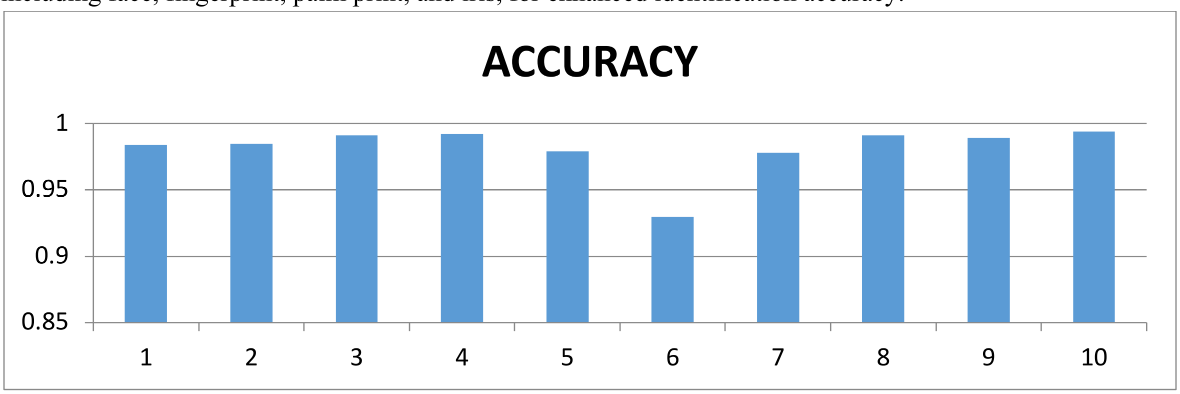

Fig 2: Accuracy of proposed method for 10 persons

The accuracy values for most individuals remain consistently high, exceeding 97%. Specifically, individuals 3, 4, 8, 9, and 10 exhibit the highest accuracy, approaching or slightly exceeding 99%, reflecting the robustness and precision of the hybrid YOLOv7 model in processing and fusing multimodal biometric data. A notable dip is observed for individual 6, where the accuracy drops to approximately 93%, indicating possible reduced feature distinctiveness in one or more biometric modalities. Nonetheless, the model maintains reliable performance across the board, with an overall average accuracy above 97%.

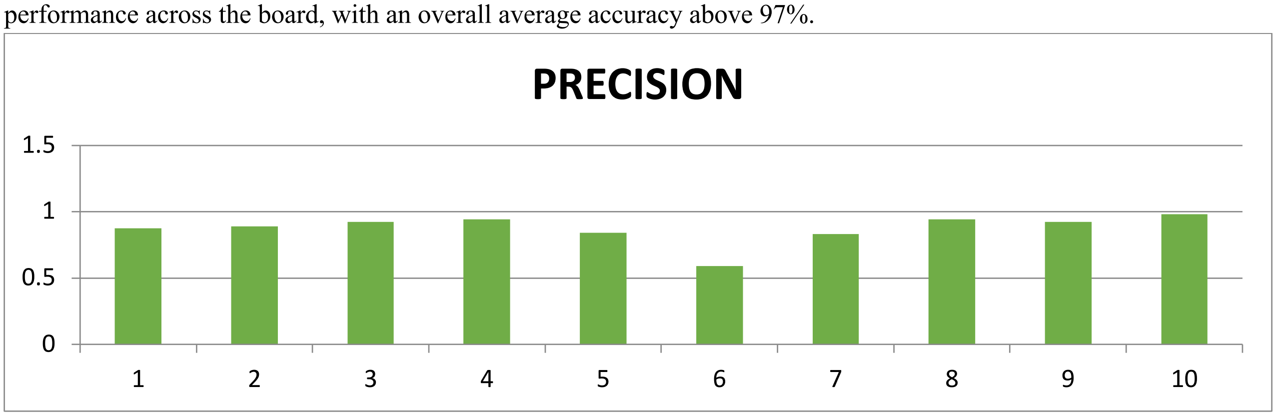

Fig 3: Precision of proposed method for 10 persons

Fig-3 presents the precision values of the proposed hybrid YOLOv7 multimodal biometric system for 10 individuals. Precision remains consistently high for most subjects, with individuals 4, 8, 9, and 10 achieving values near or above 0.95, indicating low false-positive rates. However, a noticeable drop is observed for individual 6 (around 0.6), suggesting potential misclassifications. Overall, the system demonstrates strong precision performance across most cases.

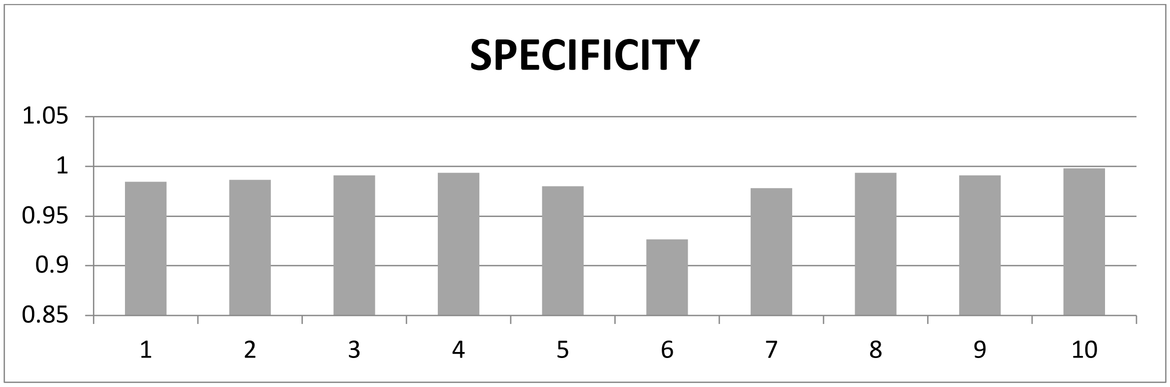

Fig 4: Specificity of proposed method for 10 persons

Fig-4 shows that the proposed hybrid YOLOv7 multimodal biometric system achieves consistently high specificity across most individuals, with values above 0.98 for the majority. The highest specificity is observed for individuals 4, 8, and 10, nearing 1.0, indicating excellent true negative rates. Individual 6 shows the lowest specificity (around 0.93), suggesting a slightly higher rate of false positives in that case.

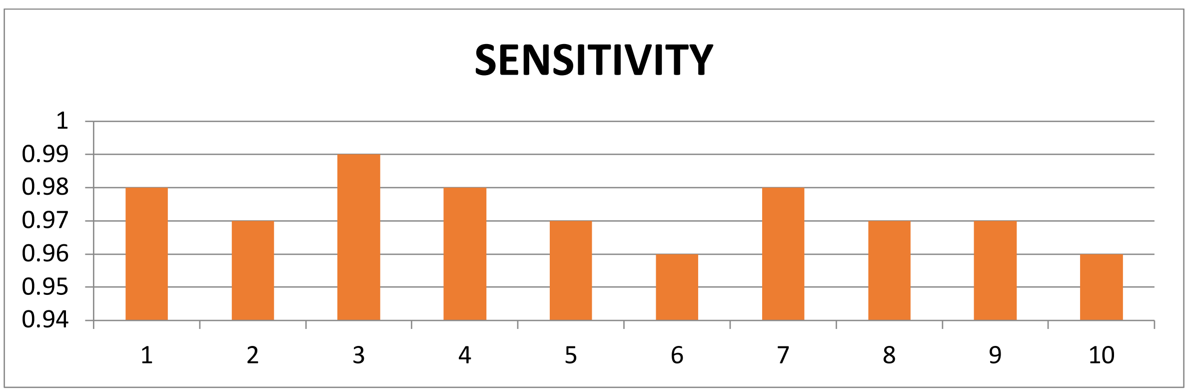

Fig 5: Sensitivity of proposed method for 10 persons

Fig-5 reflects the true positive rate of the hybrid YOLOv7 multimodal biometric system across 10 individuals. Most values remain high, especially for individuals 1, 3, 4, and 7 (around 0.98–0.99), indicating strong detection capability. The lowest sensitivity is observed for individuals 6 and 10 (approximately 0.96), suggesting slightly reduced ability to correctly identify genuine biometric matches.

This analysis confirms that the proposed hybrid YOLOv7 framework is highly effective for multimodal biometric recognition, offering excellent accuracy for diverse individuals, while also highlighting areas for potential enhancement in edge cases.

5. Discussion

Accuracy remains consistently high for all individuals, averaging above 97%, indicating that the model effectively integrates multiple biometric traits (face, fingerprint, palm print, iris) to enhance identification performance. Precision is strong overall, with most values above 0.85, reflecting a low false-positive rate. However, individual 6 showed a noticeable drop (~0.6), suggesting potential noise or data inconsistency in that case. Specificity is excellent, nearing 1.0 for most subjects, which confirms the model's ability to correctly reject imposters and minimize false alarms. Sensitivity is also robust, mostly ranging between 0.96 to 0.99, affirming the system's high capability to correctly recognize valid users. Slight dips in sensitivity for individuals 6 and 10 may indicate challenges in consistently capturing all true positives.

Overall, the hybrid YOLOv7 multimodal biometric system shows strong generalization, low error rates, and high recognition reliability, making it well-suited for secure biometric authentication applications. Minor variations in individual performance suggest room for enhancement through improved data quality or adaptive training techniques.

6. Conclusion

The overall evaluation of the proposed hybrid YOLOv7 multimodal biometric recognition system reveals promising and consistent performance across key metrics. The system demonstrates high accuracy, averaging above 97% for all individuals, highlighting its effective fusion of multiple biometric modalities—namely face, fingerprint, palm print, and iris—for reliable identification. Precision remains strong, with most values exceeding 0.85, indicating minimal false positives; however, a significant drop for individual 6 suggests possible noise or data irregularities. Specificity approaches 1.0 across most cases, showcasing the model's strong ability to reject unauthorized users and reduce false alarms. Sensitivity also performs well, with values mostly between 0.96 and 0.99, reinforcing the system's capability to detect genuine users accurately. Minor sensitivity dips for individuals 6 and 10 suggest areas for further refinement. Overall, the system proves to be accurate, dependable, and suitable for secure biometric authentication applications.

References

1. Perumal, E., & Ramachandran, S. (2015). A Multimodal Biometric System Based on Palmprint and Finger Knuckle Print Recognition Methods. International Arab Journal of Information Technology (IAJIT), 12(2).
2. Mishra, K. N., Mishra, K. N., & Agrawal, A. (2016). Veins Based Personal Identification Systems: A Review. International Journal of Intelligent Systems and Applications, 8(10), 68.
3. Jain, A.K., Ross, A.A., Nandakumar, K., Swearingen, T. (2025). Face Recognition. In: Introduction to Biometrics. Texts in Computer Science. Springer, Cham. https://doi.org/10.1007/978-3-031-61675-4_4
4. H. L. Gururaj, B. C. Soundarya, S. Priya, J. Shreyas and F. Flammini, "A Comprehensive Review of Face Recognition Techniques, Trends, and Challenges," in IEEE Access, vol. 12, pp. 107903-107926, 2024, doi: 10.1109/ACCESS.2024.3424933.
5. D. Fan, X. Liang, W. Jia, J. Chen and D. Zhang, "A Novel Hybrid Fusion Combining Palmprint and Palm Vein for Large-Scale Palm-Based Recognition," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 54, no. 7, pp. 4471-4484, July 2024, doi: 10.1109/TSMC.2024.3382877.
6. Riaz, I., Ali, A.N. & Ibrahim, H. Loss of fingerprint features and recognition failure due to physiological factors- a literature survey. Multimed Tools Appl 83, 87153–87178 (2024). https://doi.org/10.1007/s11042-024-19848-8
7. Yin, Y., He, S., Zhang, R. et al. Deep learning for iris recognition: a review. Neural Comput & Applic 37, 11125–11173 (2025). https://doi.org/10.1007/s00521-025-11109-5
8. H. K. Singh, N. Baranwal, K. N. Singh and A. K. Singh, "Using Multimodal Biometric Fusion for Watermarking of Multiple Images," in IEEE Transactions on Consumer Electronics, vol. 70, no. 1, pp. 3487-3494, Feb. 2024, doi: 10.1109/TCE.2024.3371458.
9. Y. Qiao, W. Kang, D. Luo and J. Huang, "Normalized-Full-Palmar-Hand: Toward More Accurate Hand-Based Multimodal Biometrics," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 47, no. 8, pp. 6715-6730, Aug. 2025, doi: 10.1109/TPAMI.2025.3564514.
10. S. Vatchala et al., "Multi-Modal Biometric Authentication: Leveraging Shared Layer Architectures for Enhanced Security," in IEEE Access, vol. 13, pp. 28029-28041, 2025, doi: 10.1109/ACCESS.2025.3534223.
11. I. Riaz, A. N. Ali, H. Ibrahim and I. A. Huqqani, "Multimodal Biometric Recognition System Based on Feature-Level Fusion of Dorsal Finger Crease and Finger Knuckle Print," in IEEE Transactions on Artificial Intelligence, doi: 10.1109/TAI.2025.3563315.
12. A. Hayat, A. Kumar, A. KumarBhateja and S. K. Pal, "An Approach for Multimodal Biometric Authentication using Genetic Algorithm," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-7, doi: 10.1109/ICCCNT61001.2024.10725431.
13. Byeon, H., Raina, V., Sandhu, M. et al. Artificial intelligence-Enabled deep learning model for multimodal biometric fusion. Multimed Tools Appl 83, 80105–80128 (2024). https://doi.org/10.1007/s11042-024-18509-0
14. Hattab, A., Behloul, A. Face-Iris multimodal biometric recognition system based on deep learning. Multimed Tools Appl 83, 43349–43376 (2024). https://doi.org/10.1007/s11042-023-17337-y
15. V. D. Babu, R. R. Dornala, C. Anusha, P. R. Babu, K. K. Mohan and K. V. Sumanth, "A Hybrid Multimodal Biometric Recognition System (HMBRS) based on Fusion of Iris, Face, and Finger Vein Traits," 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2024, pp. 1287-1292, doi: 10.1109/ICOSEC61587.2024.10722340.
16. R. Sharma, J. Sandhu and V. Bharti, "Exploring Feature-Based Image Classification for Human Identification in Multimodal Biometric System," 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2024, pp. 1-6, doi: 10.1109/ICRITO61523.2024.10522307.

17. L. Singh, A. Kumar and R. Golash, "Enhancing Multimodal Biometric Authentication Using Bi-LSTM Deep Learning Network," 2025 International Conference on Cognitive Computing in Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI), Greater Noida, India, 2025, pp. 363-368, doi: 10.1109/IC3ECSBHI63591.2025.10990996.