https://doi.org/10.65542/djei.v2i2.28
Dasinya Journal for Engineering and Informatics. **2026**, 2, 14.

Review

Multimodal Biometric Fusion Strategies: A Comparative Review of Current Trends, Challenges, and Future Directions

Khalid Ibrahim Jajan [1,*], Wafaa Mustafa Abduallah [2]

1 Department of Information Technology, Akre University for Applied Science, Akre, Kurdistan, Iraq; khalid.ibrahim@auas.edu.krd
2 Department of Information Technology, Duhok Polytechnic University, Duhok, Kurdistan, Iraq; wafaa.abduallah@dpu.edu.krd
* Correspondence: khalid.ibrahim@auas.edu.krd

Abstract

This paper tends to serve as a concise review of the recent literature on multimodal biometrics recognition systems, in particular, on the fusion strategies that are utilized at each stage of biometric data processing. These are feature-level, score-level, decision-level, serial, and hybrid fusions. The survey underscores that the fusion of multiple biometric traits, including fingerprints, face, iris, palmprints, voice, signature, etc., is more effective in increasing system performance, reliability, and spoof resistance than unimodal techniques. Classifying the studies based on the fusion strategies that they followed, this paper reviews the techniques, performance indicators, datasets, and application of the reviewed studies. The score-level fusion method gave the best reported accuracy of 100%, and the serial fusion obtained lower accuracy owing to the limitation of adaptability and dataset dependency. The review discusses and describes common problems and future directions for research as well. The insights drawn are targeted towards possibly the more secure, efficient and versatile multimodal biometric systems as far as the real-world applications are concerned.

Keywords: Multimodal biometric systems; biometric fusion; deep learning; identity verification; score-level fusion; feature-level fusion; biometric security.

1. Introduction

In modern security systems, user authentication is a vital element and there are many advantages for automating user authentication based on the physiological or behavioral features of users with biometrics being the best way to do this. Unimodal biometric systems use a single trait for user authentication e.g. using fingerprints, facial features or iris patterns as a trait; have numerous disadvantages which include accuracy levels that may be low, susceptibility to spoofing and performance that may be compromised in uncontrolled conditions [1]. Due to the numerous drawbacks related to unimodal biometric systems there is an increasing amount of research into multimodal biometric systems where two or more different biometric traits are utilized together to improve both the accuracy of the provided authentication and the level of difficulty to successfully exploit that authentication.

Multimodal biometric systems rely heavily on fusion strategies to combine data from many types of modalities. The different levels of fusion include feature-level, score-level, decision-level, serial, and hybrid which each carry advantages and disadvantages for specific applications or characteristics [2]. With advances in deep learning, these fusion strategies are enhanced with greater abstraction of features, flexibility of matching methods, and overall system performance.

Received: November 4, 2025; Revised: March 18, 2026; Accepted: March 18, 2026; Published: April 12, 2026; Available online: April 12, 2026.

While there have been numerous literature reviews that have comprehensively examined the various types of multi-modal biometric systems, only a small number of them classify and evaluate studies based on the type of fusion method that was utilized. As a result, there is currently a lack of literature that comprehensively reviews the multi-modal biometric systems that have been developed and/or tested between 2020 and 2025 based on the various systems of fusion—feature, score, decision, serial/parallel hybrid, etc.—and that presents a more quantitative representation via comparison between their respective methods, model architectures, biometric modalities, datasets, performance outcomes, and limitations. Therefore, this literature review aims to provide a comparative analysis of multi-modal biometric system studies in order to identify and evaluate each fusion methods' efficacy and inherent limitations; thus establishing a foundation that will facilitate future research aimed at producing more accurate and generalizable multi-modal biometric systems.

Although there have been a number of multimodal biometric investigations that have emerged in recent years, there are many unique areas of distinction in this paper that go beyond merely defining the taxonomy. First, we present a systematic and quantitative benchmark of the performance of 28 separate studies (2020 to 2025) using the same performance metrics (e.g., accuracy, EER, and AUC) to illustrate the complex trade-off among them; for example, score-based fusion methods produced the highest accuracy while feature-based fusion was the most resilient to low-quality data. Second, we provide a critical assessment of the limitations associated with datasets and identify the fact that 64% of studies use controlled laboratory datasets and also predominantly lack sufficient demographic representation—socio-technical factors have previously not been thoroughly examined in the literature. Third, we show that temporal analyses indicate 40% growth in hybrid approaches since 2022, indicating where research priorities and volumes are evolving. Fourth, we evaluate both computational complexity and real time feasibility; this is an important factor in achieving the goal of bridging academia with the industry. Fifth, the study proposes a well-structured analysis of privacy-preserving method (federated learning, homomorphic encryption) connected with specific fusion methodologies, addressing ethical aspects absent from prior surveys. And last but not least, the study attempts to point out actionable study gaps, highlighting strategy-specific pathways for lightweight structure design, adaptive fusion, and privacy preserving protocols. All these original standpoints, places this research review as both a forward-looking guide and comprehensive synthesis for developing complex, deployable, and most importantly ethically responsible multimodal biometric systems.

The leading objectives of this research review include:

1. Organizing the most recent multimodal biometric studies based on their fusion level systematically.
2. Comparatively analyzing the applied methodologies and their consequent results.
3. Identifying research gaps and challenges and examining how researchers are mitigating them to enhance multimodal biometric fusion strategies.

The remainder of this paper embraces a literature review section that categorizes and then discusses the latest studies based on the fusion level each research employs. The discussion section provides an extensive analysis of the selected academic articles, shedding light on the models applied, biometric features, datasets, performance criteria, and the most significant bottlenecks affecting the efficiency of each fusion level, while ultimately highlighting the advantages and disadvantages associated with each system as observed across the reviewed papers. This is followed by the Challenges and Limitations section, which outlines the main obstacles to the practical deployment of multimodal biometric systems. Finally, the Future Directions and Conclusion sections offer insights and recommendations for advancing this field of research and guiding future work in this area.

2. Literature Review

In the following, we review recent papers on multimodal biometric systems and different fusion techniques for biometric data at various levels. The reviewed articles were published from 2020 to 2025 and deal with different biometric modalities, namely, fingerprint, iris, face, palmprint, voice, and signature.

2.1 Paper Selection Criteria

All the study papers for this review are both systematically and transparently elicited using a protocol to identify based on specific keywords and timeframes. The literature was searched in four major databases (IEEE Xplore, SpringerLink, ScienceDirect and MDPI) using peer-reviewed publications between January 2020 and March 2025.The following keyword combinations were searched: “multimodal biometric fusion”, “feature-level fusion biometrics”," “score-level fusion biometrics”, “decision-level fusion biometrics”, “serial fusion biometrics” “hybrid fusion biometric”, “deep learning multimodal biometric” and “biometric fusion survey”.

To be considered in this research, studies must: (1) suggest or evaluate multimodal biometric systems that use a combination of at least two biometric modalities; (2) have clearly used deep learning or machine learning algorithms for fusion; and (3) provide metrics measuring the performance (accuracy, EER, AUC, etc.). Moreover, (4) publications in the English language are being considered and considered. The following are exclusion criteria: (1) biometric systems defined as unimodal; (2) publication type where the entire discussion was on either biometric acquisition or pre-processing of biometric data without mentioning multimodal fusion; (3) review articles with no original experimental data; (4) studies that do not provide adequate details about the method used in the research; and (5) duplicate publications. In other words, the current review is achieved through filtration of more than 180 academic papers. Of those, 65 that matched the criteria and protocols were originally elicited for review by eliminating duplicates and conducting a title and abstract screening. Utilizing the inclusion criteria specified, the final list of studies reviewed consisted of 28 total studies. This process ensures the selected studies will be representative of the most pertinent and rigorous studies from the previous five-year, inclusive of biometric fusion utilizing the multimodal approach.

The variation in the studies used in the research makes it impossible to formally conduct meta-analyses or aggregate all of the data to obtain a single statistical evaluation. Consequently, this report has provided an ongoing quantitative comparison to allow for the identification of patterns, strengths, and weaknesses across fusion techniques by means of systematic transparency.

When evaluating literature for inclusion, we emphasized papers that specifically apply deep learning or machine learning models to develop and/or improve existing (and/or new) fused biometric strategies, studies which use standard benchmark databases, such as CASIA, SDUM- LA-HMT and PolyU were prioritized over studies which use proprietary datasets. However, proprietary datasets were considered for inclusion if they proposed novel mechanisms and/or provided useful findings. Because of the wide variety of datasets (and evaluation conditions) used in these studies, we conducted a qualitative comparison of study performance using three common metrics (i.e., accuracy, Equal Error Rate (EER) and Area Under Curve (AUC)) measured on the study-specific test sets. Overall, this provides a concise but reliable comparative representation of generalized patterns of relative performance among the various fusion strategies tested.

To create a more systematically organized survey, labelled and classified according to their use of fusion level and attributes are the chosen articles including Serial Fusion, Decision Level Fusion, Feature Level Fusion and Hybrid or Multi-Modal. We summarize for each category the primary methodology used, model architectures, datasets utilized, accuracy scores, and limitations. By establishing this classification there is a comparison of the pros and cons of alternative fusion techniques from the perspectives of both scientific research and system design.

2.1 Feature-Level Fusion

Many researchers have used multi-level fusion of features to take advantage of the diverse replacement of biometric data. For example, Sb et al. [3] proposed the IMFIF-I-HDL scheme, which fuses facial, iris, and fingerprint feature data with the use of a Hierarchical Attention Network and Particle Swarm Optimization, reaching an overall accuracy of 98.975% when run on a Quantum-Enhanced Residual Network. Similarly, Soleymani et al. [4] introduced a quality-aware approach that integrates face, iris, and fingerprint features by accounting for sample quality, using weakly-supervised networks and novel loss functions, reporting significant improvements under challenging conditions.

In another study, Sharma et al. [5] developed a system that fuses gait and facial features using Deep Convolutional Neural Networks (DCNN), achieving 98.75% accuracy on the CASIA Gait Dataset B. Babu et al. [6] presented a hybrid model combining iris, face, and finger vein modalities, extracting complex features using adaptive CNNs and ResNets, with accuracy rates above 94% for each trait.

Patil [7] suggests a relatively original FRMSDNET algorithm for blending iris and fingerprint that has achieved the accuracy ratio of 99.4%. Ammour et al. [8] similarly merged fingerprint and ECG data applying a transformer-based structure, focusing on its persistence on spoofing. On the other hand, Byeol, Raina et al. [9] proposed a deep neural network-based model that blends pixel, score-level data, and features; a model that led to 2.2% retrieval accuracy enhancement. Similarly, Sayeed et al. [10] expanded the application of feature fusion through blending knuckle and palm vein traits with contrast improvement and the Chimp Optimization Algorithm that let to 99.85% accuracy.

Johnson and Chitra [11] suggested a forensic-based system according to the overlapped fingers, finger knuckles, and palm prints applying CS-RBFNN and BM-KMA for a better state of classification and feature integration. In another research implemented by Vensila and Wesley [12] adaptive particle swarm optimization was employed for fused face, finger vein data, and fingerprint achieving 97.14% of accuracy.

Another system, developed by Purohit and Ajmera [13], is being designed for continuous user authentication, during online exams combines keystroke patterns and face recognition. The researchers utilized a hybrid model named LCNN-Salp Swarm Optimization in parallel with a manipulated wolf optimization algorithm to achieve more efficiency through combining features. The study attempted to achieve higher security and to minimize cheating through e-learning, specifically during the Covid-19 pandemic. The system presented better accuracy and less false alarm compared to other strategies.

2.3 Score-Level Fusion

Score-level fusion has also gained attention due to its operational flexibility. Khatri and Sharma [3] reported a perfect 100% accuracy by combining iris, face, and palmprint modalities using CNNs. Medjahed et al. [14] developed a deep learning-based system integrating face and palm print scores using CNN and KNN, attaining 97.5% accuracy under noisy conditions.

Cherrat et al. [15] proposed a model using CNNs and Random Forest to fuse fingerprint, finger-vein, and face scores, achieving 99.49% accuracy. Using handwritten signature, iris and fingerprint traits, Yadav [16] demonstrated that he was able to obtain 99% accuracy in his findings. Alay and Al-Baity [17] also demonstrated 100% accuracy when conducting score level fusion of iris, face and finger vein features using convolution neural networks (CNN) using VGG-16.

Shinde and Kayte [18] fused face and fingerprint scores using the VGG16 and CNN models to achieve 99.65% accuracy using their methods; Kaur [19] developed a system to make score-level decisions to combine iris and fingerprint recognition based upon Support Vector Machines (SVM) and invariant local feature sets with an EER of less than 1.36%. Selvaraj et al. [20] combined five modalities (iris, fingerprint, palm print, ear, face) with their weighted score fusion to achieve 99.85% accuracy.

2.4 Decision-Level Fusion

A fusion model based on image texture-based feature data for both the iris and face using dual CNNs was proposed by Şerifi and Şerifi [21]. This model was able to achieve an impressive accuracy score of 99.53% from the CASIA-Iris-Distance image dataset. The localized information used in the implementation of the model allowed for increased recognition performance by mitigating the effects of pose and lighting changes.

2.5 Serial Fusion

Edwards & Hossain [22] used a Siamese NN to achieve a serial fusion approach for face, palm and fingerprint modalities in the context of deep learning and obtained a very high AUC of 0.9996, not only showing accuracy but also improved convenience for users through fewer number of biometrics being needed for user verification.

2.6 Multimodal Fusion (Non-Specific or Hybrid)

While different fusion-agreement methods have been identified from research undertaken by some authors, others have not followed a specific methodology, but instead have opted for a hybrid and/or less defined Fusion of Modalities Framework. For example, Salturk & Kahraman [23], when fusing dynamic signatures with facial recognition data using deep convolutional neural networks (CNNs), LSTMs, GRUs and Temporal Convolutional Networks (TCNs) were able to achieve accuracy levels in excess of 98% from small datasets.

Kaur & Kant [24] also developed a two phase, deep hashing model which combines three modalities (Face, Fingerprint, Iris), achieving an accuracy level of 96.15% and a precision level of 96.55%, and whose model was developed with consideration to anti-spoofing measures. Byeon et al. [9] when conducting research using a combination of many Fusion methods (Pixel, Feature and Score Fusion) on a data set of over 40,000 images, achieved an accuracy level of 99.6%.

Other Authors; Shim et al. [25], when creating ARGate - a late fusion architecture aimed at sensor reliability and the support of multi-purpose modes of multimodal data - contributed significantly to the field of fusion of features even though multimodal data alone is not considered biometric, however lessons can be drawn from ARGate in the development of biometric fusion architectures. Jha & Bansal [26] when conducting research using the combination of Face and Speech Recognition modalities, achieved an accuracy level of 97.51% for contactless authentication using the AVSpeech database, by employing a Bidirectional Long Short Term Memory (Bi-LSTM) network model.

El-Rahman & Alluhaidan [27] conducted an experiment to compare various fusion methodologies (Both Parallel and Sequential) with Fingerprint and ECG formulae, with AUC Scores of 0.96 and 0.99 respectively, demonstrating the benefits associated with biometrics through the use of the combination of physiological and behavioural modalities.

2.7 Section Summary:

In order for us to comprehend the frequency of fusion strategies used recently, we provide Figure 1 which illustrates the distribution of fusion strategies used within the 28 reviewed articles. In total, Feature Fusion (35.7%) was used most frequently among the different fusion strategies included across these articles, followed by Score Fusion (32.1%). These two strategies were effective in collecting large amounts of detailed biometric data for achieving very high levels of accuracy. Additionally, Hybrid or Multimodal Fusion strategies were present in 25% of the reviewed articles; therefore, there seems to be a developing trend of fusing multiple types of biometric system technologies to improve overall system performance.

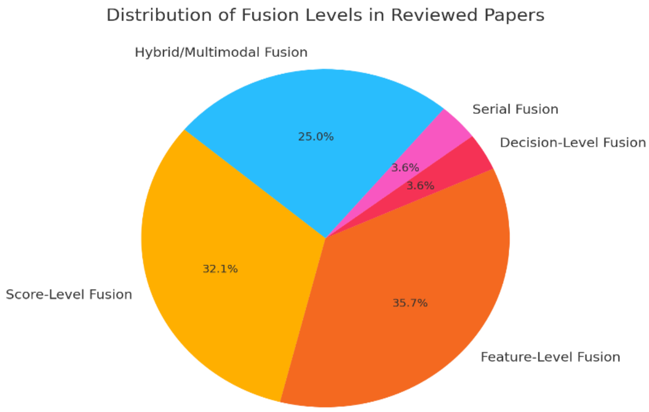

Figure 1. Distributed of fusion levels in reviewed papers.

In contrast, decision-level and serial fusion were each used in only 3.6% of studies, indicating limited application due to challenges in adaptability and performance. This distribution highlights current research focus areas and guides future exploration in fusion-based biometric systems.

3. Discussion

Through this research review it is being demonstrated that the fusion level has fundamental effects both on performance and reliability of multimodal biometric systems. Feature-level fusion enriches feature learning alongside with achieving higher ratio of accuracy simultaneously, even though such kind of fusions could be computationally sensitive and intensive in accidence to the fed data. It is worth mentioning that score-level fusion outstands as the most efficient and functional approach that would offer optimum accuracy, flexibility, and easy integration with numerous systems gaining 100% accuracy.

Despite the lack of wide application by users, Decision-level is highly resourceful to increase robustness, specifically within uncertain environments. Serial fusion has long demonstrated tangible reduction of user burden by its dynamicity and effectively eliciting the required traits, though it could be occasionally sensitive to model parameters. On the other hand, hybrid fusion approaches reflect high versatility and is capable to integrate the strong points of multiple fusion levels, yet their efficiency often is bound to specific combination of models and modalities. On the other hand, deep learning models, specifically attention-oriented systems and CNNs, have long played a focal role in success of the respective architectures. However, limitations related to dataset diversity, overfitting, and processing demands remain consistent across all fusion types.

This table compares the studies from above and is primarily designed to assist with an objective and systematic evaluation of the studies. Table 1 contains information on fusion levels, model type, biometric trait(s), dataset(s), accuracy of the study, and the limitations of each study. Presenting the study in this format gives a clear, easy to read, side-by-side comparison of studies and highlights trends, strengths, and deficiencies in the methods that were employed. Furthermore, this information is extremely helpful in providing insight into the effect of each fusion technique and model selection on system performance and can assist researchers and practitioners in identifying the best approaches to multimodal biometric recognition.

Table 1. Fusion Models Comparison.

Reference	Fusion Level	Model Used	Biometric Type	Dataset Used	Accuracy	Key Limitation
[1]	**Feature-level fusion**	**Hierarchical Attention Network, QERN**	**Face, Iris, Fingerprint**	**Not specified**	**98.975%**	**Complexity of implementation and ethical concerns**
[2]	**Feature and Score**	**Convolutional Neural Network**	**Iris, Face, Palmprint**	**PolyU-IITD, PolyU Cross-Spectral, Tufts Face**	**100%**	**Limited dataset variety, potential overfitting**
[3]	**Serial**	**Siamese Neural Network**	**Fingerprint, Palm, Face**	**XJTU Multimodal Database**	**0.9996**	**Relies on specific dataset; parameters settings may limit generalizability**
[4]	**Feature-level fusion**	**CNN**	**Face, Iris, Fingerprint**	**BIOMDATA, IJB-A, IIIT-Dehli**	**>30% improvement in accuracy**	**Dependency on the quality of input samples**
[5]	**Feature Level**	**DCNN**	**Face and Gait**	**ORL Face Database, CASIA Gait B, FEI Face Database**	**98.75%, 97.50%**	**Limited to specific datasets for validation**
[6]	**Image, Feature**	**YOLOv4-tiny, CNN**	**Face and Iris**	**CASIA-ORL, SDUMLA-HMT**	**100%**	**Limited to specific datasets**
[7]	**Decision-Level Fusion**	**Dual CNN + ULBP**	**Face & Iris**	**CASIA-Iris-Distance**	**99.53%**	**Limited dataset scope; possible overfitting**
[8]	**Feature Level**	**Adaptive CNN & ResNets**	**Iris, Face, Finger Vein**	**CASIA-Iris, Custom Face & Finger Vein datasets**	**Iris: 95.67%, Face: 94.78%, Finger Vein: 95.18%**	**Computational efficiency and dataset bias concerns.**
[9]	**Score Level**	**CNN and KNN**	**Face and Palm Print**	**FEI Face, IITD Palm Print**	**97.5%**	**Requires further exploration of real-world adaptability**
[10]	**Score Level**	**CNN, Random Forest**	**Fingerprint, Finger-vein, Face**	**SDUMLA-HMT Multimodal database**	**99.49%**	**Dependence on dataset quality and size**

[11]	**Feature and Score Level**	**CNN (VGG-32)**	**Iris, Fingerprint, Handwritten Signature**	**SDUMLA-HMT**	**99.51%**	**Limited generalizability due to dataset constraints**
[12]	**Feature/Score**	**CNN (VGG-16)**	**Iris, Face, Finger Vein**	**SDUMLA-HMT**	**99.39%/100%**	**Limited generalizability due to specific dataset**
[13]	**Feature & Score**	**AlexNet, VGG16, ResNet50**	**Finger Vein (FV), Finger Knuckle Print (FKP)**	**SDUMLA-HMT, PolyU FKP**	**99.89%**	**Dataset representativeness and computational complexity**
[14]	**Parallel, Sequential**	**CNN, Traditional Classifiers**	**ECG, Fingerprint**	**MIT-BIH, FVC2004**	**Parallel: 0.96 AUC, Sequential: 0.99 AUC**	**Limited generalizability due to dataset reliance**
[15]	**Score Level**	**VGG16 with CNN**	**Face and Fingerprint**	**KVKR dataset**	**99.65%**	**Limited dataset generalizability**
[16]	**Multimodal**	**CNN, LSTM, GRU, TCN**	**Dynamic Signatures, Facial Data**	**1,750 samples from 25 individuals**	**>98%**	**Limited sample size and hardware dependence**
[17]	**Feature-level**	**FRMSDNET**	**Fingerprint & Iris**	**CASIA-IrisV3, SOCOFing**	**99.4%**	**High computational complexity and overfitting**
[18]	**Feature-level fusion**	**Deep learning architecture with data-efficient transformers**	**Fingerprint and ECG**	**Common RS scene datasets**	**High accuracy**	**Not applicable to diverse datasets; potential overfitting.**
[19]	**Late Fusion**	**ARGate**	**Multi-modal sensor data**	**Human Activity Recognition, driver identification, KITTI**	**Up to 13.39% improvement**	**Limited to controlled datasets; real-time application not tested**
[20]	**Feature**	**Bi-LSTM**	**Face and Speech**	**DeepfakeTIMIT, AVSpeech**	**97.51%**	**Limited generalizability across environments**
[21]	**Feature-Level / Score-Level**	**CNN**	**Iris, Face**	**CASIA and ORL datasets**	**99.22% / 100%**	**Limited to specific datasets and conditions**
[22]	**Feature-level**	**Extreme Learning Machine + PSO**	**Face, Fingerprint, Finger Vein**	**Various biometric datasets**	**97.14%**	**Limited dataset diversity and real-time applicability challenges**
[23]	**Feature level**	**BM-KMA, CS-RBFNN**	**Fingerprints, Palm Prints,**	**NIST, CASIA, Finger Knuckle Dataset**	**High**	**Synthetic datasets may lack real-world representation**

			Finger Knuckles			
[24]	**Score-level**	**SVM**	**Fingerprint, Iris**	**IIITD-CLI, IIITD-Iris Spoofing, Clarkson, LivDet 2013, 2015**	**EER: <1.36%, ACER: <0.57%**	**Computational processes may be time-consuming**
[25]	**Pixel, Feature, Score**	**Deep Neural Networks**	**Face, Fingerprint, Iris**	**40,482 photos, 2712 individuals**	**99.6%**	**Reliance on simulated datasets**
[26]	**Feature-level**	**Deep Neural Network**	**Palm and knuckle vein**	**Not specified**	**99.85%**	**Computational complexity and real-time implementation challenges**
[27]	**Score-level**	**Convolutional Neural Network**	**Iris, Fingerprint, Palm Print, Ear, Face**	**CASIA-FingerprintV5, CASIA-FaceV5, CASIA-PalmprintV1, CASIA-Iris-V3, AMI Ear Dataset, MEPCO Multimodal Dataset**	**99.85%**	**Limited generalizability due to dataset constraints**
[28]	**Multimodal**	**CNN-based model**	**Face, Fingerprint, Iris**	**Four benchmark datasets**	**96.15%**	**Computational complexity**
[29]	**Feature-level**	**Hybrid LCNN-Salp swarm optimization**	**Keystroke, Face**	**Custom datasets**	**High**	**Limited dataset generalization**

When we compare the performance of biometric systems, accuracy has almost been the singular measured metric, making this the primary measure used for the analyses in this review. In order to thoroughly assess a biometric system's performance, other important metrics must be equally given consideration. Specifically, EER, FAR, FRR and AUC provide necessary insight into system security, usability, as well as the trade-offs present between different error types. Given that the bulk of the studies reviewed in this review focus on accuracy, and since most of the articles focused on the same measure of accuracy; this skewed the researcher's understanding about the performance of biometric systems and may have resulted in the potential to overlook more important performance characteristics of biometric systems. Studies should accept a broader range of measuring systems, as well as a wider range of notation standards to assist with more robust complex and detailed analysis and more appropriately inform real-world deployment choices.

When evaluating real-time authentication situations, the selected fusion technique should provide a balance between the accuracy of the result and latency or computing requirements. In many of the reviewed studies on score-level fusion, it appears to be the most effective method for this kind of application with respect to trade-offs (i.e., the method produces accurate results and can perform in many different computing environments because it requires moderate computing power) since score-level methods process previously extracted features rather than raw data. Likewise, decision-level aggregations provide a similar advantage for application in a true real-time environment. Since decision-level fusion occurs on the results of independent classifier outputs (the two classifiers are completely independent), they can be processed in parallel across different modalities and therefore produce a reduction in latencies. Conversely, because of the need to align and concatenate features prior to performing computations, and because feature-level data contains much richer information compared to all of the other aggregators, feature-level data fusion is very costly to compute and use for time-critical authentication scenarios.

Dynamic selection of modalities based on authentication confidence can help to reduce the number of modalities that require processing for each transaction through the serial fusion model. However, this model is reliant on the correct optimization of thresholds for optimal performance. Though hybrid models will provide the greatest power, they also incur high computational costs and are most appropriate for non-real-time or high-resource environments. For realistic use cases in the real-time environment, light models (e.g., MobileNet, EfficientNet) fused at the score or decision level currently provide the best opportunity to move forward.

Based on the detailed review and comparative analysis of recent multimodal biometric systems from 2020 to 2025, it is evident that the choice of fusion level significantly influences system performance, computational demands, and practical applicability.

- The use of score-level fusion has consistently delivered the highest performance levels (i.e., 100%) to date, making it the most popular approach due to its flexibility for practical applications.
- Feature-level fusion has some advantages when integrating information sources because it can provide multiple pieces of detailed information about the same object. However, feature-level fusion has some drawbacks as it requires accurate calibration of the data and incurs higher computational costs.
- Despite being less well-studied than score- or feature-level fusions, decision-level fusion provides enhancements in robustness in uncertain situations. Serial fusion may produce improvements in reducing user interaction needs; however, both of these techniques have challenges with their ability to adapt to different environments and to optimize performance.
- Hybrid fusion techniques combine the benefits of several fusion levels; however, they also tend to add complexity to system design and training.
- Each of the fusion techniques identified in this review faces a number of similar issues, including lack of sufficiently diverse datasets, high computational demands, and concerns regarding the privacy and ethical implementation of multiple fusion techniques.
- Many of the systems reviewed achieved very high accuracy rates (i.e., claims of 100% accuracy), but the potential for overfitting when a system is trained on a small subset of the available data limits the application of these accurate ratings. Additionally, computational complexity will hinder the ability to deploy these systems quickly and effectively on edge devices and in real-time.

4. Challenges And Limitations

Even though multimodal biometric systems have recently attracted researchers and appliers for their strong performance, yet they are being subject to several obstacles and drawbacks. For instance, the appliers of multimodal biometric systems might suffer shortages in real-world diverse datasets that would make their job hard. Feature-level fusion in spite of having good and accurate rate of final performance needs a very accurate data alignment. However, these biometric systems are highly computationally intensive, particularly deep learning models, which limits their potential for resource-constrained and real-time environments. Privacy issues become the most important concerns if multiple biometric characteristics are collected and stored, which change the way biometric data are protected. Further, determining the optimal fusion method generally consists of a trade-off between speed, accuracy, and robustness of the implementation.

The small number of benchmark datasets commonly used for multimodal biometric evaluations, such as CASIA, SDUMLA-HMT, and PolyU, limit the overall ability of multimodal biometric systems to generalize across different domains or applications. Even though all three benchmarks have played a vital role in providing standardised evaluations and comparisons, they are created and evaluated under laboratory conditions where lighting and other external influences are controlled, therefore not representative of real-world systems. Additionally, the continued reference to a small number of benchmark datasets creates significant risk of (1) reported performance values potentially overestimating the actual performance value of multimodal biometric systems deployed in real-world applications, and (2) training multimodal biometric models based on artefacts from the specific datasets, rather than learning the generalised biometric features to be applied in different applications/domains. Enhanced analysis and evaluation of the datasets used for developing multimodal biometric systems is warranted due to the restrictions and biases from a lack of diversity associated with the datasets collected, and as such, systematic evaluations must consider age, ethnicity, and other demographic factors when conducting evaluations. As a result, the high accuracy scores found in several studies should be taken with caution, and the area would be well served by an increased range of varied, intricate, and representative benchmark datasets that represent operational realities more accurately.

Although the tested systems show exceptional performance in a controlled environment, the practical deployment of multimodal biometric systems presents certain challenges. Latency and computational complexity are crucial, particularly on mobile devices or embedded systems, where hardware resources are limited. In addition, user friendliness and user experience aspects play a relevant role, in order to facilitate a seamless integration in real-world applications.

5. Future Directions

There have been significant advancements in multiple mode biometrics; however, a number of pressing open issues must be prioritized for future studies before they can be used in real world applications. Following our investigation, we identified the following three main challenges to focus on as necessary prior challenges to enable use of multimodal biometrics in practice: (1) address the generalization leakage caused by using only controlled data; (2) developing methods that can provide an acceptable trade-off between efficiency and accuracy for real-time use on resource-limited devices; and (3) create trust and transparency within multimodal biometrics via explainability, privacy preservation, and ethical design. The following sections will further discuss this list of priorities and additional considerations for further research directions.

- **First, lightweight and efficient model design** is essential for real-time authentication on portable and resource-constrained devices. Future work should focus on developing deep learning architectures that optimally balance accuracy with computational complexity for deployment on edge devices and mobile platforms.
- **Second, diverse and representative benchmarks** are urgently needed. The field's heavy reliance on a few controlled datasets (e.g., CASIA, SDUMLA-HMT) limits generalizability to real-world conditions. Public datasets that fairly represent different populations, environmental conditions, and sensor variations would significantly improve benchmarking and promote innovation.
- **The third requirement of architectural diversity** involves research into a new phase of investigation. Research on current architectures is very imbalanced, with most research still using convolutional neural networks (CNNs) and little research on alternative architectures (i.e., transformer networks, graph neural networks (GNNs)). Transformers appear to be more suitable for representing long-range dependencies and cross-modal relationships because of their inherent design, while graph neural networks have the ability to represent interactions between modalities with complex graphs. Future researches should evaluate CNN-transformer hybrid architectures and transformer-only architectures for use in fusion scenarios that involve sequential data, such as voice or gait, or for scenarios that have complicated cross-modal dependencies.
- **Fourth,** it is necessary that XAI techniques be embedded into Multimodal Biometric Systems. None of the reviewed research included interpretability of the model, yet they face greater demand for transparency via regulations (such as how GDPR provides for a "right to explanation"). Explainability is critical in building user trust, allowing debugging of a system, detecting bias, and providing accountability — especially when high-stakes applications are involved (e.g., border control and law enforcement). Future studies should utilize techniques like Attention Visualization, Saliency Maps, or SHAP values so that insights into decision-making processes and contributions from specific modalities can be disseminated.
- The fifth type of fusion techniques is the dynamic and adaptive fusions which provide a greater expectation of increased complexity than the traditional fixed approach of fusing based on doctrine with the introduction of dynamic fusion values based on contextual parameters. In support of this statement are three examples: (1) Fusion based on confidence knowledge where model uncertainty estimates gate the amount of contributions from the different modalities being fused; (2) Quality driven adaptive weighting where higher quality modalities are given

more weight when fusing their contribution at the score level; (3) Context switching frameworks wherein different fusion methods and strategies are used based on different environmental contexts, for example, switching from a face dominant fused output to a voice dominant fused output in low light conditions. Improvements in the mixture of experts along with improving the meta-learning aspects of development will create much more productive and effective implementations of the types of integrated approaches discussed here.

- **Sixth,** Systems will require system-level solutions to privacy issues. Given the multi-modal nature of the data involved, such as voice and face, future systems will need to utilize Federated Learning, Homomorphic Encryption, and Differential Privacy in order to build user trust and meet regulatory requirements.
- **Seventh,** the design and testing of biometric systems should take into account the need for Ethical Design (designing in compliance with ethical standards) and Bias Mitigation. Equality, transparency and fairness must be studied to ensure that all segments of the population are treated fairly by a biometric system and not discriminated against.

Developing and deploying robust, deployable, and socially responsible multimodal biometric systems (that can move from laboratory environments to practical real-world applications) will require the identification of solutions to these inter-related challenges: generalization, efficiency, trust, architecture innovation, adaptability, privacy, and ethics.

6. Conclusion

The findings of this paper indicate that multimodal systems, which integrate multiple types of biometrics, provide considerably better accuracy, robustness, and protection against spoofing than unimodal systems (i.e., one type of biometric). We found that most researchers use score-level fusion of biometric data as their preferred fusion method, while very few researchers have investigated serial and decision-level fusion methods because of concerns about either their performance or their adaptability. Although a great deal of research has been conducted and significant advances have been made in the application of deep learning to enhance the performance of multimodal biometric systems, there are still bottlenecks that impede developing and applying these systems to real-world situations. These include computational complexity of fusion, concerns about privacy law violations, and variation among the datasets used to train/fine-tune/develop the biometric systems. Furthermore, the literature reviewed in this research suggests there may be potential demographic or selection biases in the studies reviewed, specifically in the arena of face recognition that would provide generalization and fairness.

In order to deal with these bottlenecks, we focus on the vitality of factors such as computational efficiency, ethnic design, and inclusive appraisal protocols. Among all the available privacy-saving strategies, federated learning presents unprecedented solutions to secure sensitive data while enabling distributed model training. This research review offers a well-arranged synthesis of the most updated multimodal fusion strategies, shedding light on trends, critical gaps, and performance insights that could lead other researchers to develop both more deployable and reliable biometric structural architectures.

References

1. Purohit, H.; Dadhich, M.; Ajmera, P.K. Analytical Study on Users' Awareness and Acceptability towards Adoption of Multimodal Biometrics (MMB) Mechanism in Online Transactions: A Two-Stage SEM-ANN Approach. *Multimed Tools Appl* **2023**, *82*, 14239–14263, doi:10.1007/s11042-022-13786-z.
2. Deol, G.J.S.; Priyadarsini, P.I.; Nallagattla, V.R.G.; Amarendra, K.; Seelam, K.; Latha, B.R.A. A Multimodal Face and Fingerprint Authentication System Using Fuzzy Set Exponential Water Wave Optimization. *Journal of The Institution of Engineers (India): Series B* **2024**, doi:10.1007/s40031-024-01073-4.
3. Sb, J.; Nk, D.; Sm, L.; Sb, P.; Ka, H. Integrated Multimodal Face, Iris and Fingerprint Biometric Fusion for Advanced Person Identification Using Hybrid Deep Learning. **2024**, doi:10.5281/zenodo.12699235.
4. Soleymani, S.; Dabouei, A.; Taherkhani, F.; Iranmanesh, S.M.; Dawson, J.; Nasrabadi, N.M. Quality-Aware Multimodal Biometric Recognition; Quality-Aware Multimodal Biometric Recognition. *IEEE Trans Biom Behav Identity Sci* **2021**, *PP*, doi:10.48550/arXiv.2112.05827.
5. Sharma, A.; Jindal, N.; Thakur, A.; Rana, P.S.; Garg, B.; Mehta, R. Multimodal Biometric for Person Identification Using Deep Learning Approach. *Wirel Pers Commun* **2022**, *125*, 399–419, doi:10.1007/s11277-022-09556-7.

6. Babu, V.D.; Dornala, R.R.; Anusha, C.; Babu, P.R.; Mohan, K.K.; Sumanth, K.V. A Hybrid Multimodal Biometric Recognition System (HMBRS) Based on Fusion of Iris, Face, and Finger Vein Traits. In Proceedings of the Proceedings of the 5th International Conference on Smart Electronics and Communication, ICOSEC 2024; Institute of Electrical and Electronics Engineers Inc., 2024; pp. 1287–1292.
7. Patil, P.G. *International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING FRMSDNET Classifier for Multimodal Feature Fusion Biometric Authentication*; 2023; Vol. 2024;.
8. Ammour, N.; Bazi, Y.; Alajlan, N. Multimodal Approach for Enhancing Biometric Authentication. *J Imaging* **2023**, *9*, doi:10.3390/jimaging9090168.
9. Byeon, H.; Raina, V.; Sandhu, M.; Shabaz, M.; Keshta, I.; Soni, M.; Matrouk, K.; Singh, P.P.; Lakshmi, T.R.V. Artificial Intelligence-Enabled Deep Learning Model for Multimodal Biometric Fusion. *Multimed Tools Appl* **2024**, doi:10.1007/s11042-024-18509-0.
10. Sayeed, F.; Ahmed, K.R.; Swamy, S.M. Development of a Multimodal Biometric Recognition System with Feature Optimization and Deep Learning. *Multimed Tools Appl* **2025**, doi:10.1007/s11042-025-20709-1.
11. Johnson, J.; Chitra, R. Multimodal Biometric Identification Based on Overlapped Fingerprints, Palm Prints, and Finger Knuckles Using BM-KMA and CS-RBFNN Techniques in Forensic Applications. *Visual Computer* **2024**, *40*, 3217–3231, doi:10.1007/s00371-023-03023-5.
12. Vensila, C.; Boyed Wesley, A. Multimodal Biometrics Authentication Using Extreme Learning Machine with Feature Reduction by Adaptive Particle Swarm Optimization. *Visual Computer* **2024**, *40*, 1383–1394, doi:10.1007/s00371-023-02856-4.
13. Purohit, H.; Ajmera, P.K. Multi-Modal Biometric Fusion Based Continuous User Authentication for E-Proctoring Using Hybrid LCNN-Salp Swarm Optimization. *Cluster Comput* **2022**, *25*, 827–846, doi:10.1007/s10586-021-03450-w.
14. Medjahed, C.; Rahmoun, A.; Charrier, C.; Mezzoudj, F. A Deep Learning-Based Multimodal Biometric System Using Score Fusion. *IAES International Journal of Artificial Intelligence* **2022**, *11*, 65–80, doi:10.11591/ijai.v11.i1.pp65-80.
15. Cherrat, E.; Alaoui, R.; Bouzahir, H. Convolutional Neural Networks Approach for Multimodal Biometric Identification System Using the Fusion of Fingerprint, Finger-Vein and Face Images. *PeerJ Comput Sci* **2020**, *2020*, 1–15, doi:10.7717/peerj-cs.248.
16. Yadav, A.K. *Deep Learning Approach for Multimodal Biometric Recognition System Based on Fusion of Iris, Fingerprint and Hand Written Signature Traits*; 2021; Vol. 12;.
17. Alay, N.; Al-Baity, H.H. Deep Learning Approach for Multimodal Biometric Recognition System Based on Fusion of Iris, Face, and Finger Vein Traits. *Sensors (Switzerland)* **2020**, *20*, 1–17, doi:10.3390/s20195523.
18. Shinde, K.; Kayte, C. Multimodal Deep Learning Based Score Level Fusion Using Face and Fingerprint. In; 2023; pp. 140–152.
19. Kaur, B. Fingerprint and Iris Liveness Detection Using Invariant Feature-Set. *Multimed Tools Appl* **2024**, *83*, 60833–60859, doi:10.1007/s11042-023-17854-w.
20. Selvaraj, A.; Russel, N.S.; Seenivasan, M. Robust Penta-Modal Biometric Identification through Deep Learning and Weighted Score Fusion. *Iran Journal of Computer Science* **2025**, doi:10.1007/s42044-025-00234-y.
21. Şerifi, M.; Şerifi, Ü. Dual CNN and Texture-Based Face-Iris Multimodal Biometric System via Decision-Level Fusion. *Signal Image Video Process* **2025**, *19*, doi:10.1007/s11760-025-03873-7.
22. Edwards, T.; Hossain, M.S. Effectiveness of Deep Learning on Serial Fusion Based Biometric Systems. *IEEE Transactions on Artificial Intelligence* **2021**, *2*, 28–41, doi:10.1109/TAI.2021.3064003.
23. Salturk, S.; Kahraman, N. Deep Learning-Powered Multimodal Biometric Authentication: Integrating Dynamic Signatures and Facial Data for Enhanced Online Security. *Neural Comput Appl* **2024**, *36*, 11311–11322, doi:10.1007/s00521-024-09690-2.
24. Kaur, P.; Kant, C. 2-Phase Multi-Trait Biometric Authentication Model Against Spoofing Attack Using Deep Hash Model. *SN Comput Sci* **2025**, *6*, doi:10.1007/s42979-024-03513-w.
25. Shim, M.S.; Zhao, C.; Li, Y.; Zhang, X.; Zhang, W.; Li, P. Robust Deep Multi-Modal Sensor Fusion Using Fusion Weight Regularization and Target Learning. **2019**.
26. Jha, K.; Jain, A.; Srivastava, S. Feature-Level Fusion of Face and Speech Based Multimodal Biometric Attendance System with Liveness Detection. *AIP Adv* **2024**, *14*, doi:10.1063/5.0234430.
27. El-Rahman, S.A.; Alluhaidan, A.S. Enhanced Multimodal Biometric Recognition Systems Based on Deep Learning and Traditional Methods in Smart Environments. *PLoS One* **2024**, *19*, doi:10.1371/journal.pone.0291084.
28. Khatri, M.; Sharma, A. Deep Learning Approach Based on Iris, Face, and Palmprint Fusion for Multimodal Biometric Recognition System. *International Journal of Performability Engineering* **2023**, *19*, 407–416, doi:10.23940/ijpe.23.06.p6.407416.
29. Hattab, A.; Behloul, A. Face-Iris Multimodal Biometric Recognition System Based on Deep Learning. *Multimed Tools Appl* **2024**, *83*, 43349–43376, doi:10.1007/s11042-023-17337-y.

30. Daas, S.; Yahi, A.; Bakir, T.; Sedhane, M.; Boughazi, M.; Bourennane, E.B. Multimodal Biometric Recognition Systems Using Deep Learning Based on the Finger Vein and Finger Knuckle Print Fusion. *IET Image Process* **2020**, *14*, 3859–3868, doi:10.1049/iet-ipr.2020.0491.
31. Alay, N.; Al-Baity, H.H. A Multimodal Biometric System for Personal Verification Based on Different Level Fusion of Iris and Face Traits. *Biosci Biotechnol Res Commun* **2019**, *12*, 565–576, doi:10.21786/bbrc/12.3/3.