International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2350-2361
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0317

# Fusion Strategies And Open-Set Recognition In Multimodal Biometric Authentication Systems: A Comprehensive Review

*Madhav R Unnithan[1], Muhammed Sabir[2] , Sara Sherief[3] , Afiya S[4], Anish A Aziz[5]*
*[1,2,3,4] UG - Computer Science Engineering, College of Engineering Perumon, Kollam, Kerala*
*[5] Assistant Professor - Computer Science Engineering, College of Engineering Perumon, Kollam, Kerala*
***Email ID:*** *madhavrunnithan555@gmail.com[1], sabirsabu1235@gmail.com[2], sheriefsarah00@gmail.com[3] , afiyajahan642@gmail.com[4], anishaaziz@perumonec.ac.in[5]*

## Abstract
*Multimodal biometric authentication systems have gained significant attention due to their ability to enhance reliability, accuracy, and resistance to spoofing by integrating complementary biometric traits. At the same time, open-set recognition (OSR) has emerged as a critical paradigm for real-world security applications, where previously unseen or unauthorized identities may appear during deployment. Despite substantial progress in unimodal OSR and multimodal fusion strategies independently, their integration remains limited. This paper presents a review of fusion strategies and open- set recognition in multimodal biometric authentication systems. We examine unimodal open-set methods and multimodal fusion techniques across sensor, feature, score, and decision levels. A comparative analysis is performed based on architecture, threshold design, scalability, and open-set awareness. The review reveals a structural gap between open-set biometric theory and practical multimodal fusion systems, particularly in threshold calibration, enrollment scaling, and fusion-level risk modeling. Additionally, an adaptive preset-based fusion framework is outlined as a potential approach to address these challenges. Finally, key research directions are identified to guide the development of unified open-set multimodal authentication frameworks for secure real-world deployment.*
***Keywords:*** *Multimodal Biometrics; Open-Set Recognition; Fusion Strategies; Biometric Authentication, Threshold Calibration; Enrollment Scalability.*

## 1. Introduction
Biometric authentication systems utilize distinctive physiological and behavioral characteristics to verify or identify individuals. Traditional unimodal systems rely on a single biometric trait, such as face, fingerprint, or voice. While computationally efficient, unimodal systems are inherently limited by intra-class variability, environmental sensitivity, spoofing vulnerability, and non-universality. These limitations have motivated the development of multimodal biometric systems, which integrate multiple complementary modalities to improve recognition accuracy, robustness, and security [6], [8], [13]. Multimodal fusion can be performed at different stages of the biometric pipeline, including sensor, feature, score, and decision levels [6], [11]. Feature-level fusion captures inter- modal correlations, score-level fusion provides modularity and flexibility, and decision-level fusion enables simple logical integration. Recent advancements further introduce adaptive and dynamic fusion strategies to address varying environmental and signal quality conditions [3], [4]. While these approaches significantly enhance closed-set authentication performance, they often operate under the assumption that all testing identities are known in advance. In real-world security applications, however, biometric systems must operate under open-set conditions, where previously unseen individuals may attempt authentication. Open-set recognition (OSR) explicitly addresses this challenge

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2350-2361
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0317

by modeling unknown identity rejection through bounded decision regions and open space risk minimization [15]. Subsequent developments, including deep open-set models and open-world recognition frameworks, have strengthened theoretical and practical foundations for handling unknown identities [10], [16], [18]. In biometric-specific contexts, open-set face and speaker recognition studies further demonstrate the impact of enrollment scaling and threshold calibration on system performance [19], [20]. Despite these advances, open-set modeling and multimodal fusion have largely evolved as parallel research directions. Most multimodal systems implicitly rely on similarity thresholding for unknown rejection rather than explicitly modeling open-set risk. Conversely, open-set biometric research predominantly focuses on unimodal settings. The interaction between fusion strategies and open-set behavior—particularly in terms of threshold calibration, enrollment scalability, and risk propagation across modalities—remains insufficiently explored. This paper presents a comprehensive review of fusion strategies and open-set recognition mechanisms in multimodal biometric authentication systems. We systematically analyze unimodal OSR foundations, multimodal fusion architectures, and their intersection through a structured comparative analysis. The goal is to identify existing limitations and outline future research directions toward robust, scalable, and secure multimodal open-set authentication frameworks.

## 2. Background And Fundamental Concepts

### 2.1. Unimodal vs. Multimodal Biometric Systems

Biometric authentication systems utilize distinctive physiological and behavioral characteristics to verify or identify individuals. Physiological traits include facial features, fingerprints, and iris patterns, while behavioral traits encompass modalities such as voice, gait, and signature dynamics. These traits are typically evaluated based on properties such as universality, uniqueness, permanence, and collectability. Traditional systems are predominantly unimodal, relying on a single biometric trait for authentication. Although unimodal systems are computationally efficient and simpler to implement, they are inherently limited by sensitivity to noise, intra-class variability, susceptibility to spoofing, and non- universality. Performance may also degrade under adverse environmental conditions or partial occlusion. To overcome these limitations, multimodal biometric systems integrate two or more independent modalities to improve recognition accuracy and reliability. By exploiting complementary information, multimodal approaches reduce ambiguity, enhance resistance to spoofing, and provide greater stability across varying operational environments. This transition from unimodal to multimodal frameworks forms the conceptual basis for the fusion strategies discussed in subsequent sections.

### 2.2. Authentication Paradigms

Biometric authentication systems primarily operate under two paradigms: verification and identification. Verification is a one-to-one (1:1) process in which a user claims an identity and the system confirms the match, whereas identification is a one-to-many (1:N) process where the system determines an individual's identity by comparing the input sample against all enrolled templates. Biometric systems may further operate under closed-set or open-set recognition scenarios. Closed-set recognition assumes that the subject's identity exists within the enrolled database, requiring assignment to one of the known classes. In contrast, open-set recognition allows for the possibility of unknown individuals and requires the system to reject unauthorized or unenrolled users. This setting better reflects real-world security applications and introduces additional challenges in decision thresholding and unknown identity handling. In multimodal systems, execution strategies may be sequential or parallel. Sequential execution evaluates modalities in a predefined order, potentially reducing computational cost by invoking additional modalities only when necessary. Parallel execution processes multiple modalities simultaneously and fuses their outputs to enhance robustness. Additionally, continuous authentication has emerged as an extension of traditional point-in-

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2350-2361
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0317

time verification, enabling persistent monitoring throughout an active session. These paradigms define the operational context in which multimodal and open-set biometric systems are designed and evaluated.

### 2.3. Performance Metrics

The performance of biometric authentication systems is typically evaluated using statistical error metrics derived from decision thresholds. Two fundamental measures are the False Acceptance Rate (FAR) and the False Rejection Rate (FRR). FAR represents the probability of incorrectly accepting an unauthorized individual, whereas FRR denotes the likelihood of rejecting a legitimate user. A widely used aggregate metric is the Equal Error Rate (EER), which corresponds to the operating point at which FAR and FRR are equal. Lower EER values indicate better overall system performance. Receiver Operating Characteristic (ROC) curves are commonly employed to visualize the trade-off between true acceptance rate and false acceptance rate across varying thresholds. Detection Error Tradeoff (DET) curves provide a complementary representation using logarithmic scaling, offering clearer visualization in low-error regimes. In open-set recognition scenarios, additional evaluation considerations arise, including the ability of the system to correctly reject unknown identities. Metrics such as False Negative Identification Rate (FNIR) at a given False Positive Identification Rate (FPIR), along with False Alarm Rate, are commonly used to evaluate identification performance under such conditions. Therefore, threshold tuning plays a crucial role in balancing security and usability. These performance metrics form the basis for comparative analysis of fusion strategies and open-set mechanisms discussed in later sections.

### 2.4. Fusion Levels in Multimodal Biometric Systems

Fusion in multimodal biometric systems refers to the integration of information from multiple modalities to produce a unified authentication decision. Fusion can be performed at different stages of the biometric pipeline, each with distinct advantages and challenges. Sensor-level fusion combines raw data acquired from multiple sensors before feature extraction. While this approach preserves maximum information, it requires compatible data formats and synchronized acquisition mechanisms. Feature-level fusion integrates feature vectors extracted from different modalities into a single composite representation. This approach captures rich inter-modal correlations but may suffer from high dimensionality and feature incompatibility. Score-level fusion combines matching scores generated independently by individual classifiers. Common techniques include weighted summation, normalization-based combination, and statistical modeling. Score-level fusion offers flexibility and modularity, making it one of the most widely adopted strategies in practical systems. Decision-level fusion operates at the final stage by combining binary decisions from individual modalities using logical rules such as AND, OR, or majority voting. Although computationally simple, this approach may discard detailed discriminative information available at earlier stages. The choice of fusion level significantly influences system accuracy, computational complexity, scalability, and adaptability. A comprehensive analysis of fusion strategies and their effectiveness is presented in Section 3.

## 3. Fusion Strategies In Multimodal Biometric Systems

Fusion is a key mechanism that drives improvements in multimodal biometric authentication systems. By combining multiple sources of biometric evidence, fusion strategies enhance accuracy, robustness, and resistance to spoofing attacks [1], [2]. The performance of a multimodal system is largely determined by the stage at which information is combined and the specific fusion rules applied. Sensor-Level Fusion Sensor-level fusion, also known as data-level fusion, in- volves combining raw biometric signals from multiple sensors before any feature extraction occurs. Integrating information at this early stage preserves the maximum amount of signal content and retains potentially discriminative details that may be lost in later processing stages [8], [12]. Despite its advantages, sensor-level fusion can be difficult to implement in practice because it

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2350-2361
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0317

requires compatible hardware and precise temporal synchronization among heterogeneous sensors [6], [12]. Differences in data formats and sampling rates—for instance, between high-resolution imaging devices and audio sensors—can increase system complexity and computational demands, especially in real-time authentication scenarios [4], [8].

### 3.1.Feature-Level Fusion

Feature-level fusion combines feature vectors extracted from individual modalities into a single representation, often through concatenation or transformation into a joint feature space [3], [14]. For example, facial descriptors based on Eigenfaces and voice features such as Cepstral coefficients are extracted independently and then merged into a unified representation [14]. By capturing rich inter-modal correlations at this stage, feature-level fusion can improve recognition accuracy and system robustness [4], [14]. However, this approach can lead to the "curse of dimensionality," where the resulting feature vector becomes very large, increasing computational requirements and potentially reducing classifier performance [3], [6]. High-dimensional feature spaces may also demand extensive training data to avoid overfitting. To address these challenges, techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are commonly used to produce compact and discriminative fused representations [14].

### 3.2. Score-Level Fusion

Score-level fusion is one of the most widely used strategies in practical multimodal biometric systems because it strikes a balance between performance and modularity [6], [11]. In this approach, each modality generates a matching score independently, which is then normalized and combined using a predefined fusion rule [11], [12]. Typical fusion methods include simple and weighted sum rules, product rules, and more advanced machine learning- based combinations, such as using Support Vector Machines (SVM) or Artificial Neural Networks (ANN) [11], [14]. Com- pared to feature-level fusion, score-level fusion allows individual modalities to be optimized separately and avoids problems associated with incompatible raw data or high-dimensional concatenated features [6], [12]. Score normalization is a critical component in score- level fusion. Since different modalities produce scores that may vary in range or distribution—for example, face versus fingerprint—normalization ensures that no single modality dominates the final decision [6], [17]. Proper normalization improves the stability and reliability of the fused system.

### 3.3.Decision-Level Fusion

Decision-level fusion integrates the final binary outputs (accept or reject) from individual biometric matchers into a single decision [12]. Each modality independently produces a discrete outcome, which is then combined using logical rules such as AND (all modalities must accept) or OR (accept if at least one modality approves) [12]. Although simple and easy to implement, decision-level fusion is generally less informative than feature- or score-level fusion, since it discards the confidence information contained in matching scores [6], [12]. Consequently, its overall performance may be lower than fusion strategies that leverage richer intermediate data.

### 3.4.Static vs. Adaptive Fusion Strategies

An important development in multimodal biometric research is the shift from static to adaptive fusion mechanisms [3], [4]. In static fusion, fixed weights or predetermined rules are applied to each modality regardless of input quality or environmental conditions [3]. While straightforward, this approach may not perform well in dynamic real-world situations where modality reliability varies.

Adaptive fusion, on the other hand, adjusts the contribution of each modality dynamically based on contextual factors such as signal quality, environmental conditions, or confidence measures [3]. For example, if facial images are partially occluded or poorly illuminated, an adaptive system may reduce the weight of the facial modality while relying more heavily on a more reliable modality such as voice. This adaptability enhances robustness and reliability in unconstrained environments [4].

Overall, no single fusion strategy is universally optimal. Early-stage fusion may exploit richer information but often comes with higher computational and implementation costs, while later-

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2350-2361
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0317

stage methods offer greater modularity and flexibility. The rise of adaptive fusion highlights the need for systems that can respond to variability in data quality and operational conditions. Fusion strategies should therefore be designed to align with application-specific requirements, including scalability, environmental constraints, and security needs. However, in real-world scenarios, systems must also handle unknown or unseen identities, highlighting the importance of open-set recognition.

## 4. OPEN-SET RECOGNITION IN BIOMETRIC AUTHENTICATION SYSTEMS

Open-set recognition (OSR) addresses a critical limitation of conventional closed-set classifiers, which assume that all testing identities are known during training. In practical biometric authentication systems, previously unseen individuals may appear during deployment, requiring the system not only to correctly recognize enrolled users but also to explicitly reject unknown identities. The distinction between Known Known Classes (KKCs) and Unknown Unknown Classes (UUCs) is formally articulated in [10], which establishes the conceptual foundation for open-set learning.

### 4.1.Theoretical Foundations of Open-Set Recognition

Scheirer et al. describe open-set recognition as the need to correctly identify known inputs while also avoiding incorrect acceptance of unknown inputs [15]. This challenge is captured by the term open space risk, which refers to the risk of assigning labels to regions where no known training data exists.

Traditional classifiers such as standard SVMs often form unbounded decision regions, which can lead to overgeneralization. As a result, unknown inputs may be incorrectly classified as known classes. To address this, OSR aims to balance classification error and open space risk, expressed as:

$$R(f) + \lambda D(f) \qquad (1)$$

where R(f) represents open space risk and D(f) represents classification error [15]. This highlights the need for controlled decision boundaries that reduce incorrect acceptance of unknown samples. The concept of openness, introduced in [15], indicates how many unseen classes appear during testing compared to training. As openness increases, the recognition task becomes more difficult, which is important in real-world biometric systems.

### 4.2. Discriminative Open-Set Approaches

Early open-set methods focused on restricting decision regions so that unknown inputs are not wrongly classified as known classes. The 1-vs-Set Machine [15] does this by using two boundaries instead of one, creating a limited region that reduces overgeneralization. Later methods used threshold-based rejection along with Extreme Value Theory (EVT). As discussed in [10], approaches like Weibull SVM and the Extreme Value Machine (EVM) use score values to estimate whether an input belongs to a known class, improving unknown rejection. However, choosing proper thresholds and tuning these models is still mostly based on trial and error [10]. This makes reliable decision-making difficult in real-world systems.

### 4.3. Deep Neural Network-Based Open-Set Methods

Deep neural networks trained with Softmax typically assume closed-set conditions and can assign high confidence even to unknown inputs. To address this, OpenMax [18] extends Softmax by introducing an explicit “unknown” class. It adjusts the output probabilities based on how different an input is from known training data, helping reduce incorrect acceptance of unknown samples and better control open space risk [15]. As discussed in [10], similar deep open-set approaches modify output layers, loss functions, or feature representations to improve unknown detection. However, despite these improvements, most methods still depend on empirically chosen thresholds for final decision-making, which limits their reliability in practical systems.

### 4.4.Open-World Recognition

While OSR focuses on rejecting unknown samples, OpenWorld Recognition (OWR) extends this idea by incorporating incremental learning and scalability. Bendale and Boult define OWR as a system that can

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2350-2361
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0317

recognize known classes, detect unknown inputs, and continuously update itself by adding new identities [16]. Their Nearest Non-Outlier (NNO) algorithm replaces Softmax probabilities with distance-based functions, ensuring that inputs far from known classes are rejected [16]. It also allows new classes to be added without retraining the entire model, making it suitable for evolving biometric systems. Although OWR introduces useful capabilities such as incremental learning, most existing multimodal authentication systems do not explicitly adopt such mechanisms. However, these ideas present an interesting direction for future extensions, particularly in scenarios involving dynamic enrollment and large-scale deployment.

### 4.5. Open-Set Recognition in Biometric Modalities

Open-Set Face Recognition: Gunther ¨ et al. studied open-set face identification and distinguished between enrolled identities, known unknowns, and unknown unknowns [19]. They showed that simple cosine similarity thresholding is not reliable for open-set identification, mainly because score values vary across different identities. The Extreme Value Machine (EVM) performs better than cosine similarity and LDA by estimating whether a sample belongs to a known class using EVT [19]. Evaluation is also extended beyond closed-set CMC curves to include Detection and Identification Rate (DIR) and False Alarm Rate (FAR), with a focus on low FAR performance [19]. Open-Set Speaker Identification: Peri et al. introduced VoxWatch, a large-scale benchmark for Open-Set Speaker Identification (OSI) [20]. Their study shows that as the number of enrolled users increases, non-target scores shift upward, leading to higher False Alarm Rates. Even strong deep speaker models show performance drop as enrollment size grows [20]. They also show that score calibration improves performance, and score fusion gives additional gains, although scaling issues still remain. These results highlight the difficulty of maintaining reliable performance in large-scale open-set

### 4.6.Evaluation Protocols and Implications for Multimodal Systems

Accuracy alone is inadequate for evaluating open-set systems. Geng et al. emphasize the need to separately measure known-class accuracy and unknown rejection performance [10], while Scheirer et al. advocate the use of F-measure to avoid misleading performance estimates [15]. In biometric contexts, metrics such as FAR, FRR, EER, and DIR are essential for characterizing operational trade-offs [19], [20]. Although unimodal open-set recognition has been extensively studied [10], multimodal OSR remains underexplored. Each modality produces different score distributions and be- haves differently as the number of users increases, as observed in open-set face [19] and speaker recognition [20]. Consequently, multimodal systems must address open space risk not only at the modality level but also at the fusion level, particularly under large enrollment conditions.

## 5. comparative analysis of existing approaches

This section presents a comparative analysis of the reviewed unimodal open-set recognition (OSR) methods and multi- modal biometric fusion systems. Although these approaches differ in scope, the comparison highlights their complementary strengths and limitations. The analysis is structured around architectural design, fusion strategy, open-set awareness, thresh- old modeling, scalability, evaluation metrics, and system-level limitations.

### 5.1.System-Level Comparison

Table 1 summarizes the key characteristics of the surveyed approaches. Threshold strategies and limitations are summarized based on reported methodologies and observed system behavior.

### 5.2. Analysis of Open-Set Awareness

Unimodal open-set recognition (OSR) methods explicitly focus on rejecting unknown identities [15], [16], [18]– [20], [22]. For example, Su et al. [22] use FNIR@FPIR and percentile-based thresholding to handle unknown inputs. Gunther et al. [19] apply Extreme Value Machine (EVM) probabilities to control open space risk, while Peri et al. [20] show that increasing the number of enrolled users causes score distributions to shift, leading to higher false alarms. Theoretical foundations for OSR are established in [15], [16], [18], with a broader

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2350-2361
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0317

overview provided in [10]. In contrast, most multimodal biometric systems are designed under closed-set assumptions [1]–[3], [5], [11], [14]. Unknown identities are typically handled using simple similarity thresh- olds rather than explicit open-set modeling. Only Irfan et al. [21] explore open-set identification in a multimodal setting using a Bayesian approach, but their evaluation is limited to a small-scale dataset.

### 5.3.Fusion Strategy Comparison

Feature-level fusion [3], [11], [14] combines features from different modalities into a single high-dimensional representation. While this can improve accuracy in closed-set settings, it does not directly handle unknown inputs or control open space risk. Score-level fusion [1], [11] combines similarity scores from each modality, usually using simple methods like weighted sum. However, if one modality produces a high incorrect score, it can increase the final fused score and lead to false acceptance. Score normalization techniques [17] try to reduce this effect. Decision-level and serial fusion approaches [2], [5] combine final decisions from each modality using logical rules or step-by-step verification. Although these methods can improve security, they rely on manually chosen thresholds and do not provide strong support for open-set conditions. General fusion strategies are discussed in [6], [8], [13].

### 5.4.Threshold Modeling and Scalability

A key difference between unimodal OSR methods and multimodal systems lies in how thresholds are designed. OSR approaches typically use percentile-based or probabilistic thresholds derived from non-mated score distributions [15], [16], [18]–[20], [22], allowing better control over unknown rejection. Periet al. [20] also show that fixed thresholds become less reliable as the number of enrolled users increases, highlighting the need for proper calibration. In contrast, most multimodal systems use fixed or classifier- dependent thresholds without considering how score distributions change as more users are added [1]–[3], [5], [11], [14]. As a result, they may not perform reliably under large-scale or open-set conditions, and none of the reviewed multimodal works explicitly analyze open-set scaling Behavior.

### 5.5. Synthesis

The comparative analysis reveals three major observations:

- Open-set recognition methods are theoretically grounded but predominantly unimodal [15], [16], [18]–[20], [22].
- Multimodal fusion strategies improve closed-set accuracy but rarely incorporate open-set principles [1], [3], [5], [11], [14].
- Enrollment scaling and threshold calibration remain largely unexplored in multimodal biometric systems.

These findings highlight a structural gap between open-set biometric theory and practical multimodal fusion architectures, motivating further research in unified open-set multimodal authentication frameworks.

**Table 1** COMPARATIVE ANALYSIS OF REVIEWED BIOMETRIC AND FUSION APPROACH

| Reference | Modalities | Fusion Level | Open-Set Explicit | Threshold Strategy | Evaluation Metrics | Key Limitation |
|---|---|---|---|---|---|---|
| Su et al. [22] | Face, Gait, Re-ID | Unimodal | Yes | FPIR-percentile based | FNIR@FPIR, Rank-based | Not multimodal |
| Gunther et al. [19] | Face | Unimodal | Yes | EVM Proba-bility thresh-old | DIR, FAR | No fusion analysis |
| Peri et al. [20] | Speaker | Unimodal | Yes | Fixed + calibrated thresh-olds | FAR, Detection Error | Watchlist scaling degradation |

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2350-2361
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0317

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Irfan et al. [21] | | | Face + Soft Biometrics | Bayesian Fusion | Yes | Bayesian posterior thresh-old | DIR, Recognition Rate | Small-scale dataset |
| Jain et al. [5] | | | Face + Fingerprint + Speech | Decision-Level | No | Score thresh- old | Accuracy | Closed-set assumption |
| Alsaedi et al. [3] | | | Face + Voice | Feature-Level (Dynamic) | No | Implicit classifier threshold | Accuracy, FAR, FRR | No open-set evaluation |
| Choi & Park [1] | | | Face + Gesture | Score-Level | No | Similarity threshold | Accuracy | No scalability analysis |
| Andrian et al. [2] | | | Face + Voice | Serial Decision-Level | No | Dual threshold voice | Accuracy | Closed-set design |
| Byahatti & Shet- tar [11] | | | Face + Voice | Feature, Score, Rank, Decision | No | Euclidean distance threshold | FAR, FRR, ROC | No unknown modeling |
| Abozaid [14] | et | al. | Face + Voice | Feature + Score Fusion | No | Classifier based threshold | Accuracy | No open-set robustness |

## 6. Research Gaps And Future Directions

The comparative analysis highlights several structural gaps between open-set recognition theory and multimodal biometric fusion architectures. While significant progress has been made in unimodal OSR and multimodal fusion independently, their integration remains limited, particularly in handling unknown identities and fusion-level decision making. This section out- lines key research directions derived from the surveyed literature.

### 6.1.Integration of Open-Set Risk Modeling into Multimodal Fusion

(OSR) methods explicitly handle un- known inputs by controlling open space risk through threshold calibration and statistical modeling [15], [16], [18], [19], [22]. In contrast, most multimodal biometric systems focus mainly on improving closed-set accuracy using feature, score, or decision-level fusion [1], [3], [5], [11], [14], without explicitly addressing unknown identity rejection. A key research gap is the lack of open-set awareness within the fusion process itself. Instead of applying simple thresholds after fusion, future systems should incorporate rejection mechanisms directly into the fusion stage. This can involve modeling how non-target scores behave after fusion, or designing fusion strategies that limit incorrect acceptance of unknown inputs and maintain controlled decision boundaries.

### 6.2.Enrollment Scaling and Watchlist Robustness in Multi- modal Systems

Open-set speaker identification studies show that increasing the number of enrolled users shifts non-target score distributions and raises false alarm rates [20]. However, similar analysis is largely missing in multimodal biometric systems. Future work can evaluate how multimodal systems behave as enrollment size increases, particularly how fused scores change under such conditions. Studying whether certain fusion strategies remain stable at larger scales, along with simple threshold adjustments based on enrollment size, may help improve overall reliability.

### 6.3.Threshold Calibration in Fused Score Spaces

Choosing the right threshold is a key challenge in open- set biometric systems. Unimodal OSR methods often use structured approaches such as FPIR-controlled or percentile- based thresholds [22], which are designed to better handle unknown inputs. In contrast, most multimodal systems rely on fixed or empirically chosen thresholds [1]–[3], [11], which may not remain reliable across different conditions. When multiple modalities are combined, the final decision is made on a fused score rather than individual modality scores. This makes threshold selection more complex, as the fused score can vary depending on modality behavior and reliability. Therefore, there is a need for more systematic approaches to threshold design in multimodal

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2350-2361
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0317

systems. Future work can explore methods that adapt thresholds based on confidence levels, score distributions, and the relative reliability of each modality. Such approaches can help improve stability and reduce incorrect acceptance of unknown inputs.

### 6.4. Lack of Unified Multimodal Open-Set Benchmarks

While unimodal open-set benchmarks are well established [19], [20], [22], similar large-scale benchmarks for multimodal systems are limited. Existing studies are often small-scale and scenario-specific [21].

Future work could focus on developing standardized multi- modal open-set benchmarks with consistent evaluation metrics such as FNIR@FPIR. This would enable fair comparison across systems and improve reliability under real-world conditions.

### 6.5.Toward Unified Open-Set Multimodal Authentication Frameworks

The surveyed literature indicates that open-set modeling and multimodal fusion have evolved largely in parallel. Bridging this divide requires architectures that treat unknown identity rejection as a first-class design objective rather than an implicit consequence of low similarity scores. Future multimodal authentication systems may benefit from modular architectures in which individual modalities maintain calibrated open-set decision functions, followed by fusion mechanisms that preserve bounded risk properties at the system level. Investigating such unified frameworks represents a promising direction for advancing robust real-world biometric authentication.

### 6.6.Proposed Adaptive Preset-Based Open-Set Fusion Framework

### 6.7.Motivation

Existing multimodal biometric systems typically rely on fixed fusion strategies and static thresholds, assuming uniform modality reliability and user behavior. However, real-world authentication scenarios are inherently dynamic, with variations in environmental conditions, sensor quality, and user-specific constraints. Moreover, open-set recognition is often handled using simple thresholding mechanisms without explicitly modeling inter-modal interactions or contradictions. This motivates the need for a flexible, context-aware fusion framework that integrates open-set principles directly into the decision process.

### 6.8. Framework Overview

We propose an adaptive preset-based multimodal fusion framework that incorporates context awareness and explicit open-set decision making. The core idea is to condition the fusion process based on predefined presets associated with user characteristics or operational conditions, enabling dynamic adjustment without modifying the underlying system architecture.

The framework consists of the following components:

- Primary Modality-Based Context Identification: A reliable modality (e.g., face) is used to establish an initial identity hypothesis. Based on this, a corresponding preset configuration is retrieved.
- Preset-Driven Fusion Adaptation: Each preset defines modality weights, decision thresholds, and sensitivity parameters. This allows the system to dynamically adjust the contribution of each modality depending on context.
- Contradiction-Aware Decision Mechanism: The framework incorporates a mechanism to detect inconsistencies between modalities. Modalities that strongly contradict the primary hypothesis are penalized or as- signed reduced influence during fusion.
- Open-Set Aware Final Decision: The fused score is evaluated against a calibrated threshold to determine acceptance or rejection. This enables explicit rejection of unknown identities ensuring robustness under open-set conditions.

### 6.9.Key Features

The proposed framework introduces several important characteristics:

- Preset-Based Adaptation: Allows the system to dynamically adjust fusion

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2350-2361
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0317

parameters such as modality weights and thresholds without modifying the underlying architecture.

- Context-Aware Fusion: Adapts modality contributions based on reliability, environmental conditions, and user-specific constraints.
- Contradiction Handling: Enhances robustness by identifying and penalizing modality outputs that are inconsistent with the overall identity hypothesis.
- Integrated Open-Set Behavior: Performs unknown identity rejection as part of the fusion process, rather than relying on separate post-processing steps.

### 6.10. Advantages

Compared to traditional multimodal biometric systems, the proposed framework:

- Improves robustness in dynamic environments through preset-based adaptation and context-aware fusion.
- Effectively handles modality degradation or absence by dynamically adjusting modality contributions.
- Reduces false acceptance by incorporating contradiction- aware fusion and controlled decision boundaries.
- Integrates open-set rejection directly within the fusion process, improving reliability in real-world scenarios.
- Provides flexibility for deployment across diverse user conditions without requiring changes to the core architecture.

### 6.11. Summary

The proposed adaptive preset-based fusion framework pro- vides a practical direction for integrating open-set recognition with multimodal biometric systems. By combining context- aware adaptation, contradiction-aware decision making, and fusion-level open-set handling, it addresses key limitations of traditional static fusion approaches. This framework highlights a step toward more reliable and flexible real-world authentication systems.

## Conclusion

This paper reviewed fusion strategies and open-set recognition mechanisms in multimodal biometric authentication systems. While multimodal fusion techniques significantly improve closed-set authentication performance by integrating complementary modalities, open-set recognition frameworks provide essential theoretical tools for handling unknown identities in real-world deployments. The comparative analysis demonstrates that open-set modeling is theoretically mature but predominantly unimodal, whereas multimodal systems largely assume closed-set conditions and rely on heuristic thresholding. Enrollment scalability, fused score calibration, and fusion-level risk control remain underexplored challenges. To address these limitations, this paper outlined an adaptive preset-based fusion framework that integrates context-aware adaptation, contradiction-aware decision making, and fusion- level open-set handling. Bridging the gap between multimodal fusion architectures and open-set recognition principles is essential for developing secure biometric systems that can operate reliably in dynamic, real-world environments. Future research should focus on unified frameworks that incorporate calibrated open-set decision mechanisms at both modality and fusion levels, enabling scalable and robust multimodal authentication.

## References

[1]. H. Choi and H. Park, “A multimodal user authentication system using faces and gestures,” BioMed Research International, vol. 2015, Art. no. 343475, 2015, doi: 10.1155/2015/343475.

[2]. R. Andrian and G. P. Kusuma, “Serial multimodal biometrics authentication and liveness detection using speech recognition with normalized longest word subsequence method,” International Journal on Informatics Visualization, vol. 8, no. 3, pp. 1260–1270, Sep. 2024, doi: 10.62527/joiv.8.3.2247.

[3]. N. H. Alsaedi and E. S. Jaha, “Dynamic audio-visual biometric fusion for person recognition,” Computers, Materials & Continua, vol. 71, no. 1, pp. 1283–1311, 2022, doi: 10.32604/cmc.2022.021608.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2350-2361
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0317

[4]. Z. Qin, P. Zhao, T. Zhuang, F. Deng, Y. Ding, and D. Chen, "A survey of identity recognition via data fusion and feature learning," Information Fusion, vol. 91, pp. 694–712, 2023, doi: 10.1016/j.inffus.2022.10.032.

[5]. K. Jain, L. Hong, and Y. Kulkarni, "A multimodal biometric system using fingerprint, face, and speech," Dept. Comput. Sci. Eng., Michigan State Univ., East Lansing, MI, USA, Tech. Rep. MSU-CPS-98-32, 1998.

[6]. Ross, A. K. Jain, and J.-Z. Qian, "Information fusion in biometrics," in Audio- and Video-Based Biometric Person Authentication, Lecture Notes in Computer Science, vol. 2091, 2001, pp. 354–359, doi: 10.1007/3-540-45344-X 52

[7]. S. Albalawi, L. Alshahrani, N. Albalawi, R. Kilabi, and A. Alhakamy, "A comprehensive overview on biometric authentication systems using artificial intelligence techniques," International Journal of Advanced Computer Science and Applications, vol. 13, no. 4, 2022, doi: 10.14569/IJACSA.2022.0130491.

[8]. S. Raju and V. Udayashankara, "A survey on unimodal, multimodal biometrics and its fusion techniques," International Journal of Engineering and Technology (UAE), vol. 7, pp. 689–695, 2018, doi: 10.14419/ijet.v7i4.36.24224.

[9]. R. Brown, G. Bendiab, S. Shiaeles, and B. Ghita, "A novel multimodal biometric authentication system using machine learning and blockchain," in Selected Papers from the 12th International Networking Conference, Cham, Switzerland: Springer, 2021, pp. 31–46, doi: 10.1007/978-3-030- 64758-2 3.

[10]. Geng, S.-J. Huang, and S. Chen, "Recent advances in open set recognition: A survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 10, pp. 3614–3631, Oct. 2021, doi: 10.1109/TPAMI.2020.2981604.

[11]. P. Byahatti and M. S. Shettar, "Fusion strategies for multimodal biometric system using face and voice cues," IOP Conference Series: Materials Science and Engineering, vol. 925, no. 1, Art. no. 012031, Sep. 2020, doi: 10.1088/1757-899X/925/1/012031.

[12]. S. Barde, "A survey of multimodal biometrics system," Kala Sarovar (UGC Care Group-1 Journal), vol. 23, no. 02(II), pp. 91–95, Nov.–Dec. 2020.

[13]. K. Sasidhar, K. Vijayalakshmi, K. Ramakrishna, and K. Kailasa Rao, "Multimodal biometric systems – study to improve accuracy and performance," International Journal of Computer Science and Engineering Survey, vol. 1, no. 2, pp. 19–30, Nov. 2010, doi: 10.5121/ijcses.2010.1205.

[14]. Abozaid, A. Haggag, H. Kasban, and M. Eltokhy, "Multimodal biometric scheme for human authentication technique based on voice and face recognition fusion," Multimedia Tools and Applications, vol. 78, no. 12, pp. 16345–16361, Jun. 2019, doi: 10.1007/s11042-018-7012- 3.

[15]. W. J. Scheirer, A. Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 7, pp. 1757–1772, Jul. 2013, doi: 10.1109/TPAMI.2012.256.

[16]. Bendale and T. E. Boult, "Towards open world recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, Jun. 2015, pp. 1893–1902, doi: 10.1109/CVPR.2015.7298799.

[17]. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," Pattern Recognition, vol. 38, no. 12, pp. 2270–2285, 2005, doi: 10.1016/j.patcog.2005.01.012.

[18]. Bendale and T. E. Boult, "Towards open set deep networks," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1563–1572, doi: 10.1109/CVPR.2016.173.

[19]. M. Gunther, S. Cruz, E. M. Rudd, and T. E. Boult, "Toward open-set ¨ face recognition," arXiv preprint arXiv:1705.01567, 2017.

[20]. R. Peri, S. O. Sadjadi, and D. Garcia-

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2350-2361
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0317

Romero, "VoxWatch: An open-set speaker recognition benchmark on VoxCeleb," arXiv preprint arXiv:2307.00169, 2023.

[21]. Irfan, N. Lyubova, M. Garcia Ortiz, and T. Belpaeme, "Multi-modal open-set person identification in HRI," in Proc. ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI), 2018.

[22]. Y. Su, M. Kim, F. Liu, A. K. Jain, and X. Liu, "Open-set biometrics: Beyond good closed-set models," arXiv preprint arXiv:2407.16133, 2024, doi: 10.48550/arXiv.2407.16133.