TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 24, No. 2, April 2026, pp. 648~662
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v24i2.27356

Score-level biometric information fusion with generalized power mean

Leila Hellal[1], Naceur-Eddine Boukezzoula[1], Mohamed Cheniti[2], Zahid Akhtar[3]
[1]Laboratory of Intelligent Systems (LIS), Department of Electronic, Faculty of Technology, Ferhat Abbas University, Setif, Algeria
[2]Institute of Optics and Precision Mechanics, Ferhat Abbas University Setif-1, Setif, Algeria
[3]Department of Electrical and Computer Engineering, State University of New York Polytechnic Institute, Utica, United States

Article Info

Article history:

Received Jul 6, 2025
Revised Jan 17, 2026
Accepted Jan 30, 2026

Keywords:

Biometric authentication
Genuine acceptance rate
Half total error rate
National Institute of Standards and Technology fingerprint
Score level fusion
XM2VTS database

ABSTRACT

To overcome the fundamental shortcomings of single-trait biometric systems, multimodal solutions have gained considerable interest. In this work, a score-level fusion scheme for biometric authentication is introduced, where information from multiple modalities is combined using conventional mean operators such as arithmetic, harmonic, geometric, and quadratic means, with particular attention given to the power mean formulation. The proposed framework increases system robustness while preserving low computational complexity and requiring no training phase. Performance is assessed on three well-known public datasets: National Institute of Standards and Technology (NIST)-fingerprint, NIST-face, and XM2VTS, using standard score normalization methods and commonly employed evaluation metrics. The experimental analysis shows that the quadratic mean attains a genuine acceptance rate (GAR) of 91.50% on the NIST-fingerprint dataset, while the power mean with $\alpha = 5$ achieves 82.40% on NIST-face. Furthermore, the half total error rate (HTER) on XM2VTS is reduced to 0.059. In comparison with learning-based fusion techniques, the proposed approach provides a more straightforward, computationally efficient, and dependable alternative for real-world biometric applications.

Corresponding Author:

Leila Hellal
Laboratory of Intelligent Systems (LIS), Department of Electronics, Faculty of Technology
Ferhat Abbas University Setif-1
Setif, Algeria
Email: leila.hellal@univ-setif.dz

1. INTRODUCTION

The rapid growth of digital identity fraud has urged individuals and organizations to adopt reliable identity management solutions. Biometric systems authenticate individuals by examining distinctive physiological, behavioral, or chemical traits, and they are now commonly deployed in applications, including mobile platforms and border security systems [1]. A unimodal biometric system performs identity verification using only one biometric characteristic, such as facial features, fingerprints, or iris patterns. Although such systems can perform well under ideal conditions, dependence on one biometric source exposes them to several weaknesses. Common issues include limited population coverage, degraded or noisy samples, insufficient discriminatory power, variations within the same individual, and increased vulnerability to spoofing attacks [2]. To address these shortcomings, research has demonstrated that combining multiple biometric traits leads to more reliable and accurate recognition performance [3]. These systems, known as multimodal biometric systems, integrate information from different modalities at distinct stages of processing, either prior to matching (sensor-level and feature-level fusion) or after matching (score-level, rank-level, and decision-level fusion) [4], [5]. Sensor-level fusion is employed when multiple biometric traits are captured using different sensors or when several impressions of the same trait are obtained with a single sensor. For example, Cader *et al.* [6] demonstrated systems that use various fingerprint image segments to reconstruct complete fingerprints and compute their corresponding scores. Fusion at the feature level is generally expected to provide higher recognition performance [7]. However, this approach poses significant challenges due to the heterogeneity and incompatibility among feature representations of different modalities [7], [8]. For instance, combining iris and face features often requires complex normalization and dimensionality reduction techniques to address these issues [9]. On the other hand, decision-level fusion generally yields lower performance since it combines only the final binary outcomes from individual modalities. Because the raw matching scores and intermediate information are no longer available at this stage, a significant amount of discriminative data is lost, leading to reduced flexibility and accuracy. Moreover, this approach heavily depends on the thresholding strategies and decision rules applied to each classifier, which can introduce inconsistencies when modalities exhibit different reliability levels [10].

In score-level, also known as measurement-level, fusion, the matching scores produced by multiple biometric modalities are aggregated into a unified score. This combined value is subsequently evaluated by the verification or identification module to determine the final authentication outcome. This method is widely used as it offers an effective trade-off between computational efficiency and recognition accuracy when compared with feature and decision-level fusion. Its flexibility allows integration of heterogeneous biometric sources without the need for complex feature alignment or retraining of modality-specific models, making it suitable for real-world deployments. Recent studies continue to show that score-level fusion remains a very strong practical strategy in multimodal biometric systems, especially when combining disparate modalities or matchers where raw features are incompatible or unavailable. For example, the quality-guided mixture of score-fusion experts (QME) framework achieved state-of-the-art whole-body recognition by employing a learned score-fusion strategy that effectively managed variability in score distributions and data quality across modalities, outperforming baseline methods [11]. In contrast, recent feature-level fusion approaches, such as Vekariya *et al.* [12] employing a binary chimp-optimized adaptive kernel support vector machine (BCO-AKSVM), have demonstrated excellent accuracy and low equal error rate (EER), sometimes surpassing score-level methods. However, these methods come with increased computational cost and complexity in feature extraction and dimensionality reduction. Another comparative work [13] combines both feature and score-level fusion to gain robustness, showing that while feature fusion boosts information content, adding score-level fusion helps stabilize performance across datasets. Thus, recent evidence supports the idea that score-level fusion offers a very favorable trade-off (modularity, lower data sharing demands, easier implementation).

In general, fusion techniques fall into three main groups: transformation-based, density-based, and classifier-based methods [14]. Transformation-based approaches rescale raw matching scores onto a unified domain through the use of normalization methods, including min–max scaling and z-score, tanh functions, or double-sigmoid mappings [15]. The normalized scores are then combined using rules such as sum, weighted sum, min, max, triangular norms, or product [16]–[19]. Density-based fusion uses estimated probability distributions of genuine and impostor scores to calculate likelihood ratios [20], [21]. Classifier-based methods treat combined scores as feature vectors and classify them as genuine or impostor using models such as support vector machines (SVM) [22], [23], hidden Markov models (HMM) [24], multilayer perceptrons (MLP) [25], or Bayesian classifiers with Gaussian mixture models (GMMs) [26].

Building on these categories, research efforts have continually explored new strategies to enhance score fusion performance. Early approaches relied mainly on simple combination rules, such as the weighted sum [27], while later methods employed more advanced techniques like artificial neural networks (ANN) [28] and fuzzy logic systems (FLS) [29]. Numerous studies have confirmed the effectiveness of these approaches. For instance, Ammour *et al.* [30] reported that combining face and iris scores using a weighted sum significantly improves identification accuracy. Similarly, Alharbi and Alshanbari [31] found that fusing face and voice modalities at the score level reduces the EER. However, these improvements are often accompanied by higher computational demands and greater sensitivity to outliers. Weighted fusion methods require dataset-specific parameter tuning, which limits generalization, whereas max fusion, although efficient, can be unstable due to its reliance on the highest score, which may be noisy.

Recent studies have increasingly emphasized score-level fusion within deep learning frameworks, particularly convolutional neural network (CNN)-based architectures, which effectively learn cross-modal relationships and perform non-linear fusion. For instance, Riaz *et al.* [32] demonstrated that fusing CNN embeddings with auxiliary facial marks at the score level improved face verification accuracy on the IJB-A dataset. Similarly, Shinde and Kayte [33] showed that multimodal CNN fusion of face and fingerprint features using a weighted score combination outperformed unimodal recognition. Moreover, Farhadipour *et al.* [34] further confirmed the effectiveness of score-level fusion by integrating FaceNet and VoiceNet embeddings, achieving superior multimodal recognition compared to single-modality systems. Consequently, researchers are revisiting traditional fusion operators and statistical frameworks that offer competitive accuracy with lower computational costs, highlighting the need for a simple yet efficient fusion strategy that balances performance and complexity.

In response to these challenges, mean operators emerge as a simple yet powerful alternative for score-level fusion. Classical means; arithmetic, geometric, quadratic, and harmonic, serve as fundamental statistical tools capable of summarizing complex data while mitigating the effects of noise and outliers [35], [36]. Their simple mathematical structure and clear interpretation make mean-based fusion methods useful for different biometric setups, such as multi-sensor, multi-instance, and hybrid systems. The generalized versions, like power means [35], make these methods even more flexible by allowing the fusion process to adapt and balance between accuracy and computational cost. Unlike many advanced fusion techniques, mean-based methods do not need training or parameter adjustment, making them suitable, efficient, and reliable for real-world biometric systems.

This study explores the potential of mean operators for score-level fusion in multimodal biometric systems. Using three datasets: National Institute of Standards and Technology (NIST) fingerprint, NIST faces, and XM2VTS, we systematically evaluate multiple mean formulations in multi-algorithm, multi-instance, and multimodal scenarios. Our contributions are threefold: first, we demonstrate the effectiveness of mean-based operators in enhancing recognition performance while minimizing complexity; second, we provide a comparative analysis of different mean formulations across multiple datasets; and finally, we highlight their resilience to noise and outliers, positioning them as a practical solution for scalable, real-world biometric authentication.

2. PROPOSED METHOD

This work seeks to strengthen both the performance and dependability of biometric recognition by integrating scores from different modalities in a simple, adaptable, and training-free manner. Instead of relying on complex fusion algorithms, the proposed method uses a mathematical operator that can easily adjust its behavior to different biometric scenarios. By changing a single parameter, the same formulation can act like several well-known fusion rules, allowing the system to adapt when one biometric source is noisy or less reliable.

This concept is based on Kolmogorov's axiomatic theory of the mean [35], [37], which provides a unified foundation for designing and analyzing various mean functions used in score-level fusion.

Let X is an input vector defined as: $X = (x_1, \cdots, x_n)$. A regular mean is a function $M: R^n \to R$ that satisfies the following fundamental properties:

- $M(x)$ is continuous and monotonically increasing with respect to its arguments
- $M(x)$ is symmetric, meaning its value is invariant under any permutation of the input elements
- The mean of identical values returns the same value, i.e., for repeated inputs x_n, $M(x_n) = x_n$
- The mean of a combined dataset remains unchanged when a subset of values is replaced by their own mean $m = M(x_1, \cdots x_{n1})$:

$$M(X) = M(m, \cdots, x_{n1+1}, \cdots, x_n) \tag{1}$$

- The function M is idempotent, meaning that applying the mean to identical input values returns the same value:

$$\forall\, x \in [0, 1]; M(x, x) = x \tag{2}$$

- Associativity is a stronger quality than bisymmetry:

$$[0,1]^4; M[M(x, y), M(z, t)] = M[M(x, z), M(y, t)] \tag{3}$$

A mean that satisfies this requirement, while remaining continuous and strictly increasing, is typically represented by the following mathematical form:

$$\forall (x, y) \in [0,1]^2, M(x, y) = g^{-1}\left[\frac{g(x)+g(y)}{2}\right] \tag{4}$$

Kolmogorov showed that if these conditions hold, the function $M(x)$ can be represented as (5) [35]:

$$M_g(x) = g^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right) \tag{5}$$

where g is the continuous monotone, g^{-1} is the inverse function, and X is the vector: $X = (x_1, \cdots, x_n)$.

The function g adjusts the dynamic values prior to their integration. The modified values are subsequently combined using a simple arithmetic average, as demonstrated in the following examples [35]:

– Arithmetic mean: when the transformation function satisfies $g(x) = x$ leads to:

$$M_x = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (6)$$

– For the geometric mean, setting $g(x) = log(x)$ for $x > 0$ leads to:

$$M_{\log x}(x) = (\prod_{i=1}^{n} x_i)^{\frac{1}{n}} \quad (7)$$

– For the harmonic mean, setting $g(x) = \frac{1}{x}$ with $x > 0$ results in:

$$M_{\frac{1}{x}}(x) = \frac{n}{\sum_{i=1}^{n} x_i^{-1}} \quad (8)$$

As shown in (5) also encompasses other cases, including the power mean. For $\alpha \in [0, \infty]$, we have [35]:

$$M_{x^a}(x) = \left\{\frac{1}{n}\sum_{i=1}^{n} x_i^a\right\}^{\frac{1}{a}} M(X) = M(m, \cdots, x_{n1+1}, \cdots, x_n) \quad (9)$$

The Kolmogorov axiomatic framework ensures that the mean operators satisfy important mathematical properties such as monotonicity, continuity, and symmetry, which makes them suitable for reliably combining different biometric scores. Based on this framework, the generalized power mean introduces a parameter (α) that controls how the scores are combined. This allows one formula to represent several well-known fusion rules. When (α) is small or negative, the lowest score has more influence on the final result, which is useful when one biometric source may be weak or noisy. When (α) is positive and increases, the higher scores gain more influence, improving performance when one modality is strong and highly reliable. Therefore, by adjusting (α), the power mean can flexibly adapt the fusion behavior to different biometric conditions [35], [37]. As shown in Table 1, the generalized mean can smoothly transition from minimum to maximum behavior, depending on the chosen value of (α).

Table 1. Illustrations of various mean functions corresponding to different values of (a) [35]

a	M (x, y)	Comment
- ∞	min (x, y)	Limit value
-1	$\frac{2xy}{x+y}$	Harmonic mean
0	$(xy)^{\frac{1}{2}}$	Geometric mean
+1	$\frac{x+y}{2}$	Arithmetic mean
+2	$\sqrt{\frac{x^2+y^2}{2}}$	Quadratic mean
+∞	max (x, y)	Limit value

To evaluate the benefit of emphasizing high-quality information, this study also investigates the fusion performance using a power mean with ($\alpha = 5$). This configuration increases the influence of strong and reliable scores while suppressing the effect of uncertain or noisy ones [38]. In practical biometric scenarios, where the quality of fingerprint or face samples may fluctuate. This strategy enhances recognition accuracy without increasing computational complexity, thereby maintaining the simplicity and efficiency of the proposed method.

The mean operator is also practical from a computational point of view. Since it does not require learning, weighting, or complex comparisons, it can be calculated very quickly. This makes it suitable for real-time biometric systems where fast decision-making is needed [38]. In addition, the fusion can be processed in parallel for multiple users or multiple score inputs, which helps maintain high system performance even when the number of users increases.

Figure 1 outlines the proposed score-level fusion process using mean operators. The system begins by capturing biometric data, such as a fingerprint and a face image, from the same individual. Each biometric source is processed separately to extract discriminative features. These features are then compared with the corresponding enrolled templates in the database to generate matching scores [1], [4], [5]. Because the scores may come from different systems and may not share the same numerical scale, a normalization step is applied to place them within a unified range [15], [16], [39]. After normalization, the scores are combined using the generalized mean operator to obtain a single fused score that represents the final matching confidence. This fused score is then evaluated against a predefined threshold. If it meets or exceeds this threshold, the user is granted access; otherwise, the attempt is denied. By incorporating information from multiple biometric sources in this structured way, the system enhances decision accuracy and robustness.

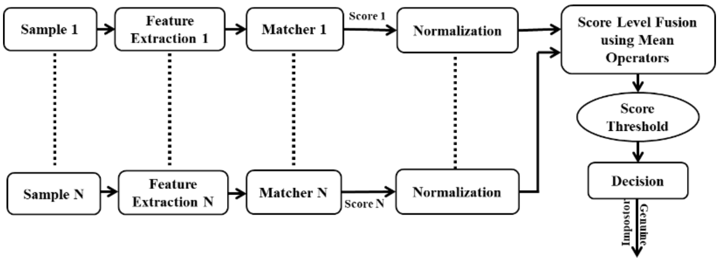

Figure 1. Overview of the proposed score-level fusion framework using mean operators

Power mean fusion, also called the generalized mean, has recently become popular because it can flexibly combine different kinds of information. By adjusting one parameter, the operator can change how strongly it reacts to high or low input values, making it useful in many applications. For example, in deep learning, it is used in the generalized mean (GeM) pooling technique to balance the behavior of average and max pooling, which helps improve feature representation in image retrieval tasks [40], [41]. In image processing, power mean fusion has been successfully applied in multi-exposure, multi-focus, and medical image fusion, where controlling the power parameter helps achieve better contrast and detail [42]. These strengths make the power mean a promising choice for biometric systems, where information from different sources must be combined efficiently. Its adaptable structure is particularly suitable for score-level fusion, helping to produce a single, reliable score that improves the final recognition decision.

3. METHOD

This section presents the datasets, experimental setup, and evaluation metrics used to assess the performance of the proposed fusion approach. The aim is to guarantee that the results are reliable, unbiased, and reproducible.

3.1. Database and experimental protocols

Evaluation of the proposed approach was conducted using three biometric datasets: the NIST fingerprint database, the NIST face database, and the XM2VTS multimodal database. These datasets contain fingerprint and face samples from a large number of users, allowing us to evaluate both single-modality and multimodal fusion performance under different conditions.

3.1.1. NIST-fingerprint database

The NIST fingerprint database is a well-known and publicly available resource that includes fingerprint samples from many individuals. It provides two fingerprint scores for each person, one from the right index finger and one from the left index finger. Since each user has two samples, a genuine comparison is made by matching the index fingers of both the right and left hands of the same person, resulting in 6,000 genuine scores (one per user). To generate impostor scores, fingerprints from different users are compared. Each user's fingerprint is matched against those of 5,999 other users, producing a total of 35,994,000 impostor scores. Overall, the dataset comprises 6,000 genuine match scores and 35,994,000 impostor match scores [43], [44].

3.1.2. NIST-face database

The NIST-Face database is an important resource for facial recognition research, providing detailed similarity score vectors for in-depth analysis. It contains two distinct sets, each comprising 6,000 samples collected from 3,000 individual users. Similarity scores were generated using two different face recognition algorithms, referred to as matcher C and matcher G. Overall, the database includes 6,000 genuine match scores, corresponding to one comparison per subject pair, and 17,994,000 impostor match scores, resulting from comparisons of each sample with the remaining 2,999 subjects [43], [44]. To ensure uniformity and facilitate fair comparison between the two matchers, tanh-estimator normalization was applied to their respective similarity scores. This normalization step enhances the consistency and reliability of the performance evaluation for both face-matching systems.

3.1.3. The XM2VTS dataset and the evaluation protocols

For both speaker and face verification tasks, the XM2VTS database was utilized according to the protocol described in [45]. The dataset includes 295 individuals, divided into 200 genuine subjects, 70 test impostors, and 25 evaluation impostors. For face analysis, three feature types were employed: face histogram (FH), discrete cosine transform small (DCT_s) (small DCT features from 40×32 images), and DCT_b (larger DCT features from 80×64 images). For speech, features included linear filter-bank cepstral coefficients (LFCC), phase auto-correlation mel filter-bank cepstral coefficients (PAC-MFCC), and spectral subband centroid (SSC), all combined with temporal derivatives to improve performance. To ensure systematic training and evaluation, the database was divided into three subsets:

- LP train: this set trains client models by learning patterns from genuine user data.
- LP eval: this set determines decision thresholds, helping classifiers distinguish between genuine users and impostors.
- LP test: this final set assesses overall system performance using unseen data.

The training and evaluation procedures are defined by two protocols: Lausanne protocol I (LP1) and Lausanne protocol II (LP2). Both protocols use the same test set [45], [46], ensuring fair performance comparison. LP1 provides eight samples per user, whereas LP2 includes five Table 2. To preserve data integrity, one faulty speech file and 70 test impostor entries were removed, yielding a total of 111,800 impostor scores [47].

Table 2. Overview of the Lausanne protocols for the XM2VTS database [47]

Data sets	Lausanne protocols		Fusion protocol
	LP1	LP2	
Train client accesses	3	4	-
Eval client accesses	600 (3×200)	400 (2×200)	Fusion dev
Eval impostor accesses	40.000 (25×8×200)		Fusion dev
Test client accesses	400 (2×200)		Fusion eva
Test impostor accesses	112.000(70×8×200)		Fusion eva

For classification, two machine learning models are employed: GMMs and MLPs. GMMs are probabilistic models that handle variable-length feature sequences by first training a universal background model with the expectation-maximization (EM) algorithm, then adapting it to each user through maximum-a-posteriori (MAP) adaptation [48]. In contrast, MLPs are trained with stochastic error backpropagation on fixed-length feature vectors to distinguish genuine users from impostors [49]. Both models output a match score representing the likelihood that the input matches the claimed identity. This score is compared against a decision threshold, calibrated with the LP evaluation set, to balance the false acceptance rate (FAR) and false rejection rate (FRR). High scores indicate genuine matches, while low scores suggest impostors. GMMs are particularly effective for variable-length speech data, whereas MLPs excel at capturing discriminative patterns in fixed-size face features [50]. Combining their strengths enhances score accuracy, verification performance, and robustness against impostor attempts.

3.2. Normalization and evaluation metrics

Before applying the mean-based fusion rule, the matching scores from different biometric systems must be normalized to ensure they are comparable. In this study, min-max normalization is applied due to its simplicity and effectiveness when the score range is known [15], [39]. The transformation maps all scores into the interval [0, 1], as expressed in:

$$S' = \frac{S - S_{min}}{S_{max} - S_{min}} \tag{10}$$

where S' the normalized score, S is the original score; S_{min} and S_{max} are the minimum and maximum scores, respectively.

After normalization and fusion, system performance is evaluated using standard biometric verification metrics. The FAR measures how often an impostor is incorrectly accepted as a genuine user, whereas the FRR measures how often a genuine user is incorrectly rejected [51], [52]. They are calculated as follows:

$$FAR = \frac{FA}{NI} \tag{11}$$

$$FRR = \frac{FR}{NC} \tag{12}$$

False acceptance (FA) and false rejection (FR) represent false acceptance and false rejection, respectively, while number of impostor accesses (NI) and the number of client accesses (NC) correspond to the total client and impostor accesses [52], [53]. In practical terms, a secure biometric system should maintain a very low FAR to reduce the risk of unauthorized access, while also achieving a low FRR to ensure user convenience and system usability. To further assess performance, the genuine acceptance rate (GAR) represents the fraction of genuine users correctly accepted:

$$GAR = 1 - FRR \tag{13}$$

For a comprehensive performance comparison, GAR and FAR curves are plotted as receiver operating characteristic (ROC) curves. A ROC curve positioned nearer to the upper-left corner signifies a higher-performing system with a lower rate of acceptance errors [52], [54]. The decision threshold plays a key role in determining the balance between security (low FAR) and accessibility (low FRR). Lowering the threshold increases convenience but may weaken security, while increasing it leads to stronger security at the cost of more rejections. By analyzing these curves, the proposed fusion strategy demonstrates its ability to improve the performance of the individual modalities, providing enhanced discrimination capability without increasing system complexity. For the XM2VTS multimodal database, the system's performance is assessed using the half total error rate (HTER), which is the primary metric specified in the XM2VTS benchmark protocol, defined as (14).

$$HTER = \frac{FAR+FRR}{2} \tag{14}$$

A lower HTER value reflects a more accurate and secures system, as it indicates both fewer false acceptances and fewer false rejections [45], [47]. This makes HTER particularly suitable for multimodal biometric evaluation, where a strong balance between usability and security is required.

To evaluate whether the performance improvement achieved by the proposed fusion method is statistically significant, a paired t-test is conducted between the best individual matcher and the fused scores [55], [56]. This test examines whether the mean difference between the two sets of genuine scores is greater than zero. The t-statistic is calculated using the (15).

$$t = \frac{\bar{d}}{S_d/\sqrt{n}} \tag{15}$$

where $\bar{d}$ represents the mean difference between paired genuine scores, S_d denotes the standard deviation of these differences, and n is the total number of paired genuine comparisons [56]. P-values below 0.05 are regarded as statistically significant, suggesting that the observed performance gains arise from the fusion strategy and not from random effects.

4. EXPERIMENTAL RESULTS AND DISCUSSION

This section reports the experimental assessment of the proposed mean-based score fusion approach. The fusion performance is analyzed using standard biometric metrics, and the impact of the power parameter (α) on recognition accuracy is assessed. The following subsections detail the results obtained on each dataset.

4.1. Performance of mean operator-based fusion on NIST fingerprint database

The NIST fingerprint database is used first to evaluate the performance of the proposed score-level fusion approach in a multi-instance biometric scenario. Since the left and right fingerprints of the same user may vary in quality, relying on only one instance can lead to reduced accuracy or inconsistent decisions. By combining the matching scores through the generalized mean operator, the system can benefit from the strengths of both fingerprint instances and obtain a more reliable identification outcome. The following analysis compares the performance of individual instances against several mean-based fusion rules to highlight the benefits of the proposed approach.

Figure 2 presents the ROC curves derived from the NIST fingerprint database for both the left and right index fingers, along with the proposed score-level fusion using mean operators. Individually, the left and right index fingers achieve GARs of 75.5% and 83%, respectively, at a FAR of 0.01%, highlighting the performance variation between the two biometric traits.

When the mean operator-based fusion is applied, the overall recognition improves significantly. The results show that the arithmetic mean reaches a GAR of 91.20%, while the harmonic mean yields 85.50% at the same security threshold. The best performance is achieved using the quadratic mean with a GAR of 91.50%, followed closely by the power mean with ($\alpha = 5$) at 90.90%. These improvements confirm that score-level fusion effectively exploits complementary information in multi-instance fingerprints.

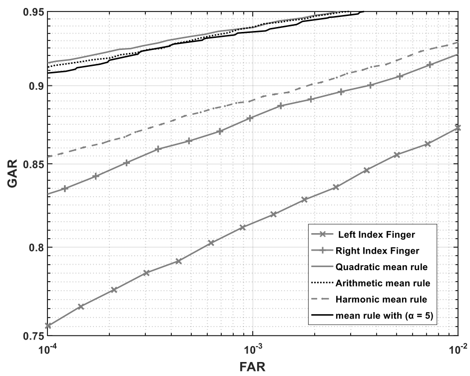

Figure 2. ROC curves of individual matchers and mean-based score-level fusion techniques

The ROC curves also shift noticeably toward the upper-left corner compared to the individual modalities, indicating stronger discrimination capability and reduced acceptance errors. The superior performance of the quadratic and power means can be explained by their ability to give more weight to high-confidence scores, while minimizing the negative influence of uncertain matches, an observation consistent with prior findings in multimodal fusion studies [10], [38], [44]. Table 3 reinforces these observations by demonstrating that the quadratic and high-power mean operators consistently achieve lower error rates than the unimodal fingerprint systems.

Table 3. Performance comparison of various fusion approaches on the NIST-fingerprint database

Score-level fusion methods	GAR (%)
Max rule [57]	90.30
Min rule [57]	79.60
SVM [58]	91.40
Likelihood ratio [58]	91.40
Entropy-with-Frank $p = 0.01$ [59]	87.77
Entropy-with-Hamacher $p = 0.01$ [59]	85.77
S-sum using max rule [44]	90.75
S-sum using probabilistic t-norm [44]	89.00
S-sum using Hamacher t-norm [44]	75.50
S-sum using Yager t-norm with $p = 10.3$ [44]	90.00
S-sum using Schweizer and Sklar t-norm with $p = 0.9$ [44]	89.00
Weighted quadratic arithmetic mean (WQAM) using cos(s) with $r = 11$ [58]	91.60
Proposed system with arithmetic mean	**91.20**
Proposed system with harmonic mean	**85.50**
Proposed system with quadratic mean	**91.50**
Proposed system with mean rule ($\alpha = 5$)	**90.90**

The results demonstrate that adjusting the parameter (α) allows the generalized mean operator to progressively enhance recognition performance. The best result is obtained using the quadratic mean, which reaches a GAR of 91.50% at a FAR of 0.01%, outperforming the individual fingerprint instances and the other tested operators. As presented in Table 3, this performance also exceeds the results reported in earlier studies, confirming the effectiveness of the proposed fusion strategy in multi-instance fingerprint recognition. These observations align with the findings of Abderrahmane *et al.* [60], supporting the claim that emphasizing high-confidence scores leads to more stable and accurate decisions. Additionally, the method is computationally lightweight, avoiding complex non-linear functions and requiring no training phase, which makes it appropriate for use in real-time biometric applications. Overall, these results show that the proposed mean operator-based fusion provides a practical and efficient approach to improving recognition reliability and system security.

4.2. Evaluation of mean-based fusion methods on the NIST face database

This section presents the performance evaluation of the proposed score-level fusion method using the NIST face database in a multi-algorithm biometric setup. Figure 3 shows the ROC curves obtained from two independent face matchers, referred to as matcher C and matcher G. At a strict security threshold of FAR = 0.01%, the matchers achieve GARs of 63.00% and 72.10%, respectively. This noticeable performance gap indicates that each matcher captures different discriminative facial information, making the fusion of their scores highly beneficial.

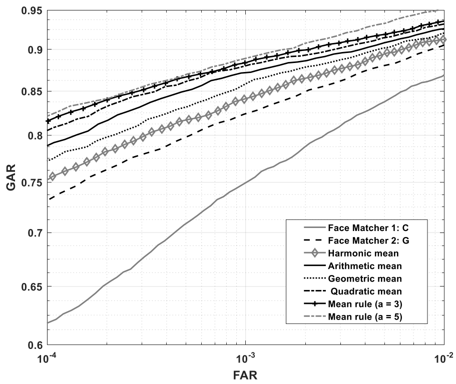

Figure 3. ROC curves of single matchers and mean-based score-level fusion techniques

To assess the adaptability of the proposed fusion strategy, six score combination settings were evaluated by varying the power parameter (α). As reported in Table 4, increasing α consistently enhances recognition accuracy. This confirms that the generalized power mean effectively balances score contribution according to the reliability of each matcher.

Table 4. Comparison of various fusion methods on the NIST face database

Score-level fusion methods	GAR (%)
Max rule [57]	71.50
Min rule [57]	73.15
Hamacher t-norm [57]	75.90
SVM [58]	77.30
Likelihood ratio [58]	77.20
S-sum using max rule [44]	76.00
S-sum using Hamacher t-norm [44]	75.75
Asym-AO2 using Aczel-Allsina with $p = 0.7$ and $m = 3$ [60]	76.20
Asym-AO2 using Algebraic product with $m = 1.2$ [60]	76.33
Asym-AO2 using Hamacher with $m = 10$ [60]	76.10
Proposed system with arithmetic mean	**78.75**
Proposed system with quadratic mean	**80.63**
Proposed system with harmonic mean	**75.10**
Proposed system with geometric mean	**77.50**
Proposed system with mean rule ($a = 3$)	**81.70**
Proposed system with mean rule ($a = 5$)	**82.40**

Among all tested configurations, ($\alpha = 5$) achieved the highest performance, reaching a GAR of 82.40% at FAR = 0.01%. This improvement demonstrates that assigning stronger influence to high-confidence matches leads to more discriminative and reliable fusion outcomes. Compared to traditional rules such as max [57], as well as machine learning-based methods like SVM or likelihood ratio [48], [58]. The proposed operator offers a better trade-off between performance, simplicity, and computational efficiency. Therefore, the power mean provides a practical and effective solution for enhancing score-level fusion in face biometrics.

The comparison between the proposed score-level fusion strategy and individual biometric matchers shows a clear improvement in recognition performance. Across all tested values of (α), the generalized mean consistently surpasses both unimodal systems and conventional score fusion techniques. This advantage arises from its simplicity, robustness, and low computational cost. Unlike weighted approaches, decision-level fusion, or machine learning-based methods [48], [61], [62], the proposed operator does not require training, parameter optimization, or complex score transformation. Its built-in resistance to outliers reduces the influence of noisy biometric samples, while still retaining the most informative characteristics of each matcher. As a result, the proposed method is well adapted to practical scenarios, especially where computational capabilities are limited or biometric data quality varies.

To analyze the behavior of the match scores on the NIST Face database, descriptive statistics were computed for both genuine and impostor distributions before fusion and after applying the mean-rule strategies (arithmetic, geometric, quadratic, and harmonic means). The results show that the mean-based fusion improves class separability by slightly raising the average genuine scores and lowering the average impostor scores. A summary of the detailed statistical results for each fusion rule is provided in Table 5.

Table 5. Genuine vs impostor score statistics for different mean-rule fusion methods

Mean rule	Genuine mean (μ_g)	Genuine std (σ_g)	Impostor mean (μ_i)	Impostor std (σ_i)
Arithmetic mean	0.7909	0.0567	0.6252	0.0784
Geometric mean	0.7827	0.0613	0.5907	0.0954
Harmonic Mean	0.7660	0.0804	0.5191	0.1072
Quadratic mean	0.7807	0.0569	0.6153	0.0768

Overall, despite some differences in score dispersion, all mean operators contribute to better class separability compared to the unfused baseline, where genuine and impostor average scores were 0.6836 and 0.3940, respectively. To statistically compare the improvements, paired t-tests were applied along with effect size and 95% confidence intervals [55], [56]. The analysis revealed a statistically significant difference between the operators. For the genuine scores, the test yielded t (2999) = 111.57, $p < 0.00001$, while for impostor scores it gave t (8,996,999) = 11573.14, $p < 0.00001$. The confidence intervals also indicate stable impostors. Because the dataset is extensive, extremely small p-values are expected even with small numerical differences. Therefore, additional performance indicators, such as the ROC curve, were included to better reflect the practical improvement achieved by fusion. Overall, these findings demonstrate that mean-based score fusion enhances discrimination capability and improves impostor rejection, leading to a more reliable biometric verification system under real-world conditions.

4.3. Performance of mean operator-based fusion on XM2VTS database

The XM2VTS experiments were performed following the standard Lausanne protocols LP1 and LP2, where the system is divided into eight baseline configurations in LP1 and five baseline configurations in LP2. Each baseline consists of a binary recognition component formed by combining a specific feature extraction method with an appropriate classifier. To assess the adaptability of the proposed fusion methodology, three score-level fusion strategies were applied across different baseline combinations, as described in [45]. Tables 6 and 7 present the results of 32 fusion configurations for LP1 and LP2, respectively. Subsection (a) reports multimodal fusion across face and speech modalities, while subsections (b) and (c) focus on classifier-based fusion and fusion across heterogeneous feature sets. This experimental setup allows a comprehensive evaluation of how mean-based score fusion enhances recognition performance under different biometric conditions.

In this experiment, three mean operators were evaluated for score-level fusion: the harmonic, geometric, and quadratic means. The fusion performance was quantified using the HTER, and the results were compared with those reported in [48] to assess the benefit of the proposed method. Tables 5 and 6 present the HTER values obtained for LP1 and LP2, respectively.

To analyze the influence of the power parameter, additional experiments were conducted using three values of α (−1, 0, and 2), representing different fusion behaviors ranging from conservative to confidence-focused weighting. These results highlight how adjusting (α) directly affects multimodal score aggregation and helps identify the most effective configuration. Together, these analyses provide a comprehensive evaluation of the proposed fusion strategy on the XM2VTS database.

In our experiments, the proposed mean-based fusion further reduced the error rates for these strong baseline matchers. The best improvement was achieved using the quadratic mean ($\alpha = 2$), where the HTER for the (DCT_b, GMM) and (LFCC, GMM) combination decreased to 0.493. This confirms that the proposed strategy enhances decision reliability by strengthening confident matches and minimizing the influence of less reliable scores. These findings show that combining a suitable biometric matcher pair with the power-controlled mean operator leads to a more accurate and robust multimodal verification system than the state-of-the-art baseline methods.

Table 7 presents the results obtained under Protocol 2. Similar to LP1, the proposed mean-based fusion method improved the verification accuracy for most binary systems. While overall HTER values in LP2 are slightly higher due to stricter evaluation conditions, the quadratic mean continued to provide the best performance in many combinations. These results confirm that the fusion strategy remains effective even when the recognition task becomes more challenging. This stability across protocols demonstrates that the mean operators not only increase accuracy but also enhance the general robustness of the system.

Table 6. HTER-based performance comparison using mean operators for LP1

Fusion technique	Fusion candidates	HTER			
		Mean	$\alpha = -1$	$\alpha = 0$	$\alpha = 2$
Fusion with different modalities	(FH, MLP); (LFCC, GMM)	0.782	1.720	1.559	**0.595**
	(FH, MLP); (PAC, GMM)	1.120	1.734	1.663	1.262
	(FH, MLP); (SSC, GMM)	0.871	1.750	1.697	**0.731**
	(DCT_s, GMM); (LFCC, GMM)	0.543	0.599	0.606	**0.533**
	(DCT_s, GMM); (PAC, GMM)	1.436	1.429	1.419	**1.409**
	(DCT, GMM); (SSC, GMM)	1.146	1.613	1.280	1.152
	(DCT_b, GMM); (LFCC, GMM)	0.511	0.590	0.525	**0.493**
	(DCT_b, GMM); (PAC, GMM)	1.021	1.014	1.115	1.399
	(DCT_b, GMM); (SSC, GMM)	0.752	0.767	0.751	0.828
	(DCT_s, MLP); (LFCC, GMM)	0.840	2.143	1.658	**0.828**
	(DCT_s, MLP); (PAC, GMM)	1.138	2.885	2.276	1.360
	(DCT_s, MLP); (SSC, GMM)	1.333	3.217	2.556	**1.343**
	(DCT_b, MLP); (LFCC, GMM)	1.523	5.085	4.580	2.428
	(DCT_b, MLP); (PAC, GMM)	3.664	5.536	4.968	**2.725**
	(DCT_b, MLP); (SSC, GMM)	3.108	6.052	5.588	**2.994**
Fusion with different features	(FH, MLP); DCT_s, GMM)	1.280	1.756	1.731	1.429
	(FH, MLP); (DCT_b, GMM)	1.122	1.745	1.741	1.305
	(FH, MLP); (DCT_s, MLP)	1.513	1.595	1.480	1.634
	(FH, MLP); (DCT_b, MLP)	1.960	2.268	1.657	2.404
	(LFCC, GMM); (SSC, GMM)	1.595	1.104	1.213	1.716
	(PAC, GMM); (SSC, GMM)	4.225	4.820	4.794	4.463
Fusion with different classifiers	(DCT_s, GMM); (DCT_s, MLP)	2.388	3.354	3.093	2.492
	(DCT_b, GMM); (DCT_b, MLP)	3.063	5.662	5.365	3.337

Table 7. HTER-based performance comparison using mean operators for LP2

Fusion technique	Fusion candidates	HTER			
		Mean	$\alpha = -1$	$\alpha = 0$	$\alpha = 2$
Fusion with different modalities	(FH, MLP); (LFCC, GMM)	1.122	1.487	1.189	**0.460**
	(FH, MLP); (PAC, GMM)	1.513	1.638	1.470	**0.621**
	(FH, MLP); (SSC, GMM)	1.960	1.750	1.697	**0.731**
	(DCT_b, GMM); (LFCC, GMM)	1.836	**0.059**	0.133	0.251
	(DCT_b, GMM); (PAC, GMM)	2.388	**0.366**	0.395	0.710
	(DCT_b, GMM); (SSC, GMM)	3.672	0.322	**0.237**	0.782
Fusion with different features	(FH, MLP); (DCT_s, GMM)	1.280	1.760	1.475	**0.831**
	(LFCC, GMM); (SSC, GMM)	3.063	0.839	**0.821**	1.577
	(PAC, GMM); (SSC, GMM)	2.934	3.407	3.068	3.097

The HTER values achieved under Protocol 2 are significantly lower than those obtained in Protocol 1, indicating that LP2 represents a more favorable evaluation setup for multimodal fusion and leads to more reliable verification performance. Notably, the fusion of face and speech scores resulted in an HTER as low as 0.059, which represents the best performance obtained across all XM2VTS experiments. This demonstrates that Protocol 2 provides a more favorable setup for fusion performance because of the improved enrollment and testing conditions. Overall, these results show that reinforcing reliable biometric information through the quadratic mean leads to a more accurate and stable verification system under LP2.

Beyond the top-performing systems, most other binary combinations also showed noticeable improvements through the proposed mean-based fusion mechanism. Even when individual modalities were relatively weak, the reduction in HTER indicates that strengthening more reliable scores during aggregation consistently enhances verification accuracy across heterogeneous matcher pairs. These findings highlight the adaptability of the generalized mean, which achieved performance gains for the majority of tested configurations under both LP1 and LP2.

Across all three datasets, the proposed fusion approach consistently delivered superior performance compared to baseline and conventional fusion strategies. On the NIST fingerprint dataset, the quadratic mean outperformed earlier techniques such as weighted sum and t-norm fusion [57], [59], while achieving accuracy on par with SVM-based fusion [58] but with lower computational complexity. A similar trend was observed on the NIST face dataset, where the power mean with ($\alpha = 5$) provided a clear improvement over individual matchers and previously reported mean-rule fusion outcomes [60]. On the XM2VTS database, the quadratic mean also produced the strongest verification accuracy under the LP2 protocol, performing at a level consistent with, and in some cases better than, earlier findings reported in [47].

Compared to more advanced machine learning fusion methods, the findings show that mean-based operators deliver a strong balance between performance and efficiency. Although learning-based approaches can capture more complex relationships between modalities, they typically require large training datasets, careful parameter tuning, and powerful hardware. The generalized mean, on the other hand, has simple linear complexity, is naturally resistant to noisy matching scores, and can be executed quickly in real time. This makes it an appealing solution for large deployments or environments with limited computational resources.

In addition to enhancing recognition accuracy, security remains a critical aspect of biometric systems. The application of mean operators in score-level fusion not only improves overall performance but also increases robustness against spoofing attacks. By combining information from multiple biometric traits, these operators make it significantly harder for an attacker to successfully replicate all modalities at once. Nevertheless, threats like fake samples or score manipulation can still occur. To address these, additional safeguards such as liveness detection, spoofing-aware machine learning classifiers, and anomaly detection mechanisms should be integrated. Furthermore, encrypting score data and applying secure normalization techniques can help prevent tampering. Together, these strategies make mean-based fusion systems both accurate and more resilient to spoofing attacks.

5. CONCLUSION

This study introduces a score-level fusion approach for multimodal biometric authentication using extended mean operators. The introduced framework eliminates the need for training or complex parameter estimation, resulting in a computationally efficient and scalable approach suitable for real-world systems. Experimental results on three benchmark datasets demonstrate consistent improvements over individual modalities and conventional fusion techniques. On the NIST-fingerprint dataset, the quadratic mean achieved a GAR of 91.50% at FAR = 0.01%, indicating a strong balance between recognition accuracy and system security. For the NIST-face dataset, the power mean with ($\alpha = 5$) yielded the best performance, achieving a GAR of 82.40% and outperforming baseline matchers. On the XM2VTS dataset, the quadratic mean reduced the HTER to 0.059, confirming its effectiveness in minimizing both false acceptances and false rejections.

Regarding security, although score-level fusion cannot completely prevent spoofing or adversarial attacks, the averaging behavior of mean operators inherently mitigates the impact of abnormal or corrupted scores, enhancing system resilience. However, a key limitation of mean-based fusion is that it treats all modalities equally and does not account for differences in individual matcher reliability. This can reduce performance when some modalities are significantly less accurate or noisier, as the averaging process may dilute the contribution of the more reliable scores.

In practical deployment, the proposed approach can be easily integrated into existing multimodal systems used in areas such as border control, secure access, and mobile identity verification. Future work will focus on optimizing parametric formulations of the mean operator and evaluating the proposed strategy under challenging scenarios, including noisy or partially degraded biometric samples, to further enhance reliability and security in biometric authentication systems.

FUNDING INFORMATION

The authors declare that this research did not receive any research grant or contract from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Leila Hellal	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	✓
Naceur-Eddine Boukezzoula		✓		✓		✓	✓		✓	✓	✓	✓	✓	
Mohamed Cheniti	✓	✓	✓	✓			✓	✓		✓	✓			
Zahid Akhtar					✓	✓				✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author [LH] on request.

REFERENCES

[1] U. Sumalatha, K. K. Prakasha, S. Prabhu, and V. C. Nayak, "A comprehensive review of unimodal and multimodal fingerprint biometric authentication systems: Fusion, attacks, and template protection," *IEEE Access*, vol. 12, pp. 64300–64334, 2024, doi: 10.1109/ACCESS.2024.3395417.

[2] M. Gayathri, C. Malathy, H. A. Chandrasekar, and P. Mathialagan, "Multimodal biometric systems, its security issues, research challenges and countermeasures—technical review," in *Inventive Communication and Computational Technologies*, 2022, pp. 993–1010, doi: 10.1007/978-981-16-5529-6_75.

[3] K. Krishna Prakasha and U. Sumalatha, "Privacy-preserving techniques in biometric systems: Approaches and challenges," *IEEE Access*, vol. 13, pp. 32584–32616, 2025, doi: 10.1109/ACCESS.2025.3541649.

[4] N. Bala, R. Gupta, and A. Kumar, "Multimodal biometric system based on fusion techniques: a review," *Information Security Journal: A Global Perspective*, vol. 31, no. 3, pp. 289–337, May 2022, doi: 10.1080/19393555.2021.1974130.

[5] J. O. Atoyebi, N. S. Okomba, I. A. Adeyanju, A. O. Akinrotimi, O. O. Ajayi, and O. O. Owolabi, "Multi-biometric recognition systems: A comprehensive review," *Adeleke University Journal of Engineering and Technology*, vol. 8, no. 1, pp. 10–38, 2025.

[6] A. J. M. A. Cader, J. Banks, and V. Chandran, "Fingerprint systems: Sensors, image acquisition, interoperability and challenges," *Sensors*, vol. 23, no. 14, Jul. 2023, doi: 10.3390/s23146591.

[7] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional feature fusion," *Prepr. arXiv.2009.14082*, Nov. 2020.

[8] 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, Mar. 2023, doi: 10.1016/j.inffus.2022.10.032.

[9] M. H. Safavipour, M. A. Doostari, and H. Sadjedi, "A hybrid approach to multimodal biometric recognition based on feature-level fusion of face, two irises, and both thumbprints," *Journal of Medical Signals & Sensors*, vol. 12, no. 3, pp. 177–191, Jul. 2022, doi: 10.4103/jmss.jmss_103_21.

[10] P. Szczuko, A. Harasimiuk, and A. Czyżewski, "Evaluation of decision fusion methods for multimodal biometrics in the banking application," *Sensors*, vol. 22, no. 6, Mar. 2022, doi: 10.3390/s22062356.

[11] J. Zhu, Y. Su, M. Kim, A. Jain, and X. Liu, "A quality-guided mixture of score-fusion experts framework for human recognition," *Prepr. arXiv.2508.00053*, Jul. 2025, [Online]. Available: http://arxiv.org/abs/2508.00053

[12] V. Vekariya, M. Joshi, S. Dikshit, and S. K. M. Bargavi, "Multi-biometric fusion for enhanced human authentication in information security," *Measurement: Sensors*, vol. 31, Feb. 2024, doi: 10.1016/j.measen.2023.100973.

[13] Y. Wu and J. Hu, "Deep and shallow feature fusion in feature score level for palmprint recognition," *IET Biometrics*, vol. 2024, no. 1, Jan. 2024, doi: 10.1049/2024/5683547.

[14] M. Mondal, M. Khayati, H.-Â. Sandlin, and P. Cudré-Mauroux, "A survey of multimodal event detection based on data fusion," *The VLDB Journal*, vol. 34, no. 1, Jan. 2025, doi: 10.1007/s00778-024-00878-5.

[15] R. Pathak, P. S. Prasad, V. K. Gunjan, and V. K. Solanki, "Normalization techniques in multi modal biometric," in *ICCCE 2019*, 2020, pp. 425–431, doi: 10.1007/978-981-13-8715-9_51.

[16] S. N. Tumpa and M. L. Gavrilova, "Score and rank level fusion algorithms for social behavioral biometrics," *IEEE Access*, vol. 8, pp. 157663–157675, 2020, doi: 10.1109/ACCESS.2020.3018958.

[17] P. S. Sanjekar and J. B. Patil, "Wavelet based multimodal biometrics with score level fusion using mathematical normalization," *International Journal of Image, Graphics and Signal Processing*, vol. 11, no. 4, pp. 63–71, Apr. 2019, doi: 10.5815/ijigsp.2019.04.06.

[18] A. Herbadji *et al.*, "Combining Multiple Biometric Traits Using Asymmetric Aggregation Operators for Improved Person Recognition," *Symmetry*, vol. 12, no. 3, Mar. 2020, doi: 10.3390/sym12030444.

[19] P. Punyani, R. Gupta, and A. Kumar, "A multimodal biometric system using match score and decision level fusion," *International Journal of Information Technology*, vol. 14, no. 2, pp. 725–730, Mar. 2022, doi: 10.1007/s41870-021-00843-3.

[20] N. Susyanto, R. Veldhuis, L. Spreeuwers, and C. Klaassen, "Semiparametric likelihood-ratio-based biometric score-level fusion via parametric copula," *IET Biometrics*, vol. 8, no. 4, pp. 277–283, Jul. 2019, doi: 10.1049/iet-bmt.2018.5106.

[21] L. M. Pereira, A. Salazar, and L. Vergara, "A comparative study on recent automatic data fusion methods," *Computers*, vol. 13, no. 1,

Dec. 2023, doi: 10.3390/computers13010013.
[22] M. T. Ghazal and K. Abdullah, "Face recognition based on curvelets, invariant moments features and SVM," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, pp. 733–739, Apr. 2020, doi: 10.12928/telkomnika.v18i2.14106.
[23] V. Karthikeyan and S. S. Priyadharsini, "Adaptive boosted random forest-support vector machine based classification scheme for speaker identification," *Applied Soft Computing*, vol. 131, Dec. 2022, doi: 10.1016/j.asoc.2022.109826.
[24] A. M. Deshmukh, "Comparison of hidden markov model and recurrent neural network in automatic speech recognition," *European Journal of Engineering and Technology Research*, vol. 5, no. 8, pp. 958–965, Aug. 2020, doi: 10.24018/ejeng.2020.5.8.2077.
[25] C. V Priya, "Behavioral biometrics based authentication system using MLP-NN and MVPA," in *2021 IEEE International Power and Renewable Energy Conference (IPRECON)*, Sep. 2021, pp. 1–6, doi: 10.1109/IPRECON52453.2021.9640941.
[26] A. Ouisaadane, S. Safi, and M. Frikel, "Arabic digits speech recognition and speaker identification in noisy environment using a hybrid model of VQ and GMM," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 4, pp. 2193–2204, Aug. 2020, doi: 10.12928/telkomnika.v18i4.14215.
[27] M. Micucci and A. Iula, "Recognition performance analysis of a multimodal biometric system based on the fusion of 3D ultrasound hand-geometry and palmprint," *Sensors*, vol. 23, no. 7, Mar. 2023, doi: 10.3390/s23073653.
[28] A. Kumar, S. Jain, and M. Kumar, "Face and gait biometrics authentication system based on simplified deep neural networks," *International Journal of Information Technology*, vol. 15, no. 2, pp. 1005–1014, Feb. 2023, doi: 10.1007/s41870-022-01087-5.
[29] K. K. Singh and S. Barde, "A feasible adaptive fuzzy genetic technique for face, fingerprint, and palmprint based multimodal biometrics systems," *Journal of Current Science and Technology*, vol. 14, no. 1, Dec. 2023, doi: 10.59796/jcst.V14N1.2024.1.
[30] B. Ammour, L. Boubchir, T. Bouden, and M. Ramdani, "Face–iris multimodal biometric identification system," *Electronics*, vol. 9, no. 1, Jan. 2020, doi: 10.3390/electronics9010085.
[31] B. Alharbi and H. S. Alshanbari, "Face-voice based multimodal biometric authentication system via FaceNet and GMM," *PeerJ Computer Science*, vol. 9, Jul. 2023, doi: 10.7717/peerj-cs.1468.
[32] S. Riaz, U. Park, and P. Natarajan, "Improving face verification using facial marks and deep CNN: IARPA Janus benchmark-A," *Image and Vision Computing*, vol. 104, Dec. 2020, doi: 10.1016/j.imavis.2020.104020.
[33] K. Shinde and C. Kayte, "Multimodal deep learning based score Level fusion using face and fingerprint," in *Proceedings of the First International Conference on Advances in Computer Vision and Artificial Intelligence Technologies (ACVAIT 2022)*, 2023, pp. 140–152, doi: 10.2991/978-94-6463-196-8_13.
[34] A. Farhadipour, M. Chapariniya, T. Vukovic, and V. Dellwo, "Comparative analysis of modality fusion approaches for audio-visual person identification and verification," *Prepr. arXiv.2409.00562*, Nov. 2024.
[35] M. de Carvalho, "Mean, what do you mean?," *The American Statistician*, vol. 70, no. 3, pp. 270–274, Jul. 2016, doi: 10.1080/00031305.2016.1148632.
[36] A. Rattani, R. Derakhshani, and A. Ross, "Introduction to selfie biometrics," in *Selfie Biometrics*, 2019, pp. 1–18, doi: 10.1007/978-3-030-26972-2_1.
[37] J. M. Stoyanov, "Kolmogorov, stochastics in Bulgaria, and probabilistic problems with unexpected solutions," *Mathematics and Education in Mathematics*, vol. 53, pp. 180–197, Mar. 2024, doi: 10.55630/mem.2024.53.180-197.
[38] A. Ashraf, K. Ullah, A. Hussain, and M. Bari, "Interval-valued picture fuzzy maclaurin symmetric mean operator with application in multiple attribute decision-making," *Reports in Mechanical Engineering*, vol. 3, no. 1, pp. 301–317, Dec. 2022, doi: 10.31181/rme20020042022a.
[39] K. Cabello-Solorzano, I. Ortigosa de Araujo, M. Peña, L. Correia, and A. J. Tallón-Ballesteros, "The impact of data normalization on the accuracy of machine learning algorithms: A comparative analysis," 2023, pp. 344–353, doi: 10.1007/978-3-031-42536-3_33.
[40] J. Chen, H. Hu, H. Wu, Y. Jiang, and C. Wang, "Learning the best pooling strategy for visual semantic embedding," *Prepr. arXiv.2011.04305*, Jul. 2021.
[41] S. Shao, K. Chen, A. Karpur, Q. Cui, A. Araujo, and B. Cao, "Global features are all you need for image retrieval and reranking," *Prepr. arXiv.2308.06954*, Aug. 2023.
[42] J. Yao, Y. Li, B. Yang, and C. Wang, "Learning global image representation with generalized-mean pooling and smoothed average precision for large-scale CBIR," *IET Image Processing*, vol. 17, no. 9, pp. 2748–2763, Jul. 2023, doi: 10.1049/ipr2.12825.
[43] "NIST biometric scores set (BSSR1)," *NIST*, 2017. https://www.nist.gov/itl/iad/btg/nist-biometric-scores-set-bssr1 (accessed Oct. 24, 2025).
[44] M. Cheniti, N. Boukezzoula, and Z. Akhtar, "Symmetric sum-based biometric score fusion," *IET Biometrics*, vol. 7, no. 5, pp. 391–395, Sep. 2018, doi: 10.1049/iet-bmt.2017.0015.
[45] N. Poh and S. Bengio, "A score-level fusion benchmark database for biometric authentication," in *Audio- and Video-Based Biometric Person Authentication (AVBPA 2005)*, 2005, pp. 1059–1070, doi: 10.1007/11527923_110.
[46] B. Koch and R. Grbić, "One-shot lip-based biometric authentication: Extending behavioral features with authentication phrase information," *Image and Vision Computing*, vol. 142, Feb. 2024, doi: 10.1016/j.imavis.2024.104900.
[47] N. Poh and S. Bengio, "Database, protocols and tools for evaluating score-level fusion algorithms in biometric authentication," *Pattern Recognition*, vol. 39, no. 2, pp. 223–233, Feb. 2006, doi: 10.1016/j.patcog.2005.06.011.
[48] Q. An, S. Rahman, J. Zhou, and J. J. Kang, "A comprehensive review on machine learning in healthcare industry: Classification, restrictions, opportunities and challenges," *Sensors*, vol. 23, no. 9, Apr. 2023, doi: 10.3390/s23094178.
[49] A. H. A. Abdel-aziem and T. H. M. Soliman, "A multi-Layer perceptron (MLP) neural networks for stellar classification: A review of methods and results," *International Journal of Advances in Applied Computational Intelligence*, vol. 3, no. 2, pp. 29–37, 2023, doi: 10.54216/IJAACI.030203.
[50] A. Naseer and A. Jalal, "Multimodal objects categorization by fusing GMM and multi-layer perceptron," in *2024 5th International Conference on Advancements in Computational Sciences (ICACS)*, Feb. 2024, pp. 1–7, doi: 10.1109/ICACS60934.2024.10473242.
[51] S. Pahuja and N. Goel, "Multimodal biometric authentication: A review," *AI Communications*, vol. 37, no. 4, pp. 525–547, Sep. 2024, doi: 10.3233/AIC-220247.
[52] F. Z. Allam1, L. Hamami-Mitiche, and H. Bousbia-Salah, "Evaluation and comparison of the performance of biometric recognition," *International Journal of Industrial Engineering & Production Research*, vol. 33, no. 1, pp. 1–12, 2022.
[53] M. Abdul-Al, G. Kumi Kyeremeh, R. Qahwaji, N. T. Ali, and R. A. Abd-Alhameed, "The evolution of biometric authentication: A deep dive into multi-modal facial recognition: A review case study," *IEEE Access*, vol. 12, pp. 179010–179038, 2024, doi: 10.1109/ACCESS.2024.3486552.
[54] R. Sánchez-Marqués, V. García, and J. S. Sánchez, "A data-centric machine learning approach to improve prediction of glioma grades using low-imbalance TCGA data," *Scientific Reports*, vol. 14, no. 1, Jul. 2024, doi: 10.1038/s41598-024-68291-0.
[55] O. G. Atanda, W. Ismaila, A. O. Afolabi, O. A. Awodoye, A. S. Falohun, and J. P. Oguntoye, "Statistical analysis of a deep learning based trimodal biometric system using paired sampling t-Test," in *2023 International Conference on Science, Engineering and Business*

for Sustainable Development Goals (SEB-SDG), Apr. 2023, pp. 1–10, doi: 10.1109/SEB-SDG57117.2023.10124624.
[56] M. Al-kassab, "The use of one sample t-Test in the real data," *Journal of Advances in Mathematics*, vol. 21, pp. 134–138, Sep. 2022, doi: 10.24297/jam.v21i.9279.
[57] M. Hanmandlu, J. Grover, V. K. Madasu, and S. Vasirkala, "Score level fusion of hand based biometrics using t-norms," in *2010 IEEE International Conference on Technologies for Homeland Security (HST)*, Nov. 2010, pp. 70–76, doi: 10.1109/THS.2010.5655093.
[58] K. Nandakumar, Yi Chen, S. C. Dass, and A. K. Jain, "Likelihood ratio-based biometric score fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 342–347, Feb. 2008, doi: 10.1109/TPAMI.2007.70796.
[59] S. Moeen, M. Uddin, and J. Grover, "Score level fusion of multimodal biometrics based on entropy function," *International Journal of Computer Applications*, vol. 142, no. 9, pp. 28–33, May 2016, doi: 10.5120/ijca2016909912.
[60] H. Abderrahmane, G. Noubeil, Z. Lahcene, Z. Akhtar, and D. Dasgupta, "Weighted quasi-arithmetic mean based score level fusion for multi-biometric systems," *IET Biometrics*, vol. 9, no. 3, pp. 91–99, May 2020, doi: 10.1049/iet-bmt.2018.5265.
[61] N. Singhal, V. Ganganwar, M. Yadav, A. Chauhan, M. Jakhar, and K. Sharma, "Comparative study of machine learning and deep learning algorithm for face recognition," *Jordanian Journal of Computers and Information Technology*, vol. 7, no. 3, pp. 313–325, 2021, doi: 10.5455/JJCIT.71-1624859356.
[62] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, and D. Zhang, "Biometrics recognition using deep learning: a survey," *Artificial Intelligence Review*, vol. 56, no. 8, pp. 8647–8695, Aug. 2023, doi: 10.1007/s10462-022-10237-x.

BIOGRAPHIES OF AUTHORS

Leila Hellal received the B.Eng. degree in Electronics from the Department of Electronics, Faculty of Technology, Ferhat Abbas University Setif-1, Setif, Algeria, in 2015, and the M.Sc. degree in Telecommunication Networks and Systems from the same university in 2017. She is currently pursuing the Ph.D. degree in Electronics and Telecommunications at Ferhat Abbas University Setif-1, Setif, Algeria. Her research focuses on multimodal biometric systems, with specific interests in fingerprint and face recognition, pattern recognition, image processing, identification, and authentication using score-level fusion techniques. She has participated in several international conferences, including ICESTI'19, 2ETA'23, and ISPA 2024. She can be contacted at email: leila.hellal@univ-setif.dz.

Naceur-Eddine Boukezzoula received the Engineering degree in 1980 from University of Oran USTO (Algeria) in Electronics in 1980, then a Magister degree in Communication from Setif University (Algeria) in 1989, and a Ph.D. degree from Setif University in 2006. He teachers in university of setif at Electronics Department since 1980. He is a member of the scientific institution of the Faculty of Technology. His principal researches are about neural networks and image processing (watermarking and biometrics). He can be contacted at email: nasrbou@yahoo.fr.

Mohamed Cheniti received his Bachelor's in Electronics in 2013 and Master's in Embedded Systems in 2015 from the Ferhat Abbas Setif-1 University, Setif, Algeria. He received the Ph.D. degree in Electronic and Computer Engineering from Ferhat Abbas Setif-1 University, Setif, Algeria. He is currently a Research Assistant Professor at the Institute of Optics and Precision Mechanics, Ferhat Abbas University Setif-1, Setif, Algeria. His research interests include pattern recognition, biometrics and security systems, machine learning, image and video processing, and fuzzy logic systems. He can be contacted at email: cheniti1988@gmail.com.

Zahid Akhtar received the Ph.D. degree in Electronic and Computer Engineering from the University of Cagliari, Italy. He is currently an assistant professor with the Department of Network and Computer Security, State University of New York (SUNY) Polytechnic Institute, USA. Prior to that, he was a research assistant professor with the University of Memphis, USA, and a Postdoctoral Fellow with the INRSEMT, University of Quebec, Canada, the University of Udine, Italy, Bahcesehir University, Turkey, and the University of Cagliari. He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at the State University of New York Polytechnic Institute, Utica, United States. His research interests include the computer vision and machine learning with applications to cybersecurity, biometrics, affect recognition, image and video processing, and audiovisual multimedia quality assessment. He can be contacted at email: akhtarz@sunypoly.edu.