JIRBDAI

Journal of Institutional Research, Big Data Analytics and Innovation

Effect of Validation-Guided Weighted-Sum PCA-LDA Feature Fusion on ABC-Optimized CNN Performance for Multimodal Biometric Identification

[1*]Akande, O. V, [2]Afolabi A. O., [3]Adeosun O. O.,[4] Salahudeen S. A., [5]Oyewale S. O., [6]Araba T. A.

[1,3]Department of Computer Science, Faculty of Computing and Informatics, Ladoke Akintola University of Technology, Ogbomoso, Nigeria

[2] Department of Cyber Security, Ladoke Akintola University of Technology, Ogbomoso, Nigeria

[4,5,6] Federal Polytechnic Ayede, Oyo State

Corresponding Author: akandeov@federalpolyayede.edu.ng

ABSTRACT

Feature-level fusion can combine complementary biometric information without increasing classifier complexity, but the benefit of a weighted fusion rule depends on how its component representations are normalized, weighted, and evaluated. This study investigated validation-guided weighted-sum Principal Component Analysis (PCA)-Linear Discriminant Analysis (LDA) fusion for face-palm-iris multimodal biometric identification. The MULB dataset provided 174 matched identities, each with 20 face, 20 right-hand/palm and 20 right-iris images (10,440 images in total). Source images were partitioned before constructing identity-consistent triplets, yielding 2,227 training, 557 validation, and 696 independent test triplets. PCA, LDA, and separate standardization transforms were fitted on training data only; both feature representations were restricted to 150 dimensions. Nine PCA:LDA weight pairs from 0.1:0.9 to 0.9:0.1 were evaluated on validation data. The selected 0.1:0.9 weighting achieved 96.04 ± 0.62% validation accuracy and 95.91 ± 0.63% macro F1, compared with 94.33 ± 0.77% and 94.13 ± 0.80% for equal weighting. With the final ABC-optimized CNN frozen, ten matched-seed runs produced independent-test accuracies of 97.19 ± 0.39% for PCA, 97.78 ± 0.29% for LDA, and 97.66 ± 0.23% for the selected fusion; corresponding macro F1 scores were 97.12 ± 0.42%, 97.71 ± 0.30%, and 97.61 ± 0.23%. Friedman tests found no significant overall difference in test accuracy (χ^2=5.389, p=0.0676) or macro F1 (χ^2=3.800, p=0.1496). The selected weighted representation therefore improved validation performance over equal weighting and PCA, while final testing showed competitive, low-variance performance rather than statistically significant superiority over LDA.

KEYWORDS: ABC Optimization, Convolutional Neural Network, Face-Palm-Iris Recognition, Feature-Level Fusion, Linear Discriminant Analysis, Multimodal Biometrics, Principal Component Analysis, Weighted-Sum Fusion

I. INTRODUCTION

Multimodal biometric recognition combines information from more than one biometric source to improve recognition reliability beyond what may be obtained from a single trait. The effectiveness of a multimodal system, however, depends not only on the availability of several biometric sources but also on how their information is represented and fused. Feature-level fusion is particularly attractive because integration occurs before classification, allowing the classifier to operate on a unified representation; at the same time, differences in feature scale, redundancy, dimensionality, and discriminative strength can reduce the benefit of fusion when they are not controlled [1]-[4]. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) provide two established but statistically different approaches to constructing compact biometric representations. PCA preserves dominant directions of variance without using class labels, whereas LDA uses class information to emphasize separation among identities. Their different objectives make them potentially complementary, but they also create a fusion problem because corresponding PCA and LDA components should not automatically

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.23415861

be assumed to encode identical information. Recent feature-fusion studies therefore emphasize dimensional compatibility, normalization, redundancy control, and the relative contribution assigned to fused sources [5]-[9].

Weighted-sum fusion combines equal-dimensional normalized feature vectors without increasing the input size. Rather than concatenating PCA and LDA into a larger vector, their outputs are combined element-wise using explicit contribution weights. Equal weighting assumes comparable contributions from both representations, whereas unequal weighting allows the stronger representation to contribute more heavily. Recent weighted and adaptive fusion studies show that recognition performance can vary materially with the relative contribution assigned to each constituent source [7], [8], [10], [11]. When differences between fusion schemes are small, experimental control becomes critical. Fitting normalization, PCA or supervised LDA before partitioning can allow information from validation or test observations to influence the learned feature space and inflate performance estimates [15]. A single neural-network run can also reflect random initialization rather than a stable representation effect. In this study, the fusion weights are stated explicitly, equal weighting is retained as a reference, all data-dependent transformations are fitted on training data only, and the competing representations are evaluated under the same classifier settings. Accordingly, this study evaluates validation-guided weighted-sum PCA-LDA fusion for face-palm-iris multimodal biometric identification. PCA and LDA are each constrained to 150 dimensions and standardized independently before fusion. Nine complementary weight pairs are compared using validation data, with 0.5:0.5 retained as an equal-weight baseline. The selected weighting is then frozen and evaluated against PCA-only and LDA-only representations using one ABC-optimized CNN architecture and ten matched random seeds. The study therefore examines both the development-stage benefit of unequal weighting and the extent to which that benefit persists under independent testing.

II. RELATED WORK

Recent multimodal biometric research has explored fusion at feature, score, and decision levels. Wang et al. [3] combined face and finger-vein information using CNN-based feature fusion, while Alshardan et al. [4] compared early, late, and score-level fusion with fingerprint and finger-vein traits. Safavipour et al. [5] reported a deep feature-level framework combining five biometric traits, and El_Rahman and Alluhaidan [14] compared traditional and deep multimodal approaches. Collectively, these studies support the value of multimodal integration but also show that the benefit depends strongly on the representation and fusion mechanism rather than on multimodality alone. Feature-level fusion is attractive because it retains information before the final classifier compresses it into a decision. Its main challenge is that fused descriptors may be dependent, redundant, or differently scaled. Li et al. [2] identify feature-level fusion, hybrid fusion, and deep fusion as major directions in hand-based multimodal recognition, while Artabaz and Sliman [6] show that feature dependency and selection can materially affect performance. Kamlaskar and Abhyankar [9] used canonical-correlation-based feature fusion to exploit relationships across modalities, illustrating the broader need to make the fusion space structurally coherent rather than merely larger.

Weighted fusion provides another route to controlling information contribution. Atenco et al. [7] used a weighted sum for deep audiovisual biometric fusion, directly optimizing the contribution of the component modalities. Ipeayeda et al. [8] likewise treated feature-fusion optimization as a search problem and reported gains when complementary information was selected more effectively. More recent adaptive approaches assign weights according to trust or context: Chitrapu et al. [10] proposed trust-adaptive multimodal fusion, while Ayeswarya and Singh [11] varied biometric contribution according to contextual conditions. These studies motivate explicit rather than implicit treatment of fusion weights. For the MULB family of data, Kadhim and Abdulameer [12] introduced a multimodal database containing face, hand, and iris observations, and later demonstrated high recognition performance from multimodal face-palm-iris fusion using deep feature extraction and support vector classification [13]. Those studies differ substantially in preprocessing, feature extraction,

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.23415861

fusion and classification, so their reported performance is not a direct benchmark for the present experiment. The narrower question addressed here is whether a compact, normalized weighted combination of PCA and LDA improves recognition when all three feature alternatives are compared under one fixed classifier and leakage-controlled evaluation protocol. The unresolved question is whether a compact weighted PCA-LDA representation retains an advantage when its coefficients are selected independently of the test set and the competing representations are evaluated under the same classifier. The present study addresses that question through explicit weight selection, an equal-weight reference, leakage-controlled preprocessing, independent testing and repeated matched-seed runs.

III. MATERIALS AND METHODS

A. Dataset Audit and Face-Palm-Iris Triplet Construction

The experiment used the full MULBv1 database [12]. MULBv1 contains homologous biometric data from 174 individuals, with 20 face images, 20 right-hand/palm images and 20 right-iris images for each person. The experiment therefore used 3,480 labelled face images, 3,480 labelled palm images and 3,480 labelled iris images, giving 10,440 biometric images representing the same 174 identities across all three modalities. Because the available files did not contain synchronized session metadata linking particular face, palm and iris acquisitions, the triplet-construction procedure was deliberately conservative. Face, palm and iris source images were first assigned to train, validation or test partitions within each identity. Triplets were then created only within the same identity and partition using deterministic one-to-one matching across the three modalities, limited by the smallest modality count. No Cartesian-product expansion was used. The resulting triplets are therefore identity-consistent rather than claimed to be session-synchronized.

Table 1. Data Partition and Triplet Summary

Partition	Face	Palm	Iris	Triplets
Training	2,227	2,227	2,227	2,227
Validation	557	557	557	557
Test	696	696	696	696

B. Data Partitioning and Leakage Control

The split followed an 80:20 development-test design, after which the development portion was divided 80:20 into training and validation data. At source-image level this produced approximately 64% training, 16% validation and 20% test data overall. The split was completed before face-palm-iris triplets, PCA, LDA or scaling were created. Each raw image therefore belonged to one partition only, and no image used to construct a test triplet appeared in training or validation. All data-dependent transforms were subsequently fitted using the 2,227 training triplets only. PCA, LDA and both feature standardizers were estimated from training data and then applied unchanged to validation and test data. This order prevents the validation or test distributions from influencing the learned feature space, consistent with recommendations for avoiding preprocessing leakage in machine-learning studies [15]. Ten independent integrity checks confirmed partition uniqueness, same-identity triplet construction, training-only fitting and the absence of test labels from all feature-fitting operations.

C. Image Preprocessing and Common Trimodal Observation

Each image was decoded in grayscale, resized to 128 × 128 pixels using area interpolation and scaled to the interval [0,1]. The face, palm and iris arrays were flattened to 16,384 values each. For each valid triplet, the three modality vectors were concatenated only to create a common 49,152-dimensional trimodal observation. This operation is distinct from the later PCA-LDA fusion: it simply represents the three biometric modalities in one observation before statistical feature extraction.

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.23415861

$$x = [face_1, \ldots, face_{16384}, palm_1, \ldots, palm_{16384}, iris_1, \ldots, iris_{16384}] \in R^{49152}. \quad (1)$$

D. PCA and LDA Feature Construction

PCA was fitted to the training trimodal observations and restricted to 150 components. The retained components explained approximately 85.07% of the training-set variance. The fitted PCA transform was then applied to validation and test observations. LDA was fitted only on the 150-dimensional training PCA representation and training identity labels. With 174 classes, up to 173 discriminant directions were theoretically available; 150 components were retained so that PCA and LDA would have identical dimensionality. The fitted LDA transform was subsequently applied to the validation and test PCA features. PCA and LDA outputs were standardized separately using StandardScaler instances fitted on the corresponding training representations. Separate normalization was necessary because weighted addition is sensitive to scale: without it, the representation with the larger numerical dispersion could dominate the fused vector regardless of the assigned coefficient. The standardized PCA and LDA matrices therefore formed comparable 150-dimensional inputs to the fusion rule.

E. Weighted-Sum Fusion and Validation-Based Weight Selection

The proposed fusion retained a fixed 150-dimensional representation and used element-wise weighted addition rather than concatenation. For a normalized PCA feature vector F_PCA and normalized LDA vector F_LDA, the fused representation was defined as

$$F_WS = w_PCA\ F_PCA + w_LDA\ F_LDA, \quad w_PCA + w_LDA = 1. \quad (2)$$

Equally indexed PCA and LDA components were not treated as semantically identical. Both vectors are equal-dimensional transformations of the same trimodal observation, independently standardized from training data and combined as an empirical linear representation. Their relative contribution was therefore determined on held-out validation data, with 0.5:0.5 retained as the neutral equal-weight reference. Nine weight pairs were evaluated: 0.1:0.9, 0.2:0.8, …, 0.9:0.1. A fixed reference Conv1D classifier was retrained from scratch for seeds 101, 202 and 303 for every candidate. The test set was not loaded into the model-selection object and no test metric entered the ranking. Mean validation accuracy was the primary selection criterion; ties were resolved by higher macro F1 and then lower accuracy standard deviation. The 0.5:0.5 candidate served as the equal-weight reference for comparison with validation-guided weighting.

F. ABC-Optimized CNN and Final Controlled Comparison

After weight selection, the CNN hyperparameters were optimized on the selected fused representation using Artificial Bee Colony search. ABC follows employed-bee, onlooker-bee and scout-bee phases to balance exploitation and exploration [17], while systematic hyperparameter search reduces dependence on ad hoc manual tuning [16], [18]. The search used training and validation data only; the final architecture was frozen before independent test evaluation. The same frozen architecture was then used without modification for PCA, LDA and selected weighted fusion, so the final comparison changed the feature representation while holding the classifier constant.

For every final run, early stopping monitored validation loss with patience 5 and restored the best validation-loss weights. ReduceLROnPlateau also monitored validation loss, using factor 0.5, patience 3 and minimum learning rate 10^{-6}. The 150-dimensional vector was reshaped explicitly as (150,1) and processed by Conv1D, avoiding the ambiguous two-dimensional convolutional interpretation of a one-dimensional feature vector.

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.23415861

Table 2: CNN and Training Configuration

Component	Setting
Input	150 × 1 sequence
Conv1D-1	128 filters, kernel 5, ReLU, same padding
Max pooling	Pool size 2
Conv1D-2	128 filters, kernel 5, ReLU, same padding
Dense	512 units, ReLU
Dropout	0.70
Output	174-way softmax
Optimizer	Adam, learning rate 0.0005
Loss	Sparse categorical cross-entropy
Batch / max epochs	32 / 30
Parameters	2,604,604

G. Repeated Evaluation and Statistical Analysis

Final evaluation used ten matched random seeds: 101, 202, 303, 404, 505, 606, 707, 808, 909 and 1010. For each seed, fresh PCA, LDA and selected-fusion CNN models were initialized and trained under the identical protocol. Matching the seed across representations produced paired observations of training stochasticity. After training and restoration of the best validation-loss weights, each model was evaluated once on the 696 fixed test triplets. No hyperparameter or fusion-weight decision was changed after test performance became available. Reported metrics include accuracy, macro precision, macro recall, macro F1, weighted F1, macro one-vs-rest false-positive rate (FPR), macro one-vs-rest false-negative rate (FNR), and classifier-level inference time. For each class c, FPR_c=FP_c/(FP_c+TN_c) and FNR_c=FN_c/(FN_c+TP_c); values were macro-averaged across the 174 classes. These multiclass quantities are not presented as biometric FAR or FRR because no verification threshold was applied. Inference timing used three warm-up passes followed by 20 timed prediction passes and therefore describes classifier computation only, not image acquisition, preprocessing, feature transformation or sensor-to-decision latency. Accuracy and macro F1 were the primary statistical outcomes. A Friedman repeated-measures test compared PCA, LDA and weighted fusion across the ten matched seeds. The pre-specified procedure required $p<0.05$ before Wilcoxon-Holm post-hoc testing. Because the omnibus tests were not significant, post-hoc significance tests were not performed. Paired mean differences, 95% confidence intervals and effect sizes were retained descriptively to quantify the magnitude and direction of the observed differences. The repeated-seed analysis measures robustness to stochastic training on one fixed split; it is not equivalent to ten independently sampled datasets.

IV. RESULTS AND DISCUSSION

A. Validation-Guided Fusion Weight Selection

Table III summarizes the validation search. Performance was not monotonic with the PCA coefficient. The strongest result occurred at w_PCA=0.1 and w_LDA=0.9, where mean validation accuracy reached 96.04% and macro F1 reached 95.91%. Equal weighting produced 94.33% accuracy and 94.13% macro F1. Thus, assigning a larger contribution to LDA improved validation accuracy by 1.71 percentage points and macro F1 by 1.77 percentage points relative to the naïve 0.5:0.5 baseline. The selected configuration also showed lower variation than the equal-weight baseline across the three selection seeds.

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.23415861

Table 3: Validation Search For PCA: LDA Fusion Weights

PCA: LDA	Val. Acc. (%)	Macro F1 (%)	Val. Loss
0.1:0.9	96.04 ± 0.62	95.91 ± 0.63	0.1925
0.2:0.8	95.51 ± 0.37	95.35 ± 0.37	0.1977
0.3:0.7	95.39 ± 0.81	95.19 ± 0.87	0.2072
0.4:0.6	94.21 ± 0.89	94.07 ± 0.84	0.2494
0.5:0.5	94.33 ± 0.77	94.13 ± 0.80	0.2451
0.6:0.4	94.15 ± 0.18	93.92 ± 0.20	0.2654
0.7:0.3	93.91 ± 0.45	93.69 ± 0.46	0.2569
0.8:0.2	94.56 ± 0.57	94.41 ± 0.67	0.2374
0.9:0.1	94.68 ± 0.94	94.51 ± 1.03	0.2307

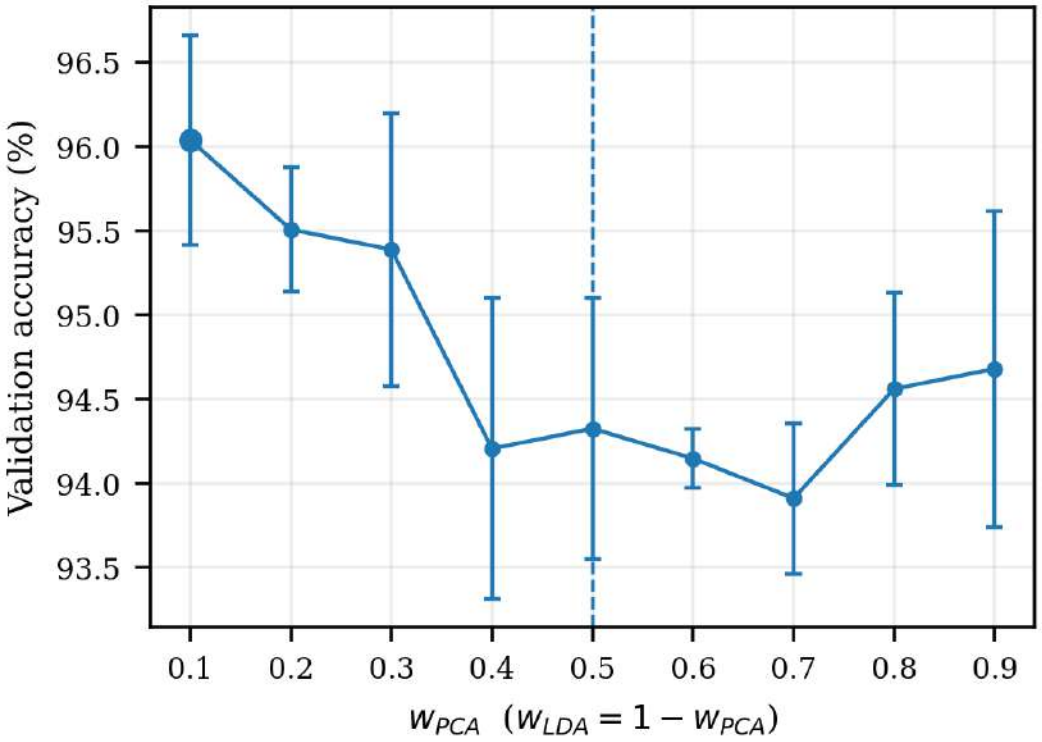


Fig. 1. Validation Accuracy across Nine PCA: LDA Weight Pairs
Error bars show sample SD across three seeds; the dashed line marks equal weighting.

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.23415861

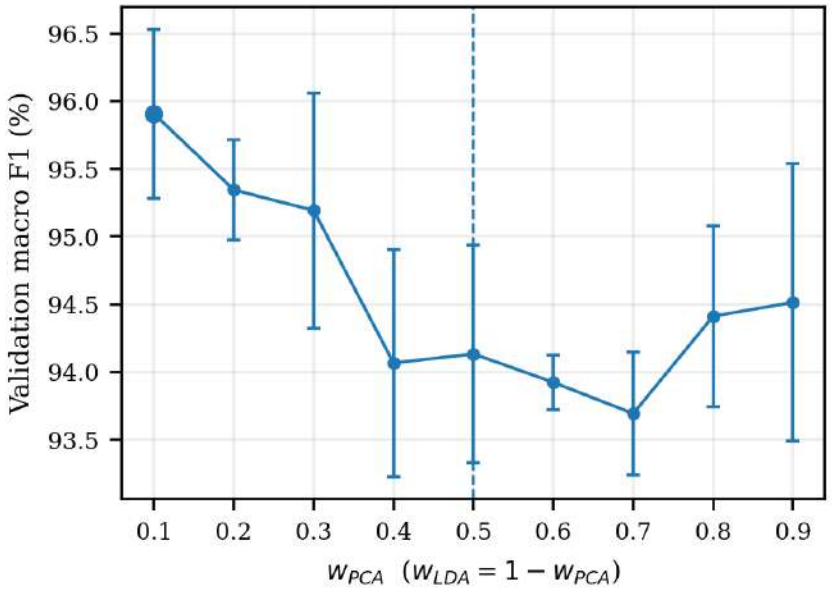


Fig. 2. Validation Macro F1 across Nine PCA:LDA Weight Pairs.
Error bars show sample SD across three seeds; the dashed line marks equal weighting.

B. Training and Validation Behaviour Under the Frozen CNN

Figures 3 and 4 show the mean training and validation trajectories across the ten final runs. All three representations converged rapidly. LDA and the selected fusion separated from PCA during the early-to-middle epochs, while the selected fusion maintained the strongest mean validation trajectory through much of training.

Table IV reports the minimum-validation-loss checkpoint from each run. At those restored checkpoints, the selected fusion achieved the highest mean validation accuracy (97.48 ± 0.35%) and the lowest mean validation loss (0.1092 ± 0.0128). LDA reached 97.25 ± 0.39% and 0.1138 ± 0.0097, while PCA reached 96.13 ± 0.58% and 0.1693 ± 0.0230. The selected fusion therefore led the development-stage validation results under the frozen CNN.

Table 4: Best-Validation-Loss Checkpoints across Ten Final Runs

Representation	Val. acc. (%)	Val. Loss	Epoch
PCA	96.13 ± 0.58	0.1693 ± 0.0230	25.3 ± 3.1
LDA	97.25 ± 0.39	0.1138 ± 0.0097	25.9 ± 4.1
0.1 PCA + 0.9 LDA	97.48 ± 0.35	0.1092 ± 0.0128	27.2 ± 3.2

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.23415861

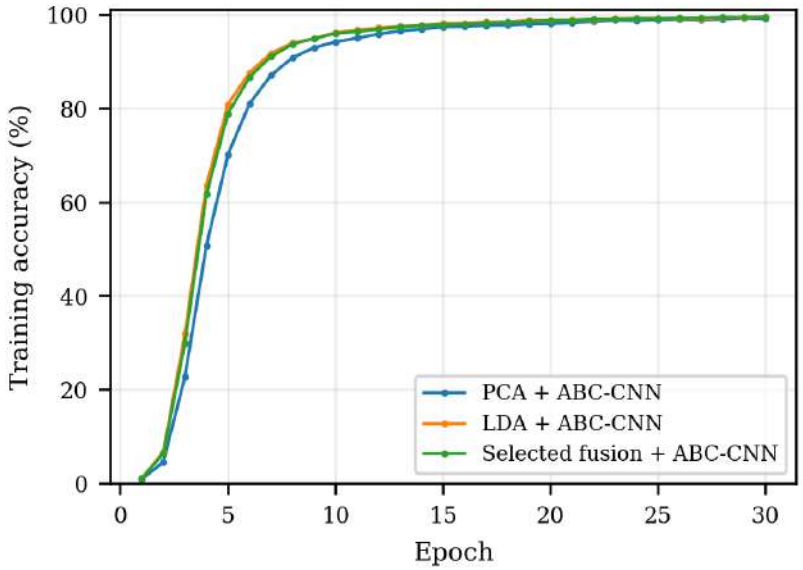


Fig. 3. Mean Training Accuracy across Ten Matched Seeds under the Frozen ABC-CNN Architecture

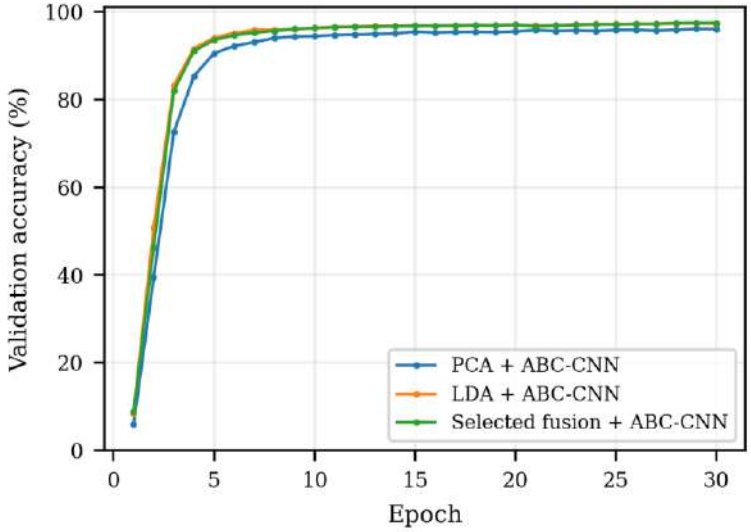


Fig. 4. Mean Validation Accuracy across Ten Matched Seeds under the Frozen ABC-CNN Architecture

C. Independent Test Performance

The independent-test results are summarized in Table V and Fig. 5. LDA obtained the highest mean accuracy at 97.78 ± 0.29%, followed closely by selected weighted fusion at 97.66 ± 0.23% and PCA at 97.19 ± 0.39%. The same ordering occurred for macro F1: 97.71% for LDA, 97.61% for fusion and 97.12% for PCA. Relative to PCA, the fused representation improved mean test accuracy by 0.47 percentage points and macro F1 by 0.49 percentage points. Relative to LDA, however, fusion was lower by 0.12 percentage points in accuracy and 0.09 percentage points in macro F1. Although the weighted representation did not achieve the highest test mean, it exhibited the lowest run-to-run variation in both primary metrics. Its test-accuracy SD was 0.23 percentage points, compared with 0.29 for LDA and 0.39 for PCA; its macro-F1 SD was likewise 0.23 percentage points. Error-rate results followed the same general pattern. Fusion reduced macro-FNR from 2.81% for PCA to 2.34%, while LDA remained lowest at 2.22%. Macro FPR values were small for all three methods and should be interpreted as multiclass one-vs-rest rates rather than operational false-accept probabilities.

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.23415861

Table 5: Independent Test Performance across Ten Matched Seeds

Representation	Acc. (%)	Macro P (%)	Macro R (%)	Macro F1 (%)
PCA	97.19 ± 0.39	97.77 ± 0.33	97.19 ± 0.39	97.12 ± 0.42
LDA	97.78 ± 0.29	98.25 ± 0.23	97.78 ± 0.29	97.71 ± 0.30
0.1 PCA + 0.9 LDA	97.66 ± 0.23	98.16 ± 0.19	97.66 ± 0.23	97.61 ± 0.23

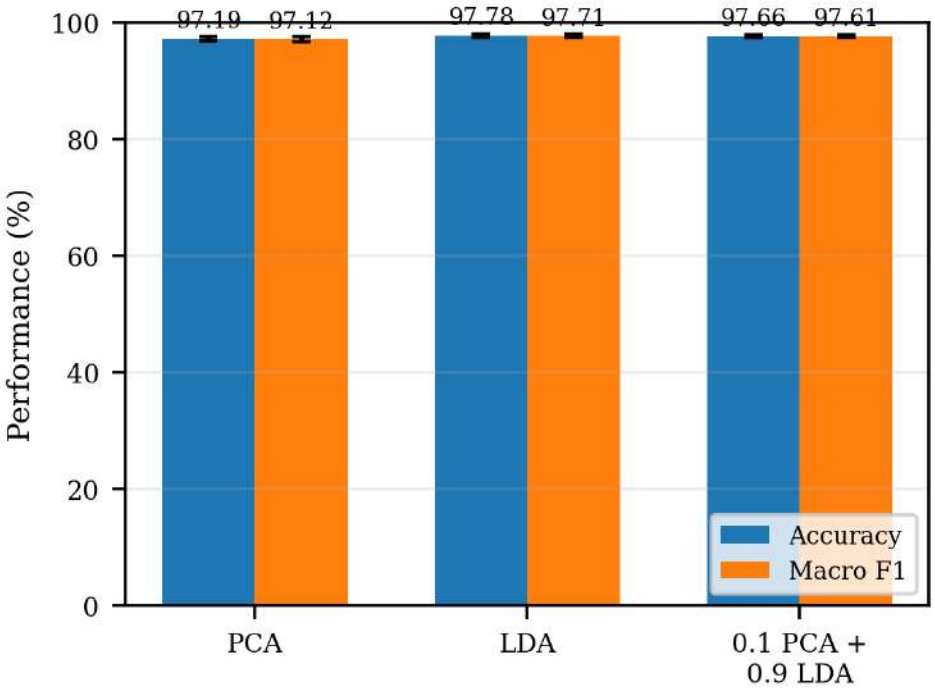


Fig. 5. Mean Independent-test Accuracy and Macro F1 across Ten Matched Runs
Error bars show sample SD.

D. Statistical Robustness of the Test Differences

The Friedman test did not reject equality of the three representations for either primary test outcome. For accuracy, χ^2=5.389 with p=0.0676; for macro F1, χ^2=3.800 with p=0.1496. Under the pre-specified analysis plan, Wilcoxon-Holm post-hoc significance testing was therefore not performed. The numerical advantage of fusion over PCA did not translate into a significant overall three-method difference at α=0.05. The paired effect estimates show the direction and magnitude of the differences. Fusion exceeded PCA by 0.465 percentage points in accuracy (95% CI: 0.098 to 0.833) and 0.490 percentage points in macro F1 (95% CI: 0.108 to 0.871), with paired Cohen dz values of 0.91 and 0.92 respectively. Against LDA, fusion was lower by 0.120 percentage points in accuracy (95% CI: -0.350 to 0.111) and 0.093 percentage points in macro F1 (95% CI: -0.343 to 0.156). The weighted representation improved upon PCA and remained close to LDA, but the final test evidence did not establish overall statistical superiority.

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.23415861

Table 6: Primary Statistical Results and Paired Effects

Outcome / comparison	Statistic or difference	Effect
Accuracy: Friedman	χ^2=5.389, p=0.0676	Not significant
Macro F1: Friedman	χ^2=3.800, p=0.1496	Not significant
Fusion − PCA accuracy	+0.465 pp	dz=0.905; r_rb=0.727
Fusion − LDA accuracy	−0.120 pp	dz=−0.371; r_rb=−0.571
Fusion − PCA macro F1	+0.490 pp	dz=0.918; r_rb=0.709
Fusion − LDA macro F1	−0.093 pp	dz=−0.268; r_rb=−0.273

E. Computational Behaviour

Classifier-level inference time was essentially unchanged by the feature representation because PCA, LDA and weighted fusion all entered the network as 150 × 1 tensors. Mean prediction time was 0.366 ± 0.011 ms/sample for PCA, 0.360 ± 0.007 ms/sample for LDA and 0.365 ± 0.012 ms/sample for selected fusion. Weighted addition therefore did not impose a meaningful classifier-side timing penalty. These measurements deliberately exclude image capture, preprocessing, PCA/LDA transformation and feature fusion, and should not be interpreted as complete end-to-end biometric-system latency.

DISCUSSION

Validation-guided weighting improved the fusion rule relative to equal weighting. The 0.1:0.9 combination increased validation accuracy by 1.71 percentage points and macro F1 by 1.77 percentage points over the 0.5:0.5 baseline, and it ranked first among the nine candidates. The dominant LDA coefficient is consistent with the stronger discriminative behaviour of LDA observed in the later comparisons. Across the ten final runs, the selected fusion also reached the highest mean validation accuracy and the lowest validation loss at the restored checkpoints. A small PCA contribution may retain variance information not emphasized by LDA, while the dominant LDA term preserves class separation. This pattern is consistent with weighted and adaptive fusion methods that assign unequal contributions to sources with different discriminative strength [7], [8], [10], [11].

On the independent test set, fusion retained a numerical advantage over PCA, while LDA achieved slightly higher mean accuracy and macro F1. The LDA-fusion difference was about one tenth of a percentage point, and fusion showed lower variability across seeds. The Friedman tests were not significant; thus, the test results support competitive performance rather than statistical superiority of fusion over LDA. The repeated matched-seed evaluation showed that the differences among the three high-performing representations were small relative to stochastic training variation. A single initialization could therefore change their apparent ordering. This pattern is consistent with broader concerns about leakage and reproducibility when machine-learning conclusions are based on narrow performance margins [15]. The weighted-sum formulation remained 150-dimensional, avoided the feature expansion associated with concatenation, and added no meaningful classifier-side inference cost. Its coefficients were explicit and selected from validation performance rather than fixed arbitrarily. The selected fusion also produced the lowest test variability in accuracy and macro F1 among the three representations. Deep, hybrid and adaptive fusion methods can model richer interactions and have reported high recognition rates [3]-[6], [8], [10], [11], [13], [14]. The weighted-sum approach trades some of that flexibility for interpretability and dimensional economy: the contribution of each representation is explicit and the fused vector remains the same size as its components. Its limitation is that linear element-wise addition cannot learn complex cross-component interactions.

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.23415861

LIMITATIONS AND FUTURE WORK

The experiment used one audited release of MULB and one fixed source-image split. The ten repeated runs vary neural-network initialization and stochastic training but do not measure uncertainty across independent population samples or alternative data splits. Future studies can combine an untouched final test set with repeated or nested cross-validation inside the development data, particularly when several fusion families are compared. The present holdout design kept the final test set isolated until all feature and hyperparameter decisions were fixed. The available face, palm and iris files shared identity labels but did not contain synchronized session metadata, so the constructed observations are identity-consistent triplets rather than claims of simultaneous acquisition. The ABC architecture was optimized using the selected fused representation and then applied unchanged to PCA and LDA; this provides a controlled common-classifier comparison but may favour the representation used during architecture search. Future work could repeat the comparison under multiple fixed architectures or nested representation-neutral model selection. Finally, this is a 174-class closed-set identification experiment. The reported macro one-vs-rest FPR and FNR should not be interpreted as conventional biometric FAR and FRR without an explicit verification threshold. Future work should evaluate subject-independent verification protocols, threshold-dependent FAR/FRR and EER, additional datasets, synchronized multimodal acquisitions, learned or attention-based fusion, and end-to-end latency on deployment hardware.

V. CONCLUSION

This study evaluated validation-guided weighted-sum PCA-LDA feature fusion for face-palm-iris multimodal biometric identification under a leakage-controlled and repeated-run protocol. PCA and LDA were fitted from training data only, separately standardized, constrained to a common 150-dimensional space and combined by genuine element-wise weighted addition. Validation search selected 0.1 PCA + 0.9 LDA, improving validation accuracy from 94.33% under equal weighting to 96.04%. Under the final frozen ABC-CNN, the selected fusion also produced the strongest development-stage validation accuracy and lowest validation loss across ten runs. On the independent test set, selected fusion achieved 97.66 ± 0.23% accuracy and 97.61 ± 0.23% macro F1, improving upon PCA but remaining marginally below LDA. Friedman tests did not establish a significant overall difference among the three representations. Validation-guided weighting therefore provides a compact and stable alternative to equal weighting and PCA alone, but the present evidence does not support statistical superiority over LDA. The results distinguish the clear validation-stage gain from the more limited difference observed on the independent test set.

REFERENCES

[1] S. Pahuja and N. Goel, “Multimodal biometric authentication: A review,” AI Communications, vol. 37, no. 4, pp. 525–547, 2024, doi: 10.3233/AIC-220247.

[2] S. Li, L. Fei, B. Zhang, X. Ning, and L. Wu, “Hand-based multimodal biometric fusion: A review,” Information Fusion, vol. 109, Art. 102418, 2024, doi: 10.1016/j.inffus.2024.102418.

[3] Y. Wang, D. Shi, and W. Zhou, “Convolutional neural network approach based on multimodal biometric system with fusion of face and finger vein features,” Sensors, vol. 22, no. 16, Art. 6039, 2022, doi: 10.3390/s22166039.

[4] A. Alshardan et al., “Multimodal biometric identification: Leveraging convolutional neural network architectures and fusion techniques with fingerprint and finger vein data,” PeerJ Computer Science, vol. 10, e2440, 2024, doi: 10.7717/peerj-cs.2440.

[5] M. H. Safavipour, M. A. Doostari, and H. Sadjedi, “Deep hybrid multimodal biometric recognition system based on features-level deep fusion of five biometric traits,” Computational Intelligence and Neuroscience, vol. 2023, Art. 6443786, 2023, doi: 10.1155/2023/6443786.

[6] S. Artabaz and L. Sliman, "Feature fusion and selection using handcrafted vs. deep learning methods for multimodal hand biometric recognition," Scientific Reports, vol. 15, Art. 29237, 2025, doi: 10.1038/s41598-025-10075-1.
[7] J. C. Atenco, J. C. Moreno, and J. M. Ramirez, "Audiovisual biometric network with deep feature fusion for identification and text prompted verification," Algorithms, vol. 16, no. 2, Art. 66, 2023, doi: 10.3390/a16020066.
[8] F. W. Ipeayeda, M. O. Oyediran, S. A. Ajagbe, J. O. Jooda, and M. O. Adigun, "Optimized gravitational search algorithm for feature fusion in a multimodal biometric system," Results in Engineering, vol. 20, Art. 101572, 2023, doi: 10.1016/j.rineng.2023.101572.
[9] C. Kamlaskar and A. Abhyankar, "Feature level fusion framework for multimodal biometric system based on CCA with SVM classifier and cosine similarity measure," Australian Journal of Electrical and Electronics Engineering, vol. 20, no. 2, pp. 205–218, 2023, doi: 10.1080/1448837X.2022.2129147.
[10] P. Chitrapu, M. K. Morampudi, and H. K. Kalluri, "A secure and explainable multimodal biometric system using trust adaptive fusion for face and fingerprint," Scientific Reports, vol. 16, Art. 14244, 2026, doi: 10.1038/s41598-026-43252-x.
[11] S. Ayeswarya and K. John Singh, "Enhancing security and usability with context aware multi-biometric fusion for continuous user authentication," Scientific Reports, vol. 15, Art. 30627, 2025, doi: 10.1038/s41598-025-14833-z.
[12] O. N. Kadhim and M. H. Abdulameer, "A multimodal biometric database and case study for face recognition based deep learning," Bulletin of Electrical Engineering and Informatics, vol. 13, no. 1, 2024, doi: 10.11591/eei.v13i1.6605.
[13] O. N. Kadhim and M. H. Abdulameer, "Biometric identification advances: Unimodal to multimodal fusion of face, palm, and iris features," Advances in Electrical and Computer Engineering, vol. 24, no. 1, pp. 91–98, 2024, doi: 10.4316/AECE.2024.01010.
[14] S. A. El_Rahman and A. S. Alluhaidan, "Enhanced multimodal biometric recognition systems based on deep learning and traditional methods in smart environments," PLOS ONE, vol. 19, no. 2, e0291084, 2024, doi: 10.1371/journal.pone.0291084.
[15] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," Patterns, vol. 4, no. 9, Art. 100804, 2023, doi: 10.1016/j.patter.2023.100804.
[16] B. Bischl et al., "Hyperparameter optimization: Foundations, algorithms, best practices and open challenges," WIREs Data Mining and Knowledge Discovery, vol. 13, no. 2, e1484, 2023, doi: 10.1002/widm.1484.
[17] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," Applied Soft Computing, vol. 8, no. 1, pp. 687–697, 2008, doi: 10.1016/j.asoc.2007.05.007.
[18] Ö. Inik, "SwarmCNN: An efficient method for CNN hyperparameter optimization using PSO and ABC metaheuristic algorithms," The Journal of Supercomputing, vol. 81, Art. 874, 2025, doi: 10.1007/s11227-025-07347-y.