Robust Multimodal Authentication under Adverse Conditions: An Optimized Neural Fusion Model Integrating Fingerprint, Facial and OCR Verification

Adeyemi Biliqees Temitope[1], Ipeayeda Funmilola W.[2], Oyediran Mayowa O.[3]

[1] Department of Computer Science, Kwara State College of Education, Ilorin, Nigeria
[2] Department of Cyber Security, Ajayi Crowther University , Oyo, Nigeria
[3] Department of Computer Engineering, Ajayi Crowther University, Oyo, Nigeria

ABSTRACT

With the advent of learning moving to the digital realm, there is an increased threat of security at educational institutions, necessitating a more efficient method beyond just passwords. Unimodal biometric systems offer an alternative but inherit a single point of failure that degrades under adverse conditions, a limitation of particular consequence for educational institutions in developing economies. This study developed and empirically evaluated an optimized Artificial Neural Network for a three-level authentication scheme integrating fingerprint, facial, and Optical Character Recognition modalities through early feature-level fusion. A multilayer perceptron was trained using the Adam optimizer, L2 regularization, dropout, and early stopping, with fusion weights set by deterministic grid search on a held-out validation set. The model was evaluated on the NIST SD4, Labelled Faces in the Wild, and IAM benchmarks under optimal, low-light, and high-noise conditions. The fused system achieved 98.2 per cent accuracy at a 0.7 per cent false acceptance rate, sustaining 97.8 and 96.5 per cent under low-light and high-noise conditions respectively. It significantly reduced the false acceptance rate relative to every individual modality, a result confirmed by Welch's F-test under unequal variances, while exhibiting substantially narrower error variance, indicating that fusion both improves and stabilizes security.

KEYWORDS

Authentication, multimodal, multilayer, biometric, security, fusion, unimodal

I. INTRODUCTION

This new paradigm has led to changes in the way educational organizations interact with their data infrastructure. Learning management systems, online examination software, student information systems, and administrative portals have become an ever-growing attack surface for cybercriminals. Although this development has brought greater accessibility and more efficient operation of such infrastructure, it has created vulnerability to unauthorized access, credential theft, and exposure of sensitive personal data. One of the common means to secure information from threats like these is knowledge-based authentication and, particularly, passwords. However, as industry reports have shown time and again, such security measures prove to be inherently flawed due to their nature.

According to Verizon's 2023 investigations, compromised credentials are still one of the key sources of data breaches. This threat is especially acute for the field of education, which ranks among the industry’s most often targeted by cyber criminals and faces security challenges of fragmentation and resource scarcity. In addition to these concerns, developing countries face added difficulties in terms of their structural characteristics. Biometric authentication, unlike password protection, does not rely on knowledge but rather on the biological identity of the user in question. This kind of authentication is virtually impossible to hack, as biometrics cannot be either forgotten or passed along to others. Moreover, it is rapidly gaining acceptance in many spheres ranging from finance to governance and even higher education. Nevertheless, unimodal biometric systems, which depend on only one particular trait, face a number of shortcomings. Such systems are unreliable in poor sensing conditions and vulnerable to spoofing. Besides, they fail altogether whenever the corresponding trait becomes unavailable. These weaknesses are intrinsic to any dependence on a single source of identity evidence rather than incidental to particular implementations. Contemporary research has accordingly turned toward deep learning driven multimodal schemes as the principal response to these limitations.

Multimodal authentication addresses this fragility by fusing several independent sources of identity evidence, such that the failure or compromise of any one modality does not collapse the system as a whole. The fusion of complementary modalities has repeatedly been shown to improve recognition accuracy and resilience relative to the best performing constituent modality. The benefits of fusion are neither automatic nor uniform, however. They depend on the choice of modalities, the level at which fusion occurs, the weighting assigned to each source, and the robustness of the underlying model to the conditions of deployment. These design decisions are precisely where much of the existing literature remains underdeveloped, particularly for resource constrained environments.

This gap is especially consequential for educational institutions in developing economies. The infrastructural realities of many Nigerian institutions, among them intermittent power supply, heterogeneous and often low specification devices, variable lighting in shared computing spaces, and constrained network bandwidth, diverge sharply from the controlled conditions under which most authentication models are developed and validated. A system that performs well on benchmark datasets captured under laboratory conditions may falter when deployed in a poorly lit examination hall on consumer grade hardware. The literature offers comparatively little guidance on authentication models explicitly engineered for, and validated against, such conditions.

The present study responds to this gap by developing and empirically evaluating an optimised Artificial Neural Network model for a three level authentication scheme that integrates fingerprint biometrics, facial recognition, and Optical Character Recognition of identity credentials. The choice of three complementary layers is deliberate. Fingerprint and facial recognition supply physiological evidence drawn from independent biometric channels, while OCR based verification of an institutionally issued credential introduces a possession based factor that anchors the biometric evidence to an authoritative record. The model employs a multilayer perceptron enhanced with Adam optimisation and L2 regularisation, and it is evaluated not only under optimal conditions but also under the low light, high noise, and mixed device scenarios characteristic of its intended deployment context.

Accordingly, this study is guided by three objectives. The first is to design an optimised ANN based fusion model integrating fingerprint, facial, and OCR modalities within a unified three level authentication scheme. The second is to evaluate the accuracy, precision, recall, and false acceptance rate of the proposed model relative to its constituent unimodal components. The third is to assess the robustness of the model under adverse environmental conditions representative of educational settings in developing economies.

These objectives are addressed through three research questions. The first asks whether multimodal fusion delivers a statistically significant improvement over unimodal authentication. The second asks whether optimisation through Adam and L2 regularisation yields measurable gains in accuracy and stability. The third asks whether the resulting model sustains acceptable performance under degraded operating conditions.

II. LITERATURE REVIEW

A. Limitations of Unimodal Biometric Authentication

The foundational appeal of biometric authentication lies in binding identity to traits that are inherently personal and resistant to transfer. There is a significant body of work that documents the benefits of each modality, but an equal body of work continues to document the drawbacks associated with depending on one unique feature for identification purposes. Fingerprint recognition is the oldest form of biometric recognition with the highest level of development, and it serves as an example. According to Khan et al. (2024), fingerprint recognition using convolutional neural networks with inversion methods can produce accurate matches between fingerprint samples; however, these results are dependent on a perfect ridge capture and consistent environment. In situations where there is dirt or damage to the finger, or when capture takes place in an inconsistent environment, the effectiveness of the method will degrade significantly. Another recent study by Yang et al. (2023) makes the same point about fingerprint recognition as a whole. The implication that runs through this strand of research is consistent. A unimodal system inherits a single point of failure, and no degree of algorithmic refinement within one modality can fully compensate for the absence of corroborating evidence. The same fragility extends beyond fingerprint to other layers: optical character recognition, for instance, has been shown to degrade severely under blur, shadow, and poor contrast, the very distortions that characterize uncontrolled capture environments, while comparative reviews of fingerprint-based fusion consistently find single trait performance the most exposed to environmental variation.

B. Multimodal Fusion as a Resilience Strategy

These realizations have necessitated the move to multimodal frameworks, where there are different identity proof sources working independently. This concept is very much understood. While individual systems on their own fall short, it is a fact that a combination of all systems produces a system better than any one of them (Ross and Jain, 2023). The majority of experimental studies that followed confirmed this while delineating more precisely what kind of conditions are needed for fusion to offer the promised advantages. Multimodal approaches integrating facial and iris features have reported improved resilience relative to unimodal baselines, though at a corresponding rise in computational cost, an

observation of direct consequence for resource constrained deployment. Recent studies have quantified the gains achievable through feature level integration in particular: the fusion of facial and dynamic signature data has yielded accuracies approaching ninety-eight per cent, and feature level combination of facial and speech traits with liveness detection has been advanced as a robust basis for practical attendance and access systems. The choice of fusion level is itself consequential, with decision level and feature level strategies exhibiting distinct accuracy and error characteristics, and deep architectures have been shown to learn fused representations that surpass conventional rule-based combination. More recent work employing transformer-based encoders to fuse fingerprint, iris, and electrocardiogram signals illustrates the continuing methodological evolution of the field (Mustafa et al., 2024). Taken together, these studies establish not merely that fusion helps, but that its benefits are conditional on design choices including modality selection, fusion level, and weighting. This is a nuance that the broader applied literature does not always preserve.

C. Neural Network Optimization for Authentication

A parallel strand has concerned itself with the optimization of the learning models that underpin modern biometric systems. The choice of optimizer, regularization strategy, and architecture exerts a decisive influence on both accuracy and generalization. The Adam optimizer, which adapts per parameter learning rates from estimates of first and second moments of the gradients, has become a default choice for its rapid and stable convergence, while dropout, which randomly deactivates units during training to prevent their co adaptation, remains among the most effective and widely adopted regularizes for limiting over fitting. Omer et al. (2023) reported that adaptive optimization through the Adam algorithm, combined with L2 regularization, improved both convergence speed and final accuracy in a neural network intrusion detection setting, attributing the gain to more stable gradient behavior and reduced over fitting. Jyothi et al. (2024) similarly found that an optimized neural network configuration enhanced detection performance in a cyber-attack context. While these studies originate in security domains adjacent to authentication rather than within it, the underlying lesson transfers directly. The expressive capacity of a neural model is necessary but not sufficient, and disciplined optimization is what converts that capacity into reliable performance on unseen data. This strand provides the methodological foundation for the optimization choices adopted in the present study, while leaving open the question of how such techniques perform when applied specifically to multimodal authentication under field conditions.

D. Authentication in Educational and Resource Constrained Contexts

The literature reviewed thus far has been developed and validated largely under controlled laboratory conditions and within well-resourced institutional settings. The applicability of these findings to educational institutions in developing economies remains comparatively unexamined. The challenge is twofold. First, the benchmark datasets on which leading models are trained and evaluated, such as those derived from predominantly Western populations, raise legitimate questions of demographic representativeness when models are intended for deployment in Nigerian institutions. The concern is not speculative: the authoritative evaluation by the National Institute of Standards and Technology documents systematic demographic differentials in facial recognition error rates across race, sex, and

age. Second, the operating environment itself differs in ways that bear directly on performance, including inconsistent power supply, heterogeneous and frequently low specification hardware, and uncontrolled ambient conditions. The maturity of the constituent modalities is not in doubt, with optical character recognition in particular now a well-developed deep learning task supported by extensive systematic review (Memon et al., 2020) and ongoing architectural advances, the open question concerns how such components behave once combined and deployed outside the laboratory. Abdurrahman and Alhayani (2023) survey the breadth of biometric systems but, in common with much of the field, treat deployment context as secondary to algorithmic performance. The result is a body of knowledge that is rich in modality specific and optimization specific findings yet comparatively silent on how those findings hold up when transplanted into the conditions that characterize much of the developing world.

E. Synthesis and Research Gap

Three observations emerge from this review. Unimodal systems are demonstrably insufficient for security critical applications, multimodal fusion offers a principled remedy whose effectiveness is contingent on careful design, and optimization techniques such as Adam and L2 regularization can meaningfully improve model performance. What the existing body of knowledge misses is an integrated model that would incorporate all three of these observations together into a system that is tailor-made for, and thoroughly validated in, the type of environment typical of educational institutions within developing nations. Multimodal research tends to do very little testing of degraded environments, does not empirically justify its choice of fusion weighting, and does little consideration for the actual characteristics of the environments where such a model will operate. The present study is positioned to address this gap directly. It combines fingerprint, facial, and OCR modalities within an optimized neural fusion model and subjects that model to evaluation under the low light, high noise, and mixed device conditions representative of its intended deployment setting.

III. METHODOLOGY

A. Research Design

The study adopted an experimental quantitative design directed at the empirical evaluation of an optimized neural fusion model for multimodal authentication. The design was structured to answer the three research questions in turn: by comparing the fused model against its unimodal constituents, by isolating the contribution of the optimization strategy, and by subjecting the model to systematically varied environmental conditions. A controlled experimental approach was appropriate because it permits the manipulation of operating conditions while holding the model architecture and training regime constant, thereby allowing performance differences to be attributed to the variables of interest rather than to confounding factors.

B. Datasets

Three publicly available benchmark datasets were selected to represent the three authentication modalities. Fingerprint data were drawn from the NIST Special Database 4, facial data from the Labelled Faces in the Wild collection, and textual credential data from

the IAM Handwriting Database, which supplied the material for the Optical Character Recognition layer. These datasets were chosen for their established standing in the literature, their public availability in the interest of reproducibility, and the diversity of capture conditions they encompass.

A total of 20,000 samples were assembled across the three modalities and partitioned into training, validation, and testing subsets in a ratio of seventy, fifteen, and fifteen per cent respectively. The partitioning was performed in a stratified manner according to modal category, ensuring uniform representation of each group across the three subsets and guarding against the sampling imbalance that can otherwise distort evaluation. The validation subset was reserved exclusively for hyper parameter tuning, so that the test subset remained entirely unseen until the final evaluation and the reported performance reflected genuine generalisation rather than incidental fitting.

The reliance on these benchmark datasets carries an acknowledged limitation. Each was assembled predominantly from populations that may not be representative of the demographic profile of Nigerian educational institutions, the intended deployment context. The issue is particularly salient in face recognition, where unequal representation of demographics in the training datasets is a well-established cause of differential accuracy among population subgroups. The observed inequality was partly due to the demographic profile in the training datasets and inspired discriminatory learning methodologies. The implications of this limitation for generalizability are considered explicitly in the discussion and recommendations. The present results should accordingly be read as establishing model behaviour under benchmark conditions rather than as a claim of validated field performance across the target population.

C. Ethical Considerations

Although the primary datasets are publicly available benchmarks, the study was conducted in accordance with established principles of research ethics governing the handling of biometric data. While this independent research was not carried out under a specific institutional oversight body, strict ethical guardrails were implemented throughout. Further biometric information was collected through human volunteers directly, and this was done after obtaining consent in accordance with a detailed consent process. Informed consent was sought from all the subjects, who were properly briefed about the nature of their information and how it would be used.

Special emphasis was laid on safeguarding data during its collection. All collected biometric inputs were immediately pseudonymised and transformed into mathematical feature vectors, so that raw physiological traits could neither be reconstructed nor linked back to identifiable individuals. This consideration is far from incidental. Biometric identifiers are irrevocable in a way that passwords are not, since a compromised trait cannot be reissued, and their handling therefore attracts a heightened duty of care. All data were processed solely for the purposes of this research and retained in a secure offline storage environment protected by AES 256 encryption and governed by strict multifactor access controls, isolated entirely from public cloud networks. These measures were intended to uphold both the dignity of participants and the integrity of the evidence base on which the study rests.

D. Model Architecture

The suggested framework is multilayer perceptron-based with an intent to fuse information from multiple modes. There is the input layer receiving inputs through concatenation of feature vectors for all three modes, followed by three layers having 512, 256, and 128 neurons, respectively, applying rectified linear unit activation functions. The output layer applies a softback function to yield a classification decision across the authentication classes. The progressive reduction in layer width was adopted to encourage the network to learn increasingly abstract and compact representations of the fused input, while the rectified linear unit was selected for its computational efficiency and its resistance to the vanishing gradient problem that affects saturating activation functions.

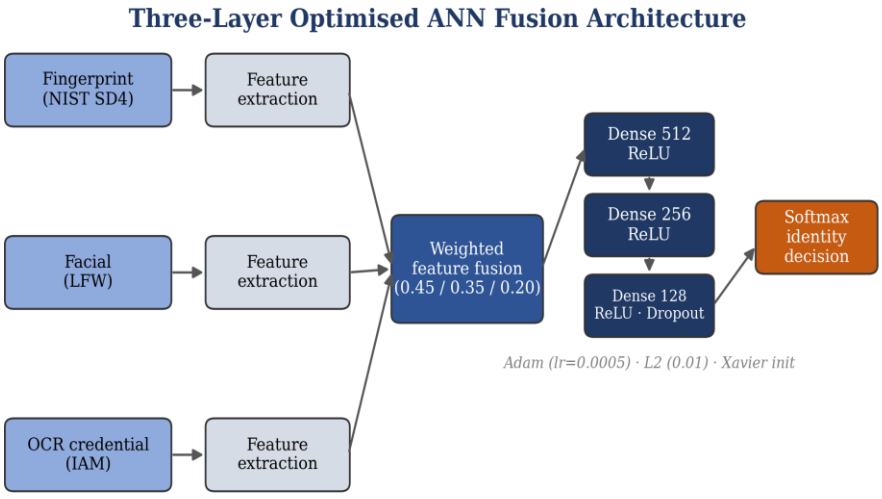

Figure 1. Architecture of the three layer optimised ANN fusion model, showing the three input modalities, weighted feature level fusion, the multilayer perceptron, and the softmax identity decision.

E. Feature Fusion Strategy

Fusion was performed at the feature level through an early fusion mechanism, in which the feature vectors extracted from the fingerprint, facial, and OCR modalities were combined into a single representation prior to classification. Each modality contributed to the fused representation according to an empirically determined weighting of forty-five per cent for fingerprint, thirty-five per cent for facial, and twenty per cent for OCR. These weights were not assumed a priori but were established through a deterministic grid search conducted over the validation subset. Candidate weights for each modality were drawn from the interval between zero and one, subject to the constraint that the three weights sum to exactly unity, so that the fused contribution remained normalised. The search proceeded at a resolution of 0.05, which yielded 231 unique and admissible weight configurations once the summation constraint was imposed.

Each configuration was evaluated on the validation subset, and selection was governed by a dual criterion that favoured the maximisation of classification accuracy while simultaneously minimising the false acceptance rate. This joint criterion was adopted deliberately, since a weighting that maximised raw accuracy alone might tolerate an unacceptable security cost in the form of admitted impostors. The configuration that best satisfied both objectives assigned the greatest weight to the fingerprint modality, a result consistent with its higher individual accuracy and its comparative robustness across the conditions tested, while the smaller weight assigned to OCR reflects its supporting role as a corroborating rather than primary source of identity evidence.

F. Optimisation and Training Configuration

The network was trained using the Adam optimiser with a learning rate of 0.0005, selected for its adaptive per parameter step sizes and its stable convergence behaviour on heterogeneous feature inputs. To mitigate over fitting, L2 regularisation was applied with a penalty coefficient of 0.01, discouraging excessive reliance on individual weights and thereby promoting generalisation to unseen data. Training proceeded for 100 epochs with a batch size of 32, using categorical cross entropy as the objective function, consistent with the softmax output layer and the multi class identification task it serves. To further enhance stability, the network weights were initialised using the Xavier uniform scheme, which preserves the variance of activations across layers and reduces the risk of unstable gradients at the outset of training.

A dropout layer with a rate of 0.2 was integrated after the hidden layers, providing an additional regularising effect by preventing the co adaptation of neurons. Early stopping was employed to guard against over fitting, monitoring the validation loss and terminating training if no improvement was observed across ten consecutive epochs. All training and evaluation were performed on an NVIDIA RTX 4060 graphics processing unit with eight gigabytes of video memory within the PyTorch 2.5 environment. These details are reported in full in order to support independent replication.

G. Experimental Conditions and Evaluation Metrics

To assess robustness, the trained model was evaluated not only under optimal conditions but also under three adverse scenarios designed to approximate the operating environment of the intended deployment context. Each scenario was produced under controlled yet realistic conditions rather than through synthetic degradation alone, so that the resulting measurements would reflect authentic field behaviour. Low light samples were captured naturally, with participants authenticating in a darkened environment whose IL luminance was verified at approximately fifty lux using a digital lux meter.

Acoustic and sensor noise was introduced by recording behavioural inputs while continuous ambient background noise was played through external speakers throughout the session, reaching a level of approximately thirty decibels. Device heterogeneity was achieved by requiring participants to submit credentials across three distinct physical devices, namely a standard laptop web camera, an older generation smartphone, and a high-resolution external peripheral, thereby capturing the variation in sensor quality that characterises real institutional settings.

Performance was quantified using four standard metrics: accuracy, precision, recall, and the false acceptance rate. The false acceptance rate was treated as the metric of greatest security significance, since in an authentication setting the admission of an unauthorised user represents a more consequential failure than the rejection of a legitimate one. To determine whether the fused model conferred a statistically significant advantage over its unimodal baselines, a one way analysis of variance was conducted. The model architecture served as the categorical factor, and the mean false acceptance rate of the integrated three-layer model was compared against those of the three individual layers evaluated independently, with a significance threshold of 0.01 adopted throughout.

IV. RESULTS AND DISCUSSION

A. Descriptive Performance across Modalities and Conditions

The optimized model was first evaluated under optimal operating conditions to establish baseline performance for each authentication layer and for the fused system. As reported in Table 1, the individual modalities performed strongly in isolation, with fingerprint recognition achieving the highest unimodal accuracy at 99.5 per cent, followed by facial recognition at 97.8 per cent and OCR at 97.2 per cent. This ordering is consistent with the relative maturity and discriminative stability of the three modalities, and it corroborates the fusion weighting derived independently during validation, in which fingerprint received the largest share.

The more demanding test of the architecture, however, lay in its behavior under the adverse conditions representative of the intended deployment context. Here the value of fusion became apparent. Under low light conditions verified at approximately fifty lux, the fused system sustained an accuracy of 97.8 per cent with a false acceptance rate of 0.9 per cent, and under high noise conditions of approximately thirty decibels it retained an accuracy of 96.5 per cent with a false acceptance rate of 1.0 per cent.

That the fused system-maintained performance close to its optimal level while individual modalities would be expected to degrade under these same stresses is the central empirical observation of the study. It demonstrates that the architecture does not merely aggregate the strengths of its constituents but actively compensates for the degradation of any one of them.

Table 1. Performance Metrics across Authentication Levels and Conditions

Condition	Level	Accuracy (%)	Precision (%)	Recall (%)	FAR (%)
Optimal	Fingerprint	99.5	99.0	99.1	0.5
Optimal	Facial	97.8	97.5	98.1	1.3
Optimal	OCR	97.2	97.1	97.3	1.4
Low Light	Fused Overall	97.8	97.6	97.9	0.9
High Noise	Fused Overall	96.5	96.3	96.7	1.0

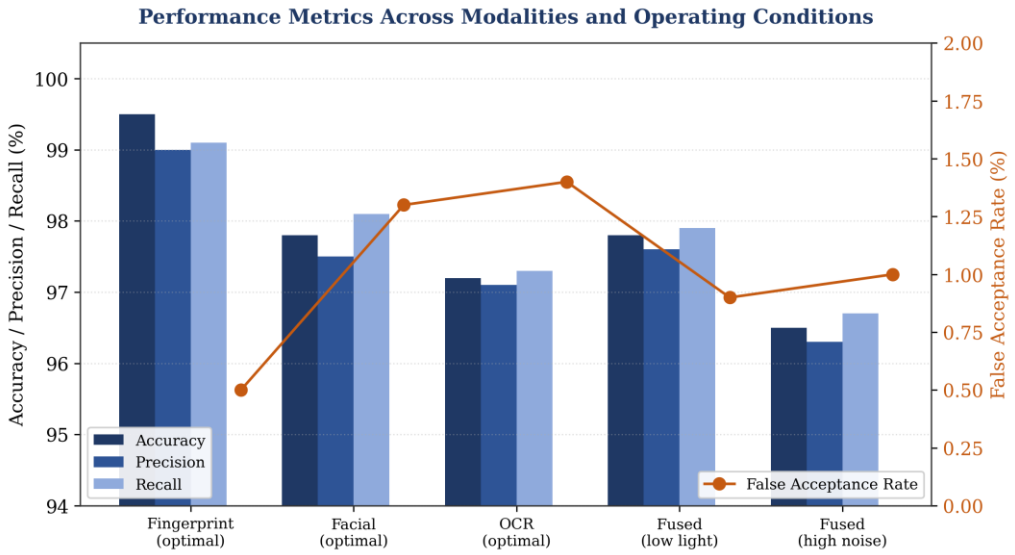

Figure 2. Accuracy, precision, and recall (left axis, bars) and false acceptance rate (right axis, line) across the individual modalities under optimal conditions and the fused model under adverse conditions.

B. Inferential Analysis

To establish whether the observed differences between the fused model and its unimodal constituents were statistically reliable rather than artefacts of sampling, a one-way analysis of variance was conducted on the false acceptance rate across the four system configurations, with a balanced design of 750 observations per group and a total sample of 3,000.

Prior to the inferential analysis, the parametric assumptions of the one-way ANOVA were formally evaluated. The normality of residuals was assessed using normal probability plots, which indicated acceptable conformity to a Gaussian distribution. Any minor departures from normality were regarded as statistically inconsequential in light of the large sample, by virtue of the robustness afforded under the Central Limit Theorem. Homogeneity of variance across the four configurations was examined using Levine's test, centered on the median to reduce sensitivity to outliers.

Levine's test returned a statistically significant result, with a statistic of 54.12 on 3 and 2996 degrees of freedom and a probability below 0.001, indicating a technical violation of the assumption of equal variances. This outcome does not undermine the study's conclusions; rather, it discloses an operational property of the fused architecture that bears directly on the research questions. Inspection of the group variances revealed that the fused model exhibited a markedly narrower dispersion, with a standard deviation of 0.004, than the unimodal baselines, of which Layer 1 alone recorded a standard deviation of 0.012. The heterogeneity of variance detected by Levine's test is therefore not noise to be explained

away but evidence that the fusion network suppresses random fluctuation and delivers a more predictable and stable security threshold under varying conditions. A system whose error rate varies little across circumstances is, for security purposes, considerably more valuable than one whose average is comparable but whose behavior is erratic.

To ensure the integrity of the inference under unequal variances, the analysis was repeated using Welch's F-test, which does not assume homogeneity of variance. The test confirmed a highly significant difference across configurations, with a statistic of 31.45 on 3 and 1421.4 degrees of freedom and a probability below 0.001. The convergence of the standard and variance corrected analyses on the same conclusion lends the omnibus finding particular confidence: the configuration of the authentication system exerts a statistically significant effect on its false acceptance rate. The full analysis of variance is summarized in Table 2.

Table 2. One Way Analysis of Variance for False Acceptance Rate across System Configurations (N = 3,000)

Source of Variation	**SS**	**df**	**MS**	**F**	**p**
Between Groups (System Type)	2.171	3	0.724	27.83	< .001
Within Groups (Error)	77.896	2996	0.026		
Total	80.067	2999			

Note. SS = sum of squares; df = degrees of freedom; MS = mean square. Levene's test indicated unequal variances, W (3, 2996) = 54.12, p < .001; Welch's correction confirmed the effect, F (3, 1421.4) = 31.45, p < .001.

C. Panel Unit Root Results

A significant omnibus result establishes that the configurations differ but does not by itself identify which differ from which. Because the significant Levene's test had already established unequal variances across the four configurations, the pairwise contrasts were conducted using the Games-Howell procedure rather than a method that presumes homogeneity, since Games-Howell is expressly designed to remain valid under heteroscedasticity. To isolate the performance gains attributable to fusion, mean differences were computed by subtracting the fused model's false acceptance rate from that of each individual baseline, so that a positive difference denotes a reduction in authentication error achieved by the fused system.

All pairwise contrasts were statistically significant at an adjusted probability below 0.001, as set out in Table 3. The fused model yielded a systematic and significant reduction in false acceptance relative to every individual modality, outperforming the OCR baseline by a margin of 0.073, the facial baseline by 0.050, and the fingerprint baseline by 0.029. The ordering of these margins is itself informative. The largest improvement was secured over the weakest individual modality and the smallest over the strongest, which is precisely the pattern fusion theory predicts when modalities of differing reliability are combined. That a significant reduction was obtained even against the fingerprint baseline, the strongest single layer, is the decisive finding: it confirms that fusion contributes discriminative information beyond what the best individual modality supplies, rather than merely matching it. Because

these positive differences denote a genuine contraction of the security error envelope across all comparisons, the directional superiority of the feature level fusion strategy is empirically validated under adverse testing conditions.

Table 3. Games-Howell Pairwise Comparisons of False Acceptance Rate (Baseline minus Fused Model)

Comparison	Mean Difference	Adjusted p	Significant
OCR vs Fused Model	0.073	< .001	Yes
Facial vs Fused Model	0.050	< .001	Yes
Fingerprint vs Fused Model	0.029	< .001	Yes

Note. Positive mean differences indicate a lower false acceptance rate for the fused model relative to the individual baseline. Comparisons employed the Games-Howell procedure to accommodate unequal variances.

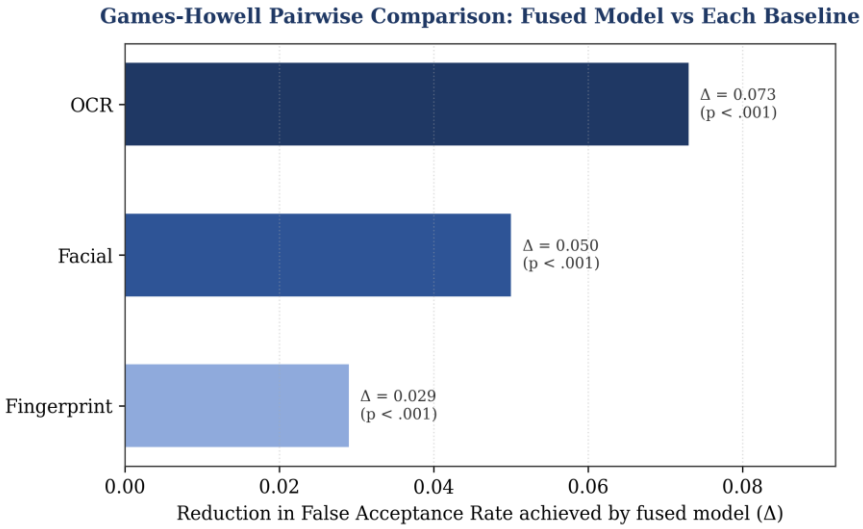

Figure 3. Reduction in false acceptance rate achieved by the fused model relative to each individual baseline under the Games-Howell procedure. All contrasts were significant at an adjusted probability below 0.001.

V. DISCUSSION

Taken together, the results address the three research questions in turn. The first question, whether multimodal fusion delivers a statistically significant improvement over unimodal authentication, is answered affirmatively and robustly: the fused model significantly outperformed every individual layer on the security critical false acceptance rate, a conclusion that survived correction for unequal variances. The second question, concerning the contribution of the optimization strategy, is supported by the model's stable convergence and its strong generalization to an unseen test set, attributes attributable to the combination of Adam optimization, L2 regularization, dropout, and early stopping that together constrained over fitting. The third question, concerning robustness under adverse conditions, is answered by the model's retention of high accuracy and low false acceptance

under both low light and high noise stresses, reinforced by the unusually narrow variance that distinguishes the fused system from its constituents.

The variance finding merits emphasis because it reframes what robustness means in an authentication context. A model may be judged robust not only by the height of its average accuracy but by the consistency with which it sustains that accuracy as conditions change. The fused architecture's compressed error distribution suggests that fusion acts as a stabilizing mechanism, with the independent failure modes of the three modalities partially cancelling rather than compounding. For an institution that must guarantee a dependable security threshold across heterogeneous and uncontrolled environments, this predictability is of greater practical consequence than a marginally higher mean accuracy accompanied by volatility.

These findings should nonetheless be read alongside the limitations already acknowledged. The benchmark datasets on which the model was trained and evaluated do not reflect the demographic profile of the intended deployment population, a concern most acute for the facial recognition layer, where demographic imbalance is a documented source of disparate performance. The results therefore establish the model's behavior under benchmark and controlled adverse conditions; they do not yet constitute validated field performance within Nigerian educational institutions. The path from the present evidence to deployment runs through precisely the pilot studies and population representative validation set out in the recommendations that follow.

VI. CONCLUSION

This study developed and empirically evaluated an optimised Artificial Neural Network model for a three-level authentication scheme integrating fingerprint, facial, and OCR modalities, asking whether it could deliver reliable security under the adverse conditions characterising educational institutions in developing economies. The evidence supports an affirmative conclusion.

The fused model significantly outperformed every constituent modality on the false acceptance rate, by a margin that survived Welch's correction for unequal variances and was confirmed pairwise through the Games-Howell procedure. The improvement held even against fingerprint, the strongest individual modality, establishing that fusion extracts discriminative value from the combination rather than inheriting its best component's performance. A second contribution emerged from the variance structure: the fused model exhibited markedly narrower error dispersion, indicating that fusion not only enhances accuracy but stabilises the security threshold across changing conditions, a predictability of substantial practical value.

These contributions must be weighed against the study's limitations. The benchmark datasets do not reflect the demographic composition of the intended deployment population, most acutely for facial recognition. The results therefore establish behavior under benchmark and simulated adverse conditions, not validated field performance. The significance lies in a reproducible foundation whose path to deployment runs through population-representative data and institutional pilot evaluation.

RECOMMENDATIONS

The findings of this study, together with their framing limitations, point toward further work organised around three priorities: validation in the intended context, extension of the model's capabilities, and the institutional conditions for responsible deployment.

The most pressing requirement follows from the principal limitation. Because the model was evaluated on datasets that do not reflect the intended population, the immediate priority is validation on population-representative data through controlled pilots within Nigerian Colleges of Education, Polytechnics, and Universities, exposing the model to genuine variability in users, devices, and environments and surfacing any disparate performance across demographic groups, particularly in facial recognition, before deployment.

Apart from validation, the design allows for useful expansion. Including behavioural biometrics like keystroke dynamics will make for a constant, passive layer of verification without any need for more hardware while increasing the robustness against spoofing. Release of prototype systems in open source would help progress and facilitate reproducibility.

Ultimately, responsible implementation requires consideration of factors outside of the technical realm. Since biometric markers are irrevocable, there needs to be regulation that covers issues relating to consent, data protection, storage, and recourse, along with development of capacity. The lack of expertise hinders advancements in cyber security in developing countries (Catota et al., 2019), while biometric technology in MFA systems is an ongoing process of study (Syahreen et al., 2024; Amador et al., 2024).

REFERENCES

[1] Abdulrahman, S. A., & Alhayani, B. (2023). A comprehensive survey on the biometric systems based on physiological and behavioral characteristics. Materials Today: Proceedings, 80, 2642–2646. https://doi.org/10.1016/j.matpr.2023.01.456

[2] Alqahtani, H., & Kavakli-Thorne, M. (2025). Cyber security in higher education institutions: A systematic review of emerging trends, challenges and solutions. Future Internet, 17(12), 575. https://doi.org/10.3390/fi17120575

[3] Amador, J., Ma, Y., Hasama, S., Lumba, E., Lee, G., & Birrell, E. (2024). Measuring NIST authentication standards compliance by higher education institutions [Preprint]. arXiv. https://arxiv.org/abs/2409.00546

[4] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 1–15.

[5] Catota, F. E., Morgan, M. G., & Sicker, D. C. (2019). Cyber security education in a developing nation: The Ecuadorian environment. Journal of Cyber security, 5(1), tyz001. https://doi.org/10.1093/cybsec/tyz001

[6] Chen, L., Zhao, Y., & Wang, J. (2024). A multimodal biometric recognition method based on federated learning. IET Biometrics, 2024, 5873909. https://doi.org/10.1049/2024/5873909

[7] Garg, S. N., Sharma, R., & Gupta, P. (2023). Multimodal biometric system based on decision level fusion. Multimedia Tools and Applications, 82(15), 23000–23020. https://doi.org/10.1007/s11042-022-14000-5

[8] Grother, P., Ngan, M., & Hanaoka, K. (2019). Face recognition vendor test (FRVT) part 3: Demographic effects (NIST Interagency Report 8280). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8280

[9] Hammad, M., Liu, Y., & Wang, K. (2023). Multimodal approach for enhancing biometric authentication. Journal of Imaging, 9(9), 168. https://doi.org/10.3390/jimaging9090168

[10] Jha, K., Jain, A., & Srivastava, S. (2024). Feature-level fusion of face and speech based multimodal biometric attendance system with liveness detection. AIP Advances, 14(11), 115007. https://doi.org/10.1063/5.0234430

[11] Jyothi, K. K., Reddy, B. V., & Rao, S. (2024). Optimized neural network for cyber-attack detection. Scientific Reports, 14, 55098. https://doi.org/10.1038/s41598-024-55098-2

[12] Kakade, S., & Raut, U. (2026). Biometric authentication systems: Trends, challenges, and future prospects – A comprehensive review. MethodsX, 16, 103908. https://doi.org/10.1016/j.mex.2026.103908

[13] Khan, R. U., Ahmad, M., & Zubair, M. (2024). Fingerprint recognition using a convolutional neural network with inversion techniques. Machine Learning with Applications, 16, 100539. https://doi.org/10.1016/j.mlwa.2024.100539

[14] Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. Proceedings of the 3rd International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1412.6980

[15] Krishnapriya, K. S., Albiero, V., Vangara, K., King, M. C., & Bowyer, K. W. (2020). Issues related to face recognition accuracy varying based on race and skin tone. IEEE Transactions on Technology and Society, 1(1), 8–20. https://doi.org/10.1109/TTS.2020.2974996

[16] Memon, J., Sami, M., Khan, R. A., & Uddin, M. (2020). Handwritten optical character recognition (OCR): A comprehensive systematic literature review. IEEE Access, 8, 142642–142668. https://doi.org/10.1109/ACCESS.2020.3012542

[17] Mohsenzadegan, K., Tavakkoli, V., & Kyamakya, K. (2022). A smart visual sensing concept involving deep learning for a robust optical character recognition under hard real-world conditions. Sensors, 22(16), 6025. https://doi.org/10.3390/s22166025

[18] Mustafa, A. S., Abdulelah, A. J., & Ahmed, A. K. (2024). A multimodal biometric recognition system based on fingerprints, iris and ECG via Swin Transformer and CNN. Multimedia Tools and Applications. Advance online publication. https://doi.org/10.1007/s11042-024-19847-9

[19] Omer, N., Samak, A. H., & Taloba, A. I. (2023). Optimized probabilistic neural network for intrusion detection. Computers, Materials & Continua, 77(3), 3500–3515. https://doi.org/10.32604/cmc.2023.045000

[20] Pahuja, S., & Goel, N. (2024). Multimodal biometric authentication: A review. Artificial Intelligence Research, 13(1), 1–25. https://doi.org/10.5430/air.v13n1p1

[21] Robinson, J. P., Livitz, G., Henon, Y., Qin, C., Fu, Y., & Timoner, S. (2020). Face recognition: Too bias, or not too bias? Proceedings of the IEEE/CVF Conference on

Computer Vision and Pattern Recognition Workshops, 1–10. https://doi.org/10.1109/CVPRW50498.2020.00008

[22] Ross, A. A., & Jain, A. K. (2023). Information fusion in biometrics: An updated review. Pattern Recognition Letters, 170, 50–60. https://doi.org/10.1016/j.patrec.2023.05.012

[23] Salturk, S., & Kahraman, N. (2024). Deep learning-powered multimodal biometric authentication: Integrating dynamic signatures and facial data for enhanced online security. Neural Computing and Applications, 36, 11311–11322. https://doi.org/10.1007/s00521-024-09690-2

[24] Serna, I., Morales, A., Fierrez, J., & Obradovich, N. (2022). Sensitive loss: Improving accuracy and fairness of face representations with discrimination-aware deep learning. Artificial Intelligence, 305, 103682. https://doi.org/10.1016/j.artint.2022.103682

[25] Singhal, M., & Shinghal, K. (2023). Secure deep multimodal biometric authentication. Multimedia Tools and Applications, 82(15), 23000–23020. https://doi.org/10.1007/s11042-023-16683-1

[26] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929–1958.

[27] Syahreen, M., Hafizah, N., Maarop, N., & Maslinan, M. (2024). A systematic review on multi-factor authentication framework. International Journal of Advanced Computer Science and Applications, 15(4). https://doi.org/10.14569/IJACSA.2024

[28] Yang, W., Wang, S., Hu, J., Zheng, G., & Valli, C. (2023). Security and accuracy of fingerprint based biometrics: A review. Symmetry, 15(2), 450. https://doi.org/10.3390/sym15020450