© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723171

An Enhanced Siamese Neural Network with Attention-Adam Optimization for Bimodal Biometric Authentication in Mobile Financial Systems

ALUKO, TOLULOPE OLUGBEMIGA[1], AYENI JOSUA AYOBAMI[2], MAKINDE OLADAYO EZEKIEL[3]

[1,2]Department of Computing Science and Digital Technology, Koladaisi University, Ibadan
[3]Department of Computer Science, Faculty of Computing, Ajayi Crowther University, Oyo, Nigeria

Abstract- The increasing use of mobile financial services has intensified the need for authentication mechanisms that combine security, usability, and computational efficiency on resource-constrained devices. Passwords and unimodal biometric systems can expose mobile financial applications to credential compromise, presentation attacks, environmental variability, and single-modality failure. This study proposes an enhanced bimodal biometric authentication framework that combines an Enhanced Siamese Neural Network (ESNN) for facial verification with hardware-backed fingerprint authentication. The facial model uses a ResNet-18 backbone enhanced with Squeeze-and-Excitation (SE) blocks for channel-wise feature recalibration. A novel Attention-Adam (Attn-Adam) optimizer is formulated using a sliding window of historical gradients and temporal attention weights to emphasize consistent optimization directions and reduce the effect of noisy gradient updates. Facial data were obtained from the Labeled Faces in the Wild (LFW) dataset, comprising 13,233 images from 5,749 identities. The thesis reports a 70:20:10 training-validation-test split and balanced positive and negative contrastive pairs. The proposed system achieved an ROC-AUC of 0.9892, EER of 4.20%, precision of 95.86%, recall of 95.74%, and F1-score of 95.80%. In the reported ablation study, the baseline ResNet-18 with standard Adam achieved AUC = 0.9152 and EER = 11.24%; adding SE blocks improved performance to AUC = 0.9581 and EER = 7.52%; and the complete ESNN with Attn-Adam achieved AUC = 0.9892 and EER = 4.20%. The model was converted to TensorFlow Lite using post-training Int8 quantization, with the reported model size decreasing from approximately 45 MB to 11 MB. The mobile authentication workflow uses serial face-then-fingerprint verification and Android Keystore/Trusted Execution Environment support for cryptographic binding. The findings indicate that combining residual feature learning, channel attention, and temporal gradient attention can improve facial verification performance while supporting deployment in mobile financial environments.

Keywords: bimodal biometrics; siamese neural network; attention-adam; resnet-18; squeeze-and-excitation; facial verification; edge ai; mobile financial security

I. INTRODUCTION

Mobile financial services have changed the way individuals access payments and other financial products. Smartphones increasingly function as interfaces to financial accounts and as local holders of sensitive authentication material. This creates a security problem: authentication must be sufficiently strong to protect financial assets while remaining fast and practical on mobile hardware. The thesis identifies password-based authentication as vulnerable to phishing, malware, social engineering and related attacks, motivating the adoption of biometric and multi-factor mechanisms (Naji & Mahmood Al-Dabbagh, 2026; NIST, 2022).

Biometric authentication provides a direct relationship between the authentication decision and a user's physiological characteristics. Face and fingerprint recognition are particularly attractive for smartphones because contemporary devices commonly provide cameras and fingerprint sensors. Nevertheless, relying on a single biometric trait creates a single point of failure. Face verification can be affected by illumination, pose, expression, occlusion and presentation attacks, while fingerprint recognition can be affected by sensor or physiological conditions. Multimodal biometric systems address this weakness by combining complementary evidence from more than one modality (Ross & Jain, 2003, 2004; U. Sumalatha et al., 2024).

A second challenge is the deployment of deep learning on resource-constrained edge devices. Large recognition models may provide strong representation capacity but can impose memory, latency and energy costs. Edge AI and quantization therefore provide an important route to privacy-preserving and low-latency biometric inference (Jacob et al., 2018; Warden & Situnayake, 2019). Siamese architectures are especially appropriate for verification because they learn a similarity function rather than a fixed set of identity classes. This permits a newly enrolled user to be verified against a stored template without retraining a conventional multiclass classifier (Bromley et al., 1993; Chopra et al., 2005; Koch et al., 2015).

The research gap identified in the thesis is the difficulty of combining robust biometric representation, optimization stability, multimodal security and mobile deployment. Existing work includes lightweight face recognition models, CNN/Siamese systems and multimodal biometric approaches, but the reviewed studies do not simultaneously provide the specific ESNN architecture, the proposed temporal gradient-attention optimizer and the serial face-fingerprint mobile security workflow described in this study (Bhavani & Karthikeyan, 2025; Kumar et al., 2023; Setyawan et al., 2025; Yuan et al., 2026).

The objective of this study is therefore to develop and evaluate an enhanced Siamese Neural Network for facial verification and integrate it with a hardware-backed fingerprint authentication stage for mobile financial authentication. The specific contributions are: (1) formulation of the Attention-Adam optimizer; (2) development of an ESNN based on ResNet-18 and SE blocks; (3) empirical evaluation using biometric verification metrics; (4) ablation analysis of the architectural and optimization components; and (5) implementation of a serial face-fingerprint authentication workflow with cryptographic binding on Android.

II. RELATED WORK

Deep learning has become central to face recognition because it supports hierarchical representation learning directly from images (LeCun et al., 2015; Goodfellow et al., 2016). Early deep face recognition systems such as DeepFace demonstrated the ability of deep convolutional representations to close the gap between machine and human face verification performance (Taigman et al., 2014). ResNet subsequently introduced residual learning and shortcut connections that facilitate optimization of deeper networks (He et al., 2015). For mobile deployment, MobileNet and MobileNetV2 introduced efficient convolutional structures and inverted residuals to reduce computational requirements (Howard et al., 2017; Sandler et al., 2018).

Siamese networks transform face recognition from fixed-class classification to metric learning. Two weight-sharing subnetworks map images into an embedding space in which samples from the same identity are expected to be close and samples from different identities are expected to be separated. Contrastive learning and related metric-learning approaches have been used extensively for verification and one-shot learning (Bromley et al., 1993; Chopra et al., 2005; Hadsell et al., 2006; Koch et al., 2015).

Attention has also become an important mechanism for emphasizing informative features. In the present architecture, attention is applied at the channel level through Squeeze-and-Excitation blocks. This design is intended to recalibrate feature channels so that identity-relevant information receives greater weight than less informative background responses. The thesis further introduces temporal attention at the optimization level: rather than treating every current gradient update independently, the optimizer compares the current gradient with recent gradients and uses the resulting similarities to form attention weights.

Recent mobile-oriented face-recognition studies reinforce the need for a balance between accuracy and efficiency. The thesis reviews Mobi-FaceNeXt, a lightweight model using depthwise separable convolution and combined losses, and FaceLiVT, a linear-attention vision-transformer approach for mobile devices. It also reviews MTCNN-DeepFace and attention-based ConvFaceNeXt systems. These approaches report strong benchmark performance, but the thesis identifies limitations involving pose, lighting, liveness detection, multimodal fusion, or validation in mobile financial environments (Godase, 2025; Ma et al., 2025; Setyawan et al., 2025; Yuan et al., 2026).

Multimodal biometrics can improve robustness by combining complementary traits. Score-level and other fusion strategies have been widely studied (Hall & Llinas, 1997; Ross & Jain, 2003, 2004; Abderrahmane et al., 2020). In the proposed system, fusion is serial rather than simultaneous: facial verification is performed first and the fingerprint stage is invoked only after the face gate is passed. The design is intended to reduce unnecessary fingerprint prompts for frames that have already failed facial verification.

Biometric template protection is another essential consideration because biometric characteristics cannot simply be replaced like passwords. Prior work has examined biometric template attacks, cancelable biometrics, fuzzy vaults and other protection mechanisms (Jain et al., 2008; Ratha et al., 2001, 2007; Juels & Sudan, 2006). The proposed implementation addresses this at the mobile security layer by using Android Keystore-backed cryptographic material and requiring biometric authentication before a protected operation can be performed.

III. MATERIALS AND METHODS

The study used a quantitative experimental design. Model development and optimization were performed in Python using TensorFlow/Keras, while the mobile client was implemented in Kotlin for Android. The facial verification model was designed for 128-dimensional embeddings. The reported development environment included an Intel i7 processor, 16 GB RAM and Windows 11 Pro. Software dependencies included Python 3.10, TensorFlow 2.16.1, NumPy 1.26.4, OpenCV-Python 4.9, scikit-learn 1.3.2 and Matplotlib 3.8.0.

The primary facial dataset was Labeled Faces in the Wild (LFW), reported in the thesis as containing 13,233 images of 5,749 people. The experimental protocol used positive and negative contrastive pairs with a 1:1 ratio. The thesis reports a 70% training, 20% validation and 10% test split. Facial samples were detected/aligned using MTCNN, cropped to the face region, resized to 128×128 pixels and normalized to the range [-1, 1].

The ESNN uses a ResNet-18 feature extractor augmented with Squeeze-and-Excitation blocks. The processing sequence consists of an input image, convolutional stem, pooling, four residual stages with SE-based channel recalibration, global average pooling, a dense projection layer and L2 normalization. The resulting embedding is a compact 128-dimensional vector. During training, the two Siamese branches share weights. During inference, a live embedding is compared with a stored enrollment embedding.

For a feature map X with spatial dimensions $H \times W$ and C channels, channel squeezing can be represented as $z_c = (1/(H W)) \Sigma_i \Sigma_j X_c(i,j)$. The excitation stage applies a nonlinear gating function $s = \sigma(W_2 \delta(W_1 z))$, where δ is a ReLU activation and σ is a sigmoid activation. The recalibrated feature map is then $Y_c = s_c X_c$. This mechanism provides channel-wise attention while preserving the residual structure.

The proposed Attention-Adam optimizer extends the usual momentum-based update by introducing temporal attention over a sliding window of recent gradients. Let g_t be the current gradient and $G_t = \{g_{t-k+1}, \ldots, g_t\}$ denote the recent gradient buffer of length k. A scaled similarity score can be written as $a_{t,i} = (g_t \cdot g_i)/\sqrt{d}$, where d is the gradient dimension. The normalized attention weight is $\alpha_{t,i} = \exp(a_{t,i}) / \Sigma_j \exp(a_{t,j})$. The synthesized gradient is $\hat{g}_t = \Sigma_i \alpha_{t,i} g_i$. The synthesized gradient is then supplied to the AdamW-style first- and second-moment updates. The intended effect is to emphasize gradient directions that are temporally consistent and reduce the influence of unstable updates.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723171

Training used contrastive metric learning. For a pair (x_1, x_2) with binary label y and Euclidean embedding distance D, the contrastive-loss formulation used by the thesis can be expressed as $L = yD^2 + (1-y)\max(0, m-D)^2$, where m is the separation margin. Genuine pairs are therefore encouraged to have small distances, while impostor pairs are penalized when their distance falls within the margin.

The mobile workflow uses a serial fusion strategy. CameraX captures the face image, ML Kit performs face detection, the face region is resized and normalized, and the ESNN generates a live embedding. The live embedding is compared with the stored template. When the face gate is passed, the camera is stopped and Android BiometricPrompt is invoked for fingerprint authentication. The fingerprint operation is delegated to secure device hardware. A protected wallet ledger is encrypted using AES-256 in GCM mode, with key material generated and protected by Android Keystore. The thesis reports that the key is configured to require user authentication before the protected cryptographic operation is released.

For edge deployment, the thesis reports post-training Int8 quantization in TensorFlow Lite, reducing the model size from approximately 45 MB to approximately 11 MB. The inference pipeline is throttled to approximately 5–6 frames per second by processing every fifth frame from a nominal 30-FPS camera stream. This design is intended to reduce battery consumption and thermal load.

3.1 System Architecture

The overall workflow is organized into perception, intelligence and security layers. Face data are acquired and preprocessed on the mobile edge; the ESNN generates a compact representation; and the hardware-backed fingerprint stage controls access to protected cryptographic material.

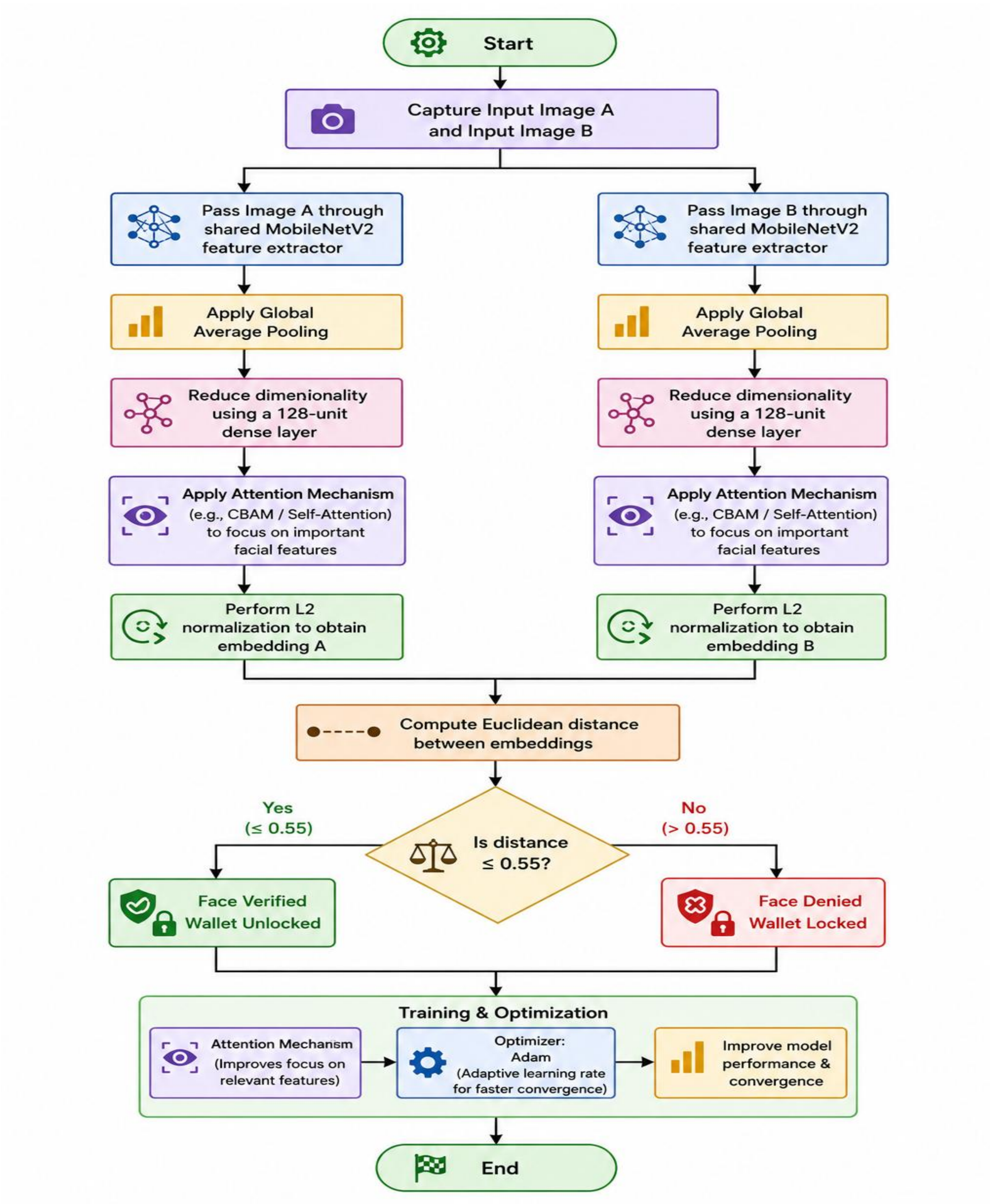

Figure 1. System flowchart for the proposed serial biometric authentication architecture.

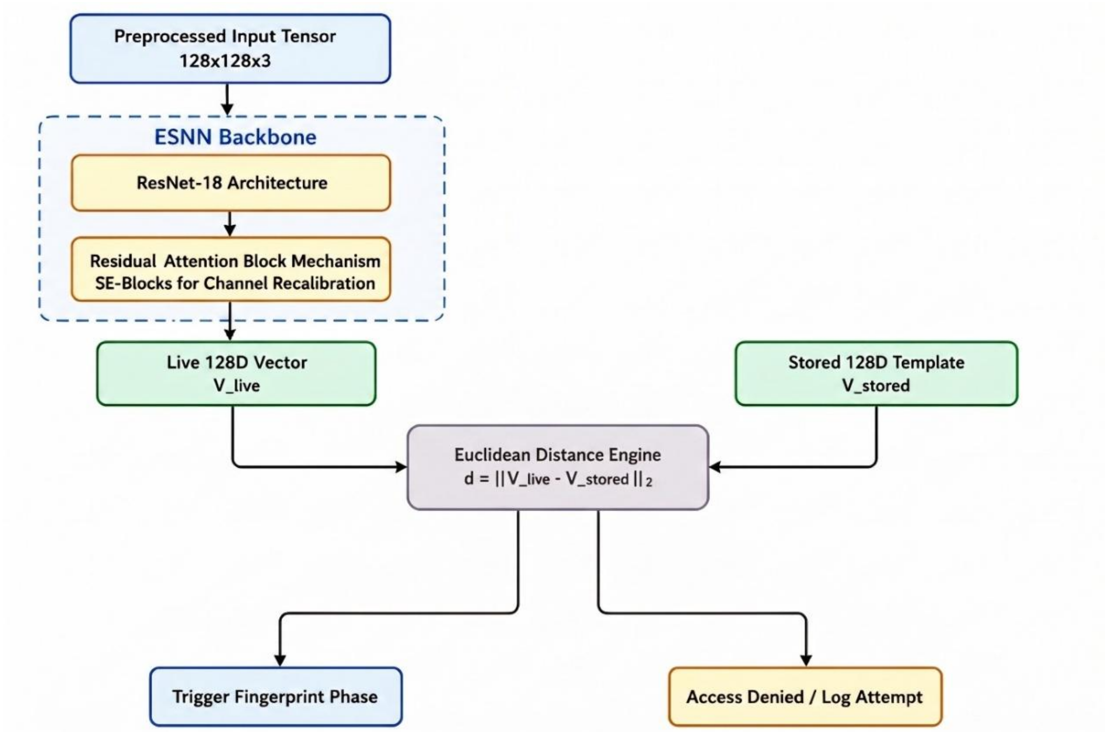

Figure 2. ESNN embedding architecture and distance-based verification workflow.

3.2 Algorithmic Formulation of Attention-Adam

Gradient buffer: $G_t = \{g_{t-k+1}, \ldots, g_t\}$

Temporal attention score: $a_{t,i} = (g_t \cdot g_i) / \sqrt{d}$

Attention weight: $\alpha_{t,i} = \exp(a_{t,i}) / \Sigma_j \exp(a_{t,j})$

Attention-weighted gradient: $\hat{g}_t = \Sigma_i \alpha_{t,i} g_i$

3.3 Evaluation Metrics

The evaluation uses ROC-AUC, EER, accuracy, precision, recall, F1-score and confusion-matrix outcomes. For binary verification, precision = TP/(TP+FP), recall = TP/(TP+FN), and F1 = 2(precision × recall)/(precision + recall). FAR and FRR are also conceptually relevant because they quantify unauthorized acceptance and legitimate-user rejection, respectively.

IV. EXPERIMENTAL RESULTS

The proposed system achieved an ROC-AUC of 0.9892 and an EER of 4.20%. The reported precision was 95.86%, recall was 95.74%, and F1-score was 95.80%. The thesis also reports an optimal threshold distance of 0.6421 in the principal performance summary. These results indicate strong separation between genuine and impostor samples under the reported LFW evaluation protocol.

The ROC analysis shows a curve approaching the upper-left region of ROC space, with AUC = 0.9892. The EER analysis reports 4.2%, representing the operating point at which false-acceptance and false-rejection rates are equal. The reported confusion matrix contains 958 true negatives, 42 false positives, 42 false negatives and 958 true positives across 2,000 evaluated verification decisions. This corresponds to an overall accuracy of 95.80%.

The cosine-similarity distributions provide a visual explanation for the performance. The thesis reports genuine-pair scores concentrated approximately between 0.75 and 1.00, with highest density around 0.85–0.90, while impostor scores are concentrated approximately between 0.00 and 0.40, with highest density around 0.15–0.20. The displayed cosine-similarity figure uses a 0.50 decision threshold and shows a substantial separation gap between the two distributions.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723171

The ablation study provides the strongest evidence for the incremental contribution of the proposed components. The baseline ResNet-18 + standard Adam configuration achieved AUC = 0.9152, EER = 11.24%, precision = 0.884 and F1 = 0.881. Adding SE blocks increased AUC to 0.9581 and reduced EER to 7.52%, with precision = 0.921 and F1 = 0.918. The complete ESNN + Attn-Adam configuration achieved AUC = 0.9892, EER = 4.20%, precision = 0.958 and F1 = 0.958.

Relative to the baseline EER of 11.24%, the final EER of 4.20% represents an approximate 62.6% reduction in error. This result supports the thesis interpretation that the combination of architectural channel recalibration and temporal gradient attention contributes more than either change alone. However, the study does not provide a separate controlled experiment isolating every possible optimizer hyperparameter or a statistical significance test across repeated independent training runs; these are therefore recommended for future validation.

4.1 Summary of Performance

Metric	Reported value
ROC-AUC	0.9892 (98.92%)
EER	4.20%
Precision	95.86%
Recall	95.74%
F1-score	95.80%
Accuracy	95.80%
Principal threshold distance	0.6421

4.2 ROC and EER Analysis

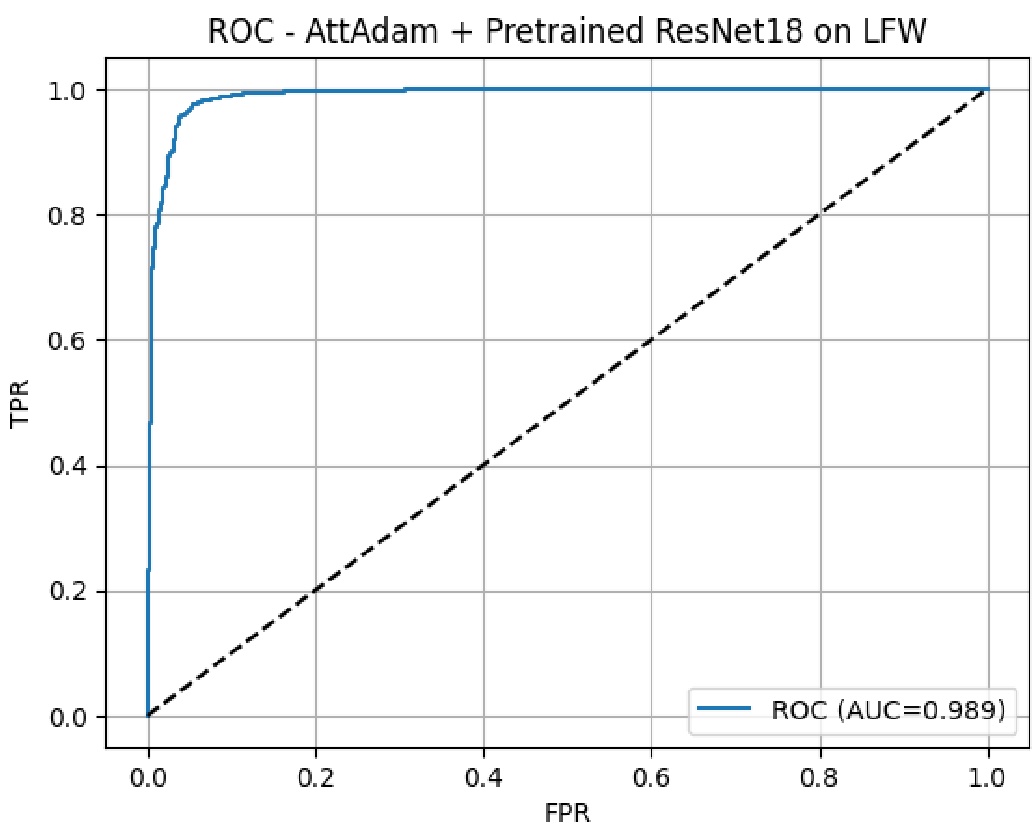

Figure 3. ROC curve for the proposed Attn-Adam model.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723171

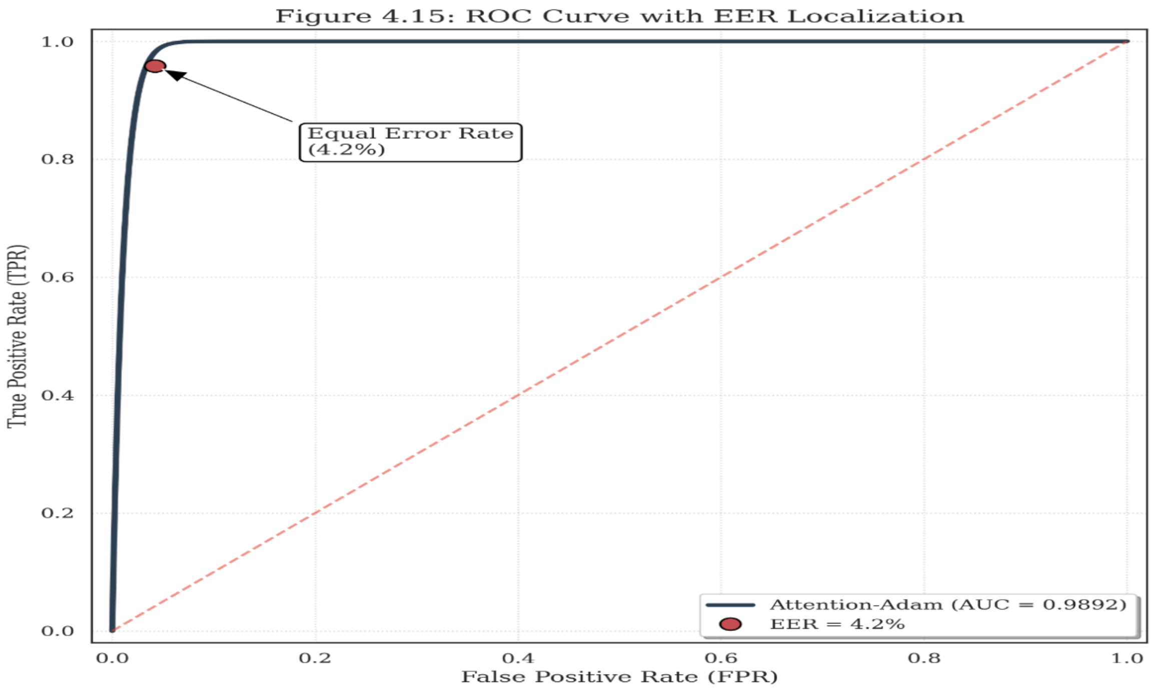

Figure 4. ROC curve with the reported equal-error-rate localization.

4.3 Confusion Matrix

Actual / Predicted	Different	Same
Different	958	42
Same	42	958

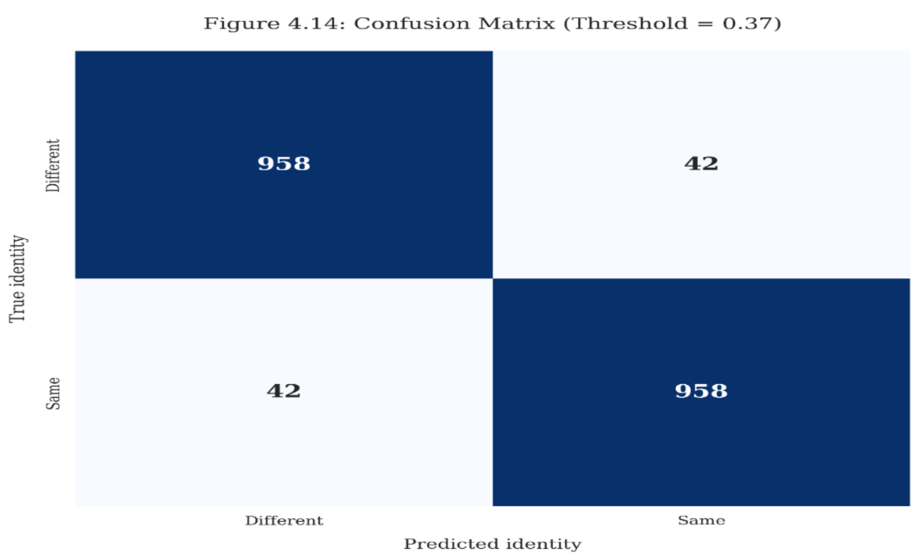

Figure 5. Confusion matrix for 2,000 reported verification decisions.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723171

4.4 Similarity Distribution

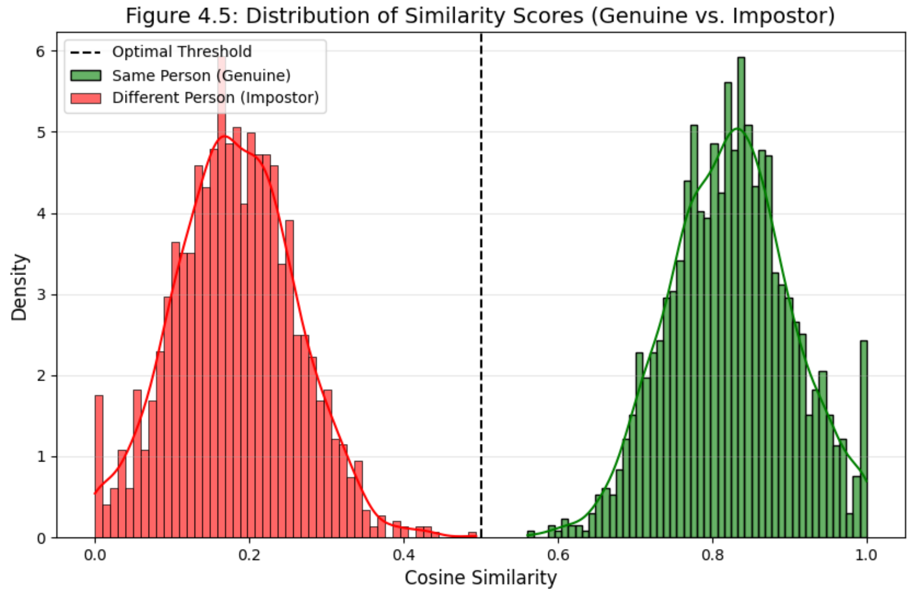

Figure 6. Distribution of genuine and impostor cosine-similarity scores.

4.5 Ablation Study

Configuration	ROC-AUC	EER	Precision	F1-score
ResNet-18 + Standard Adam	0.9152	11.24%	0.884	0.881
ResNet-18 + SE Blocks + Adam	0.9581	7.52%	0.921	0.918
ESNN + Attn-Adam	0.9892	4.20%	0.958	0.958

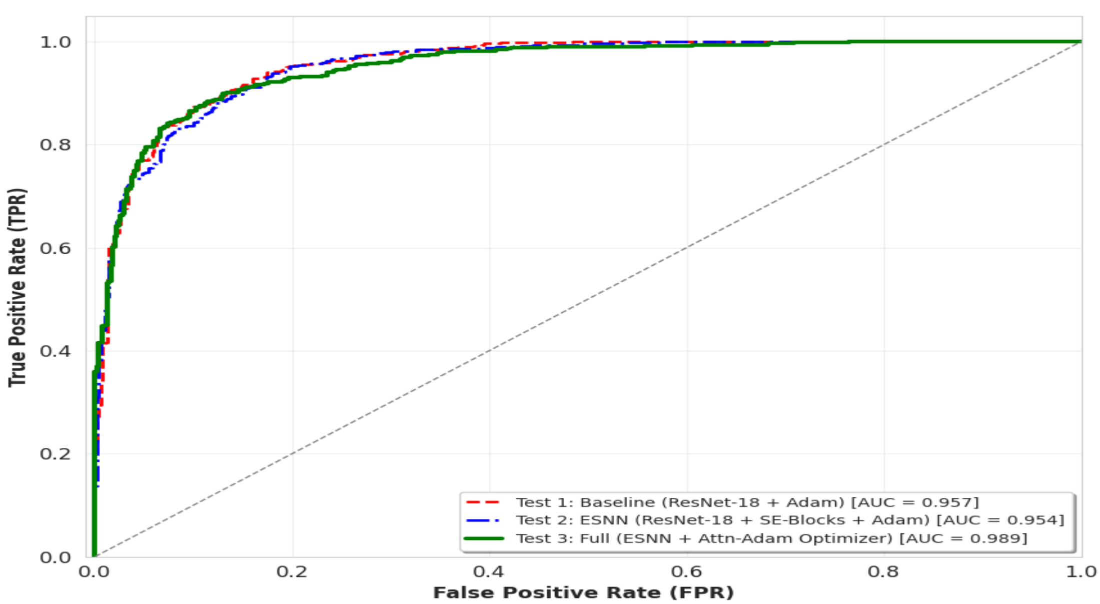

Figure 7. Ablation comparison of the baseline, SE-enhanced and full ESNN + Attn-Adam configurations.

V. DISCUSSION

The principal finding is that the proposed ESNN with Attn-Adam materially outperformed the baseline configurations in the reported experiments. The progression from AUC 0.9152 to 0.9581 and finally 0.9892 indicates that both architectural and optimization changes were associated with improved discrimination. The decrease in EER from 11.24% to 7.52% after SE integration suggests that channel-wise feature recalibration helped the model produce more separable representations. The further reduction to 4.20% after introducing Attn-Adam supports the study's hypothesis that optimization stability can affect the quality of learned metric embeddings.

The use of SE blocks is consistent with the broader principle of attention-based feature selection: informative channels can be amplified while less useful responses can be suppressed. In the context of unconstrained face images, this can help reduce the influence of background and environmental variation. The temporal gradient attention mechanism addresses a different stage of the learning pipeline. Instead of changing the input representation directly, it changes how historical optimization information is weighted. The observed improvement in the ablation study is therefore consistent with the proposed division of responsibility between architectural attention and optimization attention.

The mobile security architecture provides a second layer of protection. The neural model is used for face verification, whereas the fingerprint stage is handled by Android's biometric security framework. The cryptographic binding design means that successful application-level face verification alone is insufficient to release the protected wallet operation; the secondary biometric step is required. This is important because biometric templates themselves should not be treated as ordinary application data. The thesis therefore separates learned facial representation from hardware-backed fingerprint processing.

The results should nevertheless be interpreted within the scope of the evaluation. LFW is a public face-recognition benchmark rather than a dedicated mobile-wallet biometric dataset. The fingerprint stage is hardware-backed and is not trained jointly with the facial network. Consequently, the reported AUC, EER, precision, recall and F1 values primarily quantify the facial verification model and its associated verification decision experiments. They should not be interpreted as a complete empirical security evaluation of all possible face-fingerprint spoofing or financial fraud scenarios.

The study also contains threshold values associated with different analyses: 0.6421 is reported as the principal optimal threshold distance, 0.55 appears in the implementation flow as a face-verification gate, 0.37 is reported in the confusion-matrix analysis, and 0.50 is shown in the cosine-similarity distribution. Because the thesis does not fully document the transformation between Euclidean distance and cosine similarity or explain whether these thresholds belong to distinct experiments, the journal version treats the metrics as reported and recommends a unified threshold-calibration protocol in follow-up work. This transparency is preferable to presenting the values as if they were directly interchangeable.

VI. LIMITATIONS AND FUTURE WORK

Several limitations define the interpretation of the results. First, the principal empirical dataset is LFW, which may not represent the full range of demographic, device, illumination and operational conditions encountered in Nigerian mobile financial applications. Second, the reported experiments focus on facial verification, while fingerprint authentication is integrated through the Android hardware security layer rather than evaluated as a learned fingerprint classifier. Third, the manuscript does not report a dedicated presentation-attack detection experiment involving printed photographs, replay attacks, deepfakes or masks. Fourth, the thesis reports several operating thresholds that require formal reconciliation before deployment. Fifth, the reported results do not include repeated-run confidence intervals or statistical significance tests.

Future work should therefore evaluate the complete face-fingerprint workflow using a controlled multimodal dataset and real mobile devices. Future experiments should include presentation-attack detection, deepfake and replay testing, demographic subgroup analysis, cross-device validation, latency and energy measurements, and repeated independent training runs. Additional metric-learning losses such as angular-margin objectives could also be investigated, as suggested in the thesis discussion. Finally, the Attention-Adam optimizer should be compared with contemporary optimizers under matched training budgets and evaluated through sensitivity analysis over its gradient-window and attention parameters.

VII. CONCLUSION

This study presented an enhanced Siamese Neural Network with a novel Attention-Adam optimization mechanism for bimodal biometric authentication in mobile financial systems. The facial verification component combines a ResNet-18 backbone with Squeeze-and-Excitation blocks and produces compact 128-dimensional embeddings. The proposed Attn-Adam optimizer applies temporal attention to recent gradients before the AdamW-style parameter update. The resulting model is integrated sequentially with hardware-backed fingerprint authentication and cryptographic key protection on Android.

Under the reported LFW evaluation, the final system achieved AUC = 0.9892, EER = 4.20%, precision = 95.86%, recall = 95.74% and F1 = 95.80%. The ablation results show progressive improvement from the baseline ResNet-18 + Adam model through the SE-enhanced model to the complete ESNN + Attn-Adam system. The reported approximately 62.6% reduction in EER relative to the baseline provides the principal empirical evidence supporting the proposed optimization strategy. The work therefore contributes an architecture-level and optimization-level approach to improving deep metric learning for mobile biometric verification, while also demonstrating a practical security pathway through serial face-fingerprint authentication and hardware-backed cryptographic protection.

Declarations

Data availability. The study used the Labeled Faces in the Wild benchmark as described in the thesis. No new public dataset is claimed to have been created by this study.

Ethics statement. The reported experiments rely principally on a public benchmark dataset. The manuscript does not claim a new human-subject data-collection study.

Conflict of interest. No conflict of interest is reported in the source thesis.

Funding. No external funding information is reported in the source thesis.

REFERENCES

[1] Abderrahmane, H., Noubeil, G., Lahcene, Z., Akhtar, Z., & Dasgupta, D. (2020). Weighted quasi-arithmetic mean based score level fusion for multi-biometric systems. *IET Biometrics, 9*(3), 91–99. Wiley

[2] Akhtar, Z., Micheloni, C., & Foresti, G. L. (2015). Biometric liveness detection: Challenges and research opportunities. *IEEE Security & Privacy, 13*(5), 63–72. IEEE

[3] Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(8), 1798–1828. IEEE

[4] Bhavani, S. A., & Karthikeyan, C. (2025). An attention based deep learning with effective SVM-ConvFaceNeXt model for face recognition in unconstrained environment. *Signal, Image and Video Processing, 19*(12). Springer

[5] Bromley, J., et al. (1993). Signature verification using a "Siamese" time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence, 7*(4), 669–688. World Scientific

[6] Chopra, S., Hadsell, R., & LeCun, Y. (2005). Learning a similarity metric discriminatively, with application to face verification. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE

[7] Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. *arXiv:1801.07698*. arXiv

[8] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

[9] Hadsell, R., Chopra, S., & LeCun, Y. (2006). Dimensionality reduction by learning an invariant mapping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE

[10] Hall, D. L., & Llinas, J. (1997). An introduction to multisensor data fusion. *Proceedings of the IEEE, 85*(1), 6–23. IEEE

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv:1512.03385*. arXiv

[12] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv:1704.04861*. arXiv

[13] Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H., & Kalenichenko, D. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE

[14] Jain, A. K., Nandakumar, K., & Nagar, A. (2008). Biometric template security. *EURASIP Journal on Advances in Signal Processing, 2008*, 579416. Springer

[15] Juels, A., & Sudan, M. (2006). A fuzzy vault scheme. *Designs, Codes and Cryptography, 38*(2), 237–257. Springer

[16] Koch, G., Zemel, R., & Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition. Carnegie Mellon University

[17] Kumar, C. R., N., S., Priyadharshini, M., E., D. G., & M., K. R. (2023). Face recognition using CNN and siamese network. *Measurement: Sensors, 27*, 100800. ScienceDirect

[18] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444. Nature

[19] Liu, W., & Pan, Y. (2024). Spatio-temporal-based action face anti-spoofing detection via fusing dynamics and texture face keypoints cues. *IEEE Transactions on Consumer Electronics, 70*(1), 2401–2413. IEEE

[20] Ma, H., Guo, D., Wang, J., Li, P., Xu, C., & Wang, Y. (2025). Real-time face recognition algorithm based on lightweight neural network in the field of computer vision. *Discover Applied Sciences*.

[21] Maltoni, D., Maio, D., Jain, A. K., & Prabhakar, S. (2009). *Handbook of fingerprint recognition*. Springer. Springer

[22] Naji Ali, H., & Mahmood Al-Dabbagh, S. S. (2026). A systematic literature review on biometric authentication in mobile banking. *F1000Research, 15*, 5. F1000Research

[23] Ratha, N. K., Chikkerur, S., Connell, J. H., & Bolle, R. M. (2007). Generating cancelable fingerprint templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(4), 561–572. IEEE

[24] Ratha, N. K., Connell, J. H., & Bolle, R. M. (2001). Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal, 40*(3), 614–634. IBM Research

[25] Ross, A., & Jain, A. K. (2003). Information fusion in biometrics. *Pattern Recognition Letters, 24*, 2115–2125. ScienceDirect

[26] Ross, A., & Jain, A. K. (2004). Multimodal biometrics: An overview. In *Proceedings of the 2004 International Symposium on Electronics in the 21st Century*.

[27] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4510–4520).

[28] Setyawan, N., Sun, C.-C., Hsu, M.-H., Kuo, W.-K., & Hsieh, J.-W. (2025). FaceLiVT: Face recognition using linear vision transformer with structural reparameterization for mobile device. In *2025 IEEE International Conference on Image Processing* (pp. 1720–1725). IEEE

[29] Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE

[30] U. Sumalatha, K., Krishna Prakasha, K., Prabhu, S., & Nayak, V. C. (2024). A comprehensive review of unimodal and multimodal fingerprint biometric authentication systems: Fusion, attacks, and template protection. *IEEE Access*. IEEE

[31] Warden, P., & Situnayake, D. (2019). *TinyML: Machine learning with TensorFlow Lite on Arduino and ultra-low-power microcontrollers*. O'Reilly Media.

[32] Yuan, B., Du, C.-Q., & Li, Z.-T. (2026). Mobi-FaceNeXt: Research on an efficient face recognition algorithm based on lightweight convolutional neural networks. *Thermal Science, 30*(2 Part A), 1191–1201. Crossref

[33] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters, 23*(10), 1499–1503. IEEE