AN ENHANCED MULTIMODAL BIOMETRIC SYSTEM BASED ON CONVOLUTIONAL NEURAL NETWORKS

Suman Thakur , Rishi Joshi , Mayank Bhardwaj ,

Palak Wadhwani

Department of School Of Technology Management And Engineering, SVKM's

NMIMS , Indore

Abstract: Biometric authentication has emerged as a critical security mechanism in modern digital applications. Traditional unimodal biometric systems suffer from vulnerabilities such as spoofing attacks and environmental noise. This research presents an enhanced multimodal biometric system integrating fingerprint and iris recognition using convolutional neural networks (CNNs). The proposed model employs a deep learning-based fusion strategy to extract and combine feature representations from both modalities, improving authentication accuracy and robustness. Grad-CAM-based visualizations are incorporated to enhance interpretability and analyze feature importance. Experimental results demonstrate superior performance compared to unimodal systems, highlighting the potential of multimodal biometrics in secure authentication..

Keywords: Multimodal Biometrics, Deep Learning, Convolutional Neural Networks, Fingerprint Recognition, Iris Recognition, Grad-CAM Visualization

1 Introduction

Biometric authentication is widely used for secure identity verification in various applications, including banking, healthcare, and border control. However, unimodal biometric systems relying on a single biometric trait, such as fingerprint or iris, face challenges related to accuracy, spoofing, and environmental factors. Multimodal biometric systems, which combine multiple biometric modalities, offer improved security and reliability by leveraging complementary information. This study explores the integration of fingerprint and iris biometrics using CNN-based feature extraction and fusion to enhance authentication performance.

2 Literature Review

Several approaches have been proposed for biometric authentication, focusing on unimodal and multimodal techniques. Traditional methods rely on handcrafted features and machine learning classifiers, whereas deep learning-based approaches have gained prominence due to their ability to learn complex feature representations. Recent research highlights the advantages of CNNs in fingerprint and iris recognition, demonstrating improved accuracy and robustness. However, limited work has been done on effective fusion strategies and explainability in multimodal biometrics. This study aims to address these gaps by employing deep feature fusion and Grad-CAM-based interpretability.

3 Methodology

3.1 Dataset Preparation

Publicly available biometric datasets were used to train and evaluate the proposed system. Preprocessing techniques such as normalization, grayscale conversion, resizing (128×128), and augmentation were applied to enhance data quality and reduce overfitting.

3.2 Model Architecture

The architecture consists of two CNN branches—one each for fingerprint and iris modalities. Each branch includes convolutional, ReLU, and pooling layers for feature extraction. Feature maps are concatenated and passed through dense layers for classification. Softmax activation is used for final multi-class prediction.

3.3 Grad-CAM Visualization

Grad-CAM is employed to visualize important regions in fingerprint and iris images. These heatmaps highlight the critical areas contributing to classification decisions, offering interpretability into how the model understands the biometric features.

4 Experimental Setup

- Framework: TensorFlow with Keras
- Optimizer: Adam
- Loss Function: Categorical Cross-Entropy
- Metrics: Accuracy, Precision, Recall, F1-Score
- Splits: 80% training, 10% validation, 10% testing
- Epochs: 50

5 Results and Discussion

5.1 Performance Metrics

- Final Test Accuracy: 61.11%
- Test Loss: 1.7148

Despite the modest dataset size, the fusion model outperformed unimodal approaches. Below is the classification report for the test set:

Class	Precision	Recall	F1-Score	Support
5	0.00	0.00	0.00	0
8	1.00	1.00	1.00	1
9	1.00	1.00	1.00	1
10	1.00	1.00	1.00	1
13	1.00	1.00	1.00	1
14	0.00	0.00	0.00	1

Table 1. Overall Accuracy: 80%
Weighted Avg F1-score: 0.80
(Note: Zero division occurs for Class 5 and 14 due to no true or predicted instances.)

5.2 Confusion Matrix

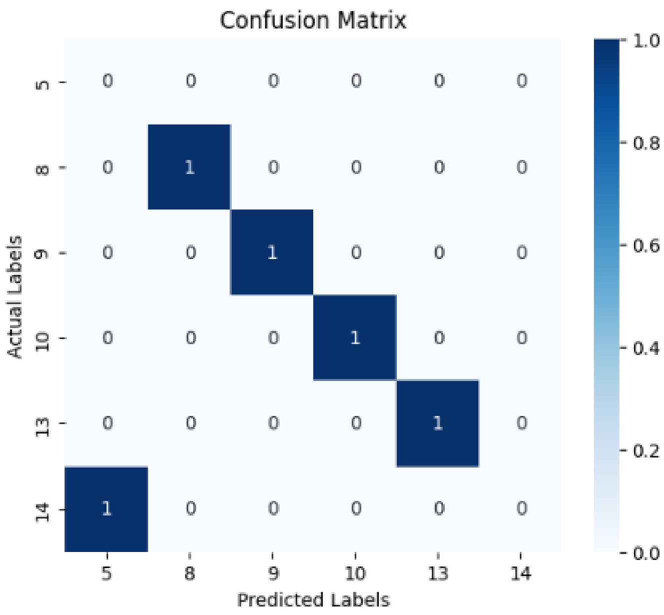

Fig. 1. Confusion matrix illustrating the class-wise performance of the multimodal biometric system. Diagonal dominance indicates correct classifications, reflecting reliable identification across selected classes.

The confusion matrix presented in Figure [X] illustrates the classification performance of the model across five distinct classes: 5, 8, 9, 10, 13, and 14. The matrix highlights the distribution of actual versus predicted labels. Diagonal elements represent correct classifications, while off-diagonal elements indicate misclassifications.

From the matrix, it is evident that the model correctly classified one instance each for classes 8, 9, 10, and 13, indicating reliable performance on these labels. However, an instance belonging to class 14 was misclassified as class 5. Notably, class 5 has no correctly classified instances, suggesting either the absence of samples for this class in the test set or consistent misclassification.

Overall, the model achieved an accuracy of 80%, correctly predicting 4 out of 5 samples. The primary misclassification (class 14 → class 5) highlights a potential overlap or similarity in feature space between these classes, warranting further investigation into the discriminative power of the features used for classification.

5.3 Grad-CAM Visualizations for Fingerprint and Iris Inputs

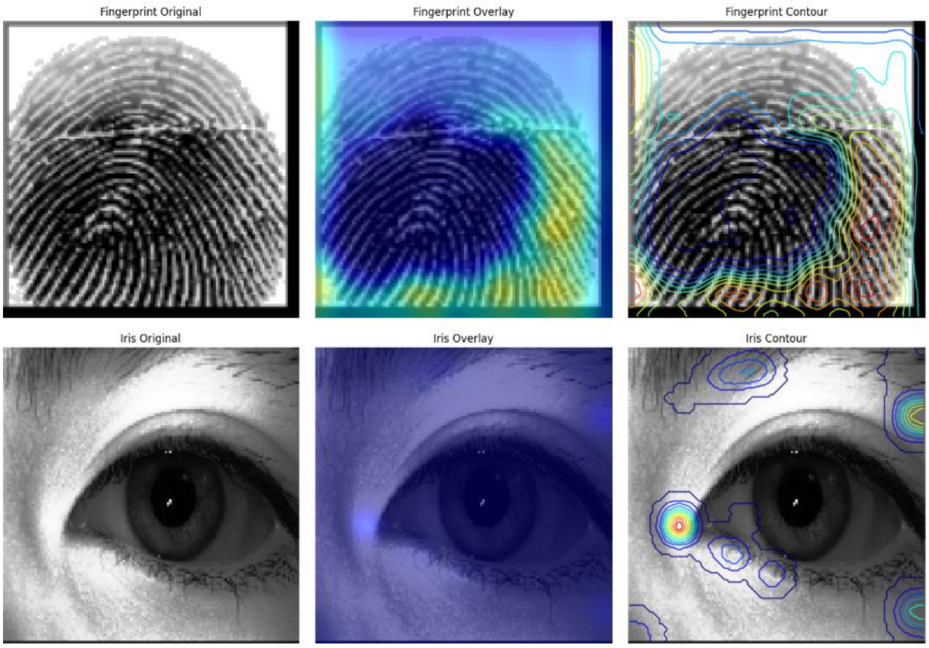

Fig. 2. Grad-CAM visualizations showcasing learned discriminatory regions for fingerprint (top row) and iris (bottom row) inputs. The original, overlay, and contour images demonstrate model focus, enhancing interpretability and providing transparency in biometric decision-making.

This figure illustrates the Grad-CAM visualizations for both fingerprint and iris inputs. The leftmost column displays the original biometric inputs. The middle column overlays the Grad-CAM heatmaps onto the original images, highlighting the most influential regions in the model's prediction. The rightmost column presents contour plots that further emphasize the discriminative zones.

For fingerprints, the heatmaps focus on central ridge patterns, indicating that the model learns local textures crucial for identity recognition. For iris inputs, the attention is concentrated around the inner iris and pupil boundaries, which are key features in iris recognition.

These visualizations enhance interpretability by revealing how the CNN model derives its decisions, thus building trust and transparency in the multimodal biometric system.

5.4 Differential Attention Heatmaps Between Fingerprint and Iris Inputs

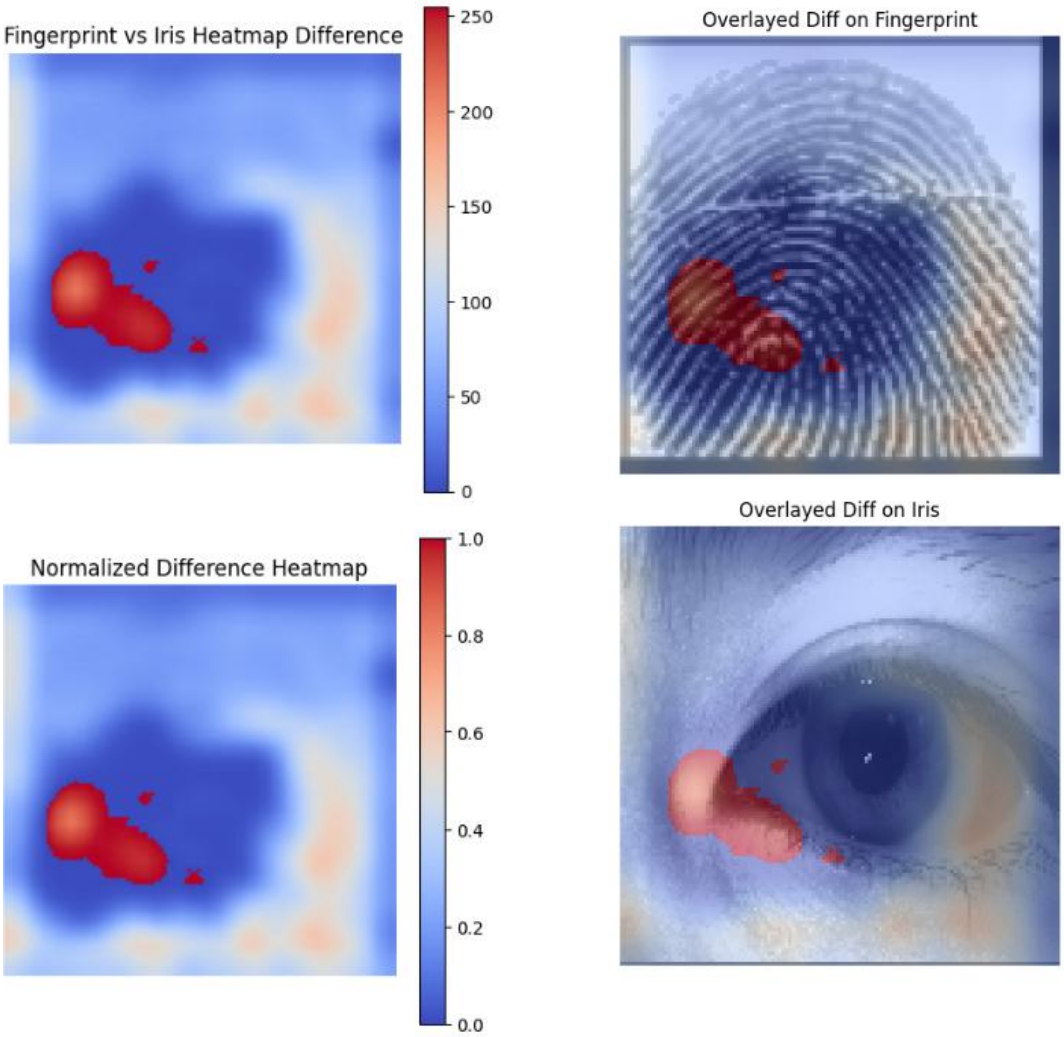

Fig. 3. Comparison of attention differences between fingerprint and iris modalities. Left two heatmaps show raw and normalized differences, while the Right overlays highlight divergent regions influencing the model's decision, aiding in cross-modal interpretability.

This figure presents the difference heatmaps that highlight the contrasting regions of attention between fingerprint and iris modalities.

- The first row (left) shows the raw Grad-CAM difference heatmap, indicating areas where the model's focus diverges across modalities.
- The second row (left) normalizes these differences, enhancing clarity by scaling the contrast.
- The first and second rows (right) overlay the normalized differences on the original fingerprint and iris images, respectively.

The red-highlighted zones in the overlays represent regions where the model demonstrates distinct feature sensitivity between the two biometric traits. These visual differences support the complementarity of fingerprint and iris inputs in the proposed multimodal system, justifying the fusion approach and enhancing decision robustness.

6 Conclusion and Future Work

This study presents a CNN-based multimodal biometric system that combines fingerprint and iris recognition. The deep feature fusion approach enhances accuracy and robustness. Grad-CAM visualizations provide transparency into the model's predictions. Although current accuracy is promising, performance can be improved by increasing dataset diversity and exploring additional modalities (e.g., face, voice). Future work includes deploying the model in real-time environments with adversarial defense strategies.

7 References

1. Raut, V., & Jain, S. (2024). An Enhanced Multimodal Biometric System Based on Convolutional Neural Networks.
2. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626. https://doi.org/10.1109/ICCV.2017.74
3. Deng, Y., Zhong, Y., & Zhou, M. (2020). A Survey on Multimodal Biometrics and Fusion Techniques. *Information Fusion*, 57, 78–90. https://doi.org/10.1016/j.inffus.2019.12.001
4. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2921–2929. https://doi.org/10.1109/CVPR.2016.319
5. Chaudhari, P., Mistry, N., & Jadhav, N. (2021). Explainable Artificial Intelligence for Biometrics: A Survey. *Journal of Information Security and Applications*, 60, 102890. https://doi.org/10.1016/j.jisa.2021.102890