

A Hybrid CNN-CBAM-Vision Transformer Framework for Fingerprint-Based ABO and Rh Blood Group Classification

Rajdeep Moulik¹, Protyusha Mondal², *Bijoyini Bagchi³, Rahul De⁴, Pritesh Patra⁵, and Riddhi Banerjee⁶

Department of Computer Science and Engineering,
Institute of Engineering and Management, Kolkata, India
Rajdeep.Moulik2023@iem.edu.in, Protyusha.Mondal2023@iem.edu.in,
*Bijoyini.Bagchi@iem.edu.in, Rahul.De2023@iem.edu.in,
Pritesh.Patra2023@iem.edu.in, Riddhi.Banerjee2023@iem.edu.in

Abstract. Fingerprint dermatoglyphics show that there is a latent ABO and Rh blood group correlation. This paper suggests a hybrid deep learning model utilizing a truncated Convolutional Neural Network (CNN), Convolutional Block Attention Module (CBAM) and Vision Transformer (ViT) to classify 8 blood groups. The model leverages local and global feature learning. Before classification, robust representation is created using segmentation, binarization, and minutiae detection. The best precision in classification of 96.25 percent is obtained with the proposed model, outperforming standalone CNN and ViT architectures.

Keywords: Deep Learning · Convolutional Neural Networks · Vision Transformer · CBAM · Fingerprint Classification · Blood Group Detection

1 Introduction

Recent works have used architectures like ResNet50 and custom CNNs to classify blood groups by fingerprint. However, CNNs have not been able to capture long-range spatial dependencies. Vision Transformers (ViT) address this global spatial relation but struggle to identify micro-level textures. Explicit attention gating which includes Convolutional Block Attention Module (CBAM) has yet to be introduced to unify CNNs and ViTs for this domain.

We identify key research gaps in existing biometric blood profiling: modern approaches are limited to either CNN or ViT architecture, they are highly vulnerable to sensor noise, and highly accurate models possess computational costs that prevent edge deployment. This paper introduces a single, lightweight hybrid architecture for local feature extraction, noise removal, and global topology modeling.

2 Methodology

2.1 Data and Preprocessing

An open-source dataset from Kaggle, containing digital fingerprints specially designed for ABO and Rh phenotypes, was utilized. The dataset comprises 4,000 enhanced images divided into a standard 80% training, 10% validation, and 10% testing split. Preprocessing ensures input quality through a unified pipeline: ridge segmentation, histogram equalization, adaptive binarization, thinning, and minutiae detection (bifurcations and endings) to form a compact feature vector.

2.2 Proposed Custom Hybrid Architecture

The framework manipulates micro-level and macro-level data via three main components.

Truncated CNN A truncated CNN extracts initial local features, such as minutiae and complex structural textures. $Y = W * X + b$

Attention Gating Mechanism (CBAM) A Channel and Spatial Attention mechanism masks non-informative regions to filter background noise and focus on discriminative ridge patterns.

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))$$

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)]))$$

Vision Transformer (ViT) Blocks Attention-gated maps are flattened and passed through ViT blocks for processing long-range dependencies.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

A final MLP classification head with a softmax activation processes the high-dimensional tokens to output specific probabilities for the 8 classes:

$$P(y_i) = \frac{e^{z_i}}{\sum e^{z_j}}$$

The network is optimized using the Cross-Entropy Loss function:

$$L = - \sum y_i \log(\hat{y}_i)$$

2.3 Experimental Setup

The framework was implemented in TensorFlow/Keras and trained on Google Colab (NVIDIA T4 GPU). Inputs were standardized to $224 \times 224 \times 3$ pixels. Optimization was performed using Adam (10^{-4} , batch size=32) with a 15-epoch early stopping patience. Online data augmentation included random rotations, spatial shifts, and shear transformations.



Fig. 1. Proposed workflow for fingerprint-based blood group classification.

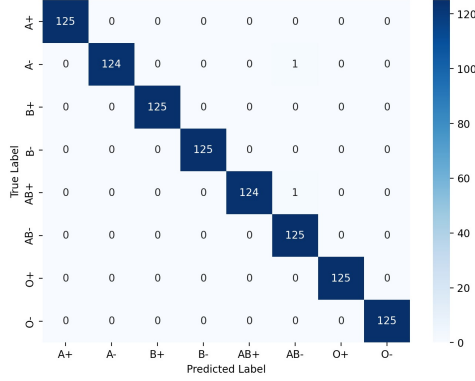


Fig. 2. Confusion matrix for 8-class prediction.

3 Results and Analysis

The proposed architecture achieved a maximum accuracy of 96.25%, with a precision of 96.33% and a macro-average recall of 96.25%. Comparative ablation studies show that combining attention-gated local features (CBAM) with global self-attention (ViT) successfully solves the non-linear mapping between fingerprint minutiae and blood types.

The confusion matrix analysis (Fig. 2) showed that the proposed model reduced inter-class misclassifications, as the non-informative background data was forcefully blocked by the CBAM module. The remaining 3.75% error rate is primarily attributed to misaligned fingerprint samples.

As detailed in Table 1, our model is significantly lighter than state-of-the-art alternatives. It resides in the ideal region of lightweight models, offering

Table 1. Performance and Complexity Comparison

Model Architecture	Acc (%)	Params (M)	Inference
Standalone ViT [1]	97.00	86.0	Slow
Heavy CNN [2]	99.00	150.0	Very Slow
Proposed Hybrid	96.25	4.2	Fast (45ms)

high accuracy and low complexity suitable for edge deployment. It possesses fewer parameters (4.2 million), allowing for an average inference time of just 45 milliseconds per image.

4 Conclusion

This framework demonstrates the feasibility of a CNN-CBAM-ViT model to properly identify ABO and Rh blood types using fingerprint dermatoglyphics. By achieving 96.25% accuracy with only 4.2 million parameters, the model is efficient enough for real-time deployment on portable fingerprint scanners. Future work will focus on improving explainability using tools like Grad-CAM to isolate exact ridge markers, and migrating the architecture to low-power IoT healthcare devices for rural medical triage.

References

1. P. Vijaykumar, N. Patil, and D. R. Ingle, "Blood group detection using vision transformer," *Int J Recent Adv Eng Sci Technol*, vol. 13, no. 1, pp. 442-449, 2025.
2. S. Yashvanth et al., "Innovative blood group detection through image processing and fingerprint recognition," *ResearchGate*, 2025.
3. A. Sultan et al., "Explainable AI (XAI) using Grad-CAM for fingerprint region interpretation," *Int J Creat Res Thoughts*, vol. 12, no. 4, 2024.
4. H. Alshehri et al., "Systematic review of deep learning advancements in fingerprint dermatoglyphics for non-invasive blood typing," *MDPI Appl Sci*, vol. 14, no. 12, 2024.

Biographies

Rajdeep Moulik is a B.Tech. student at the Institute of Engineering and Management, Kolkata, India, with research interests in machine learning and pattern recognition.

Protyusha Mondal is a B.Tech. student at the Institute of Engineering and Management, Kolkata, India, interested in deep learning and computer vision.

Bijoyini Bagchi is an Assistant Professor at the Institute of Engineering and Management, Kolkata, India. Her research interests include artificial intelligence and intelligent healthcare systems.