Fingerprint Biometrics: A Comparative Study of Deep Learning Models (CNNs vs. ViTs)

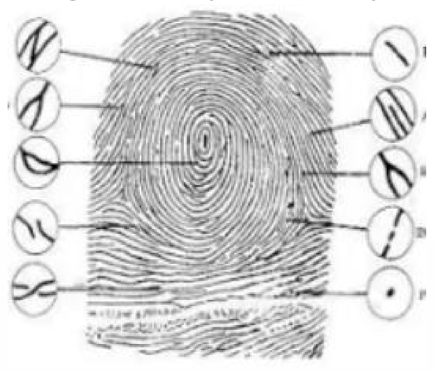
AI 2025
SUMMER SCHOOL
ai.uni-jena.de

KEMKUINI FOUODJI Gertrude Ketsia
Faculty of Mathematics and Computerscience
Friedrich Schiller Universität Jena

INTRODUCTION

Fingerprint recognition is a widely adopted biometric method due to its uniqueness and ease of acquisition. While traditional systems rely on minutiae and ridge analysis, they struggle with low-quality or partial prints—especially in children.

Deep learning offers a robust alternative by learning features directly from fingerprint images. This work compares three prominent approaches: **Convolutional Neural Networks (CNNs)** and **Vision Transformers (ViTs)**.



APPROACHES

Convolutional Neural Network

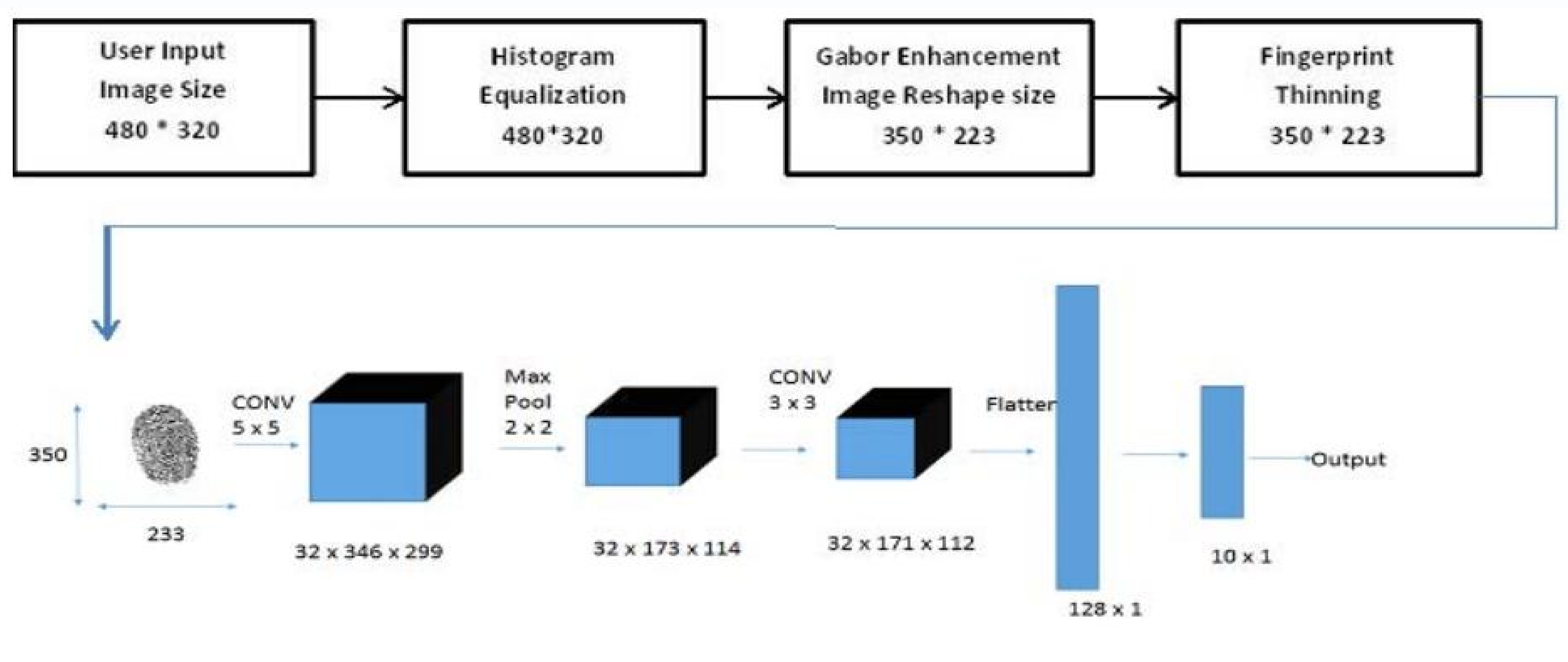
Feature Extraction:

- Input: Preprocessed fingerprint image (e.g., Gabor-filtered, normalized).
- Convolutional layers detect hierarchical features:

  Layer 1: Ridges/valleys → Layer 2: Minutiae → Layer 3: Macro-patterns (whorls, loops).
- Flattened features → fed into fully connected (FC) layers.

Classification/Verification:

- Closed-Set Identification (1:N matching):Softmax output layer predicts class probabilities (e.g., "User 1" vs. "User 2").
- Open-Set Verification (1:1 matching): Embedding vector (e.g., 128-D) compared via cosine distance.



Vision Transformers

Patch Embedding & Linear Projection
Input: Fingerprint image (e.g., 224×224 pixels).

- Split into fixed-size patches (e.g., 16×16 → 196 patches).
- Flatten patches → Linear projection into patch embeddings (e.g., 768-D).
- Add positional embeddings to retain spatial information.

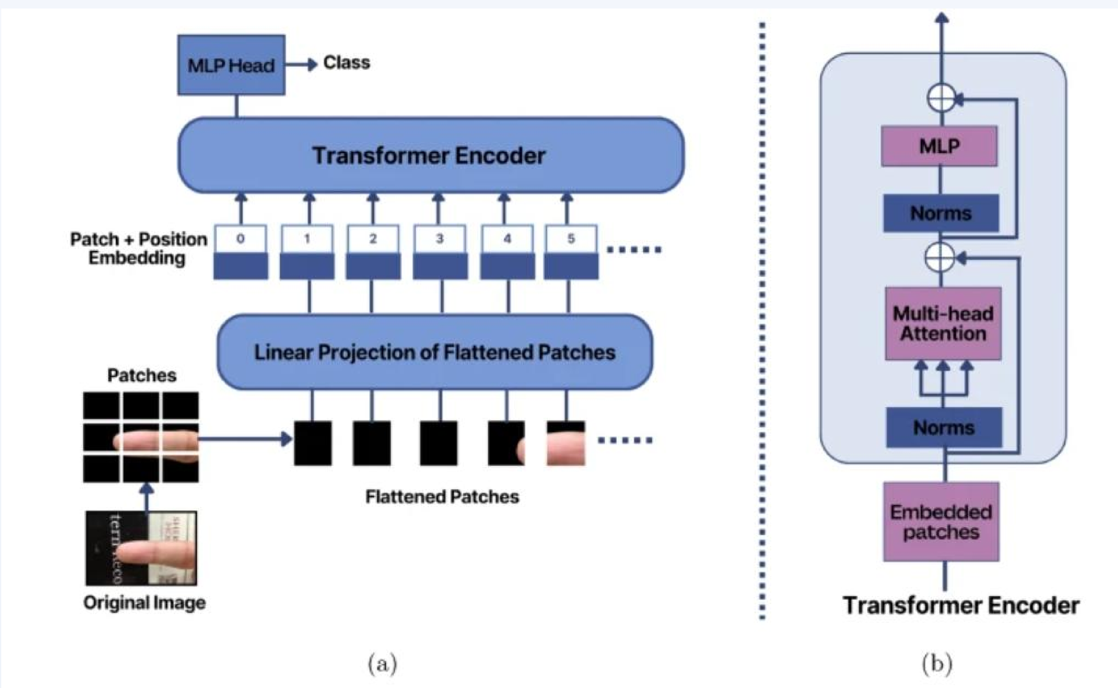
Transformer Encoder Layers

- Multi-Head Self-Attention (MSA):
  Computes global relationships between patches (e.g., links ridge breaks across distant regions).
  Heads: Parallel attention mechanisms focus on different features (minutiae,orientation).
- Layer Normalization (Norm): Stabilizes training.
- MLP Block: Non-linear transformation (GeLU activation).

Residual Connections: Skip links around each layer for gradient flow.

Classification Head

- [CLS] Token: A specialtoken (prepended to patch sequence) aggregates global fingerprint features.
- MLP Head: Final layers map the [CLS] token embedding to class probabilities (softmax)

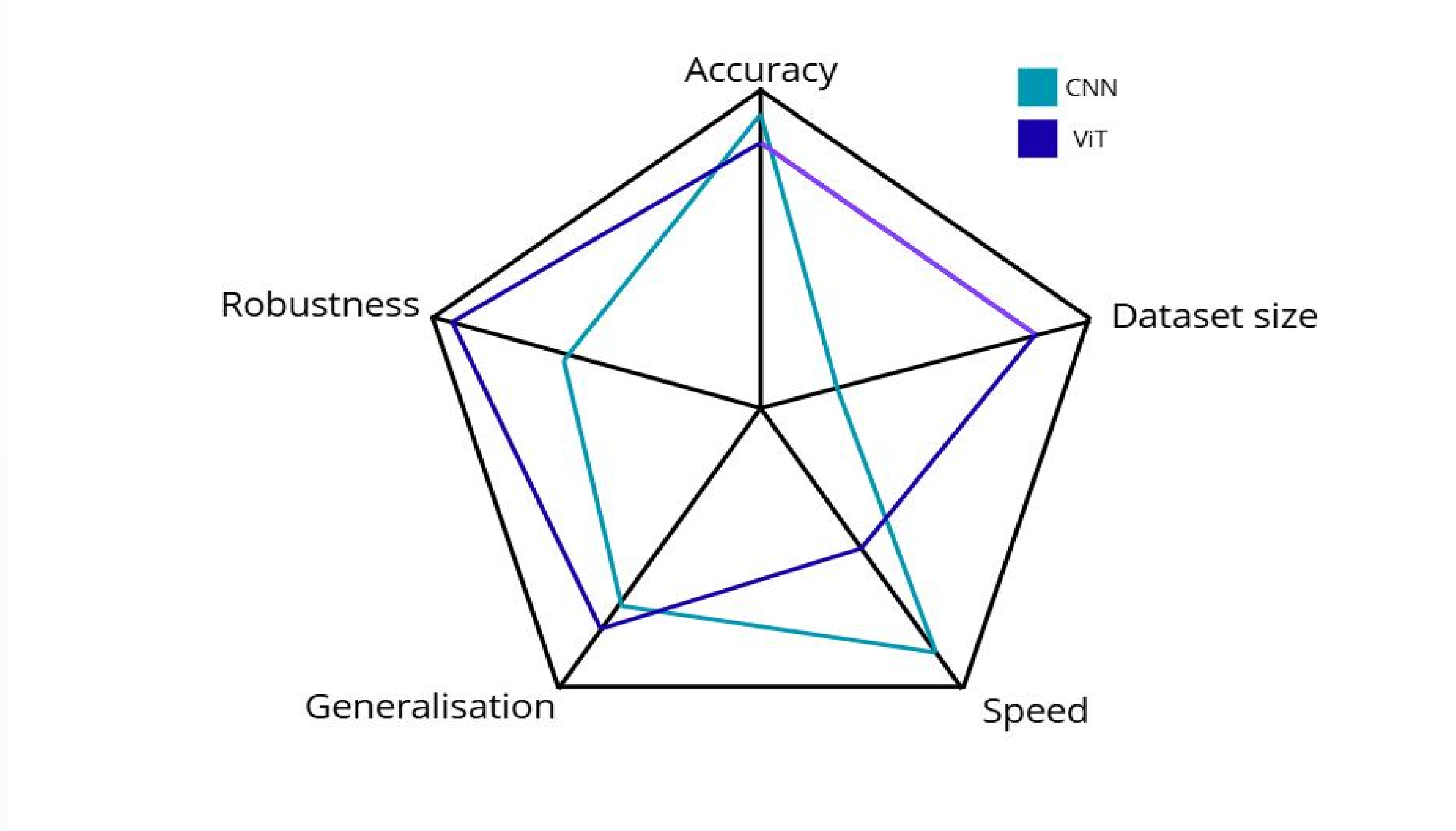


Architecture of the proposed model for contactless fingerprint classification. (a) Structure of adapted ViT model including embedding, transformer encoder, and classification head; (b) Detailed structure of the transformer encoder (*model adapted from* [5, 43])

RESULTS

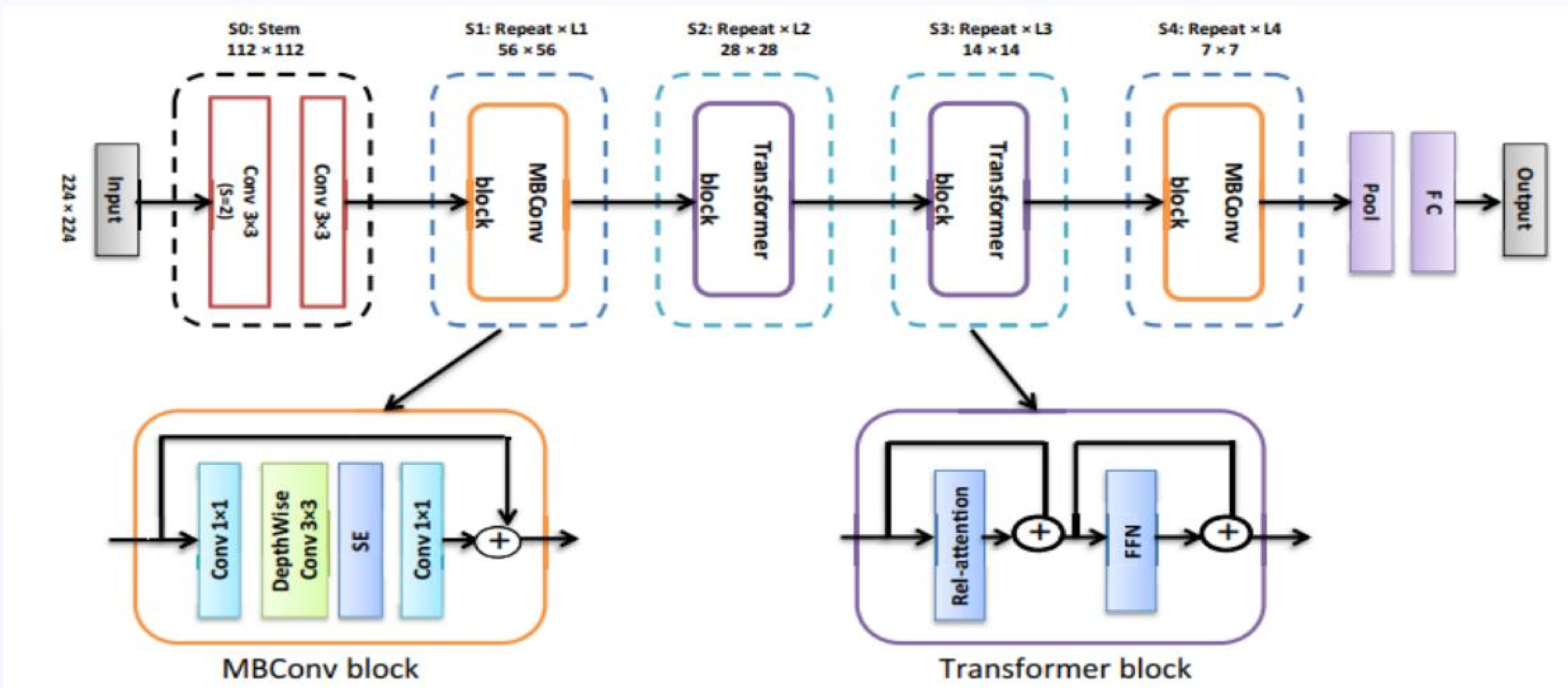
Although the two studies used different datasets and acquisition methods—contact-based in Pandya et al. and contactless in Kaplesh et al.—their results provide valuable insights into how different deep learning architectures perform under varying conditions. Building on these findings, the following section details the modeling approaches considered in our study.

Study & Model	**Pandya et al., 2018 – CNN**	**Kaplesh et al., 2025 – Vision Transformer (ViT-Base)**
Dataset Used	Custom fingerprint dataset (Futronics FS88 scanner, 280 samples, 56 classes)	ISPFDv1 (4096 images, 64 users, contactless)
Input Type	Contact-based fingerprints	Contactless fingerphoto images
Reported Accuracy	**98.21%**	**96%**



CONCLUSIONS

CNNs excel in controlled, contact-based settings, while Vision Transformers show strong robustness in unconstrained, contactless scenarios.
Combining CNNs' local feature power with Transformers' global context, and adopting lightweight models like MobileNet or EfficientNet, could bridge current gaps and deliver both accuracy and efficiency.



REFERENCES

- B. Pandya, G. Cosma, A. A. Alani, A. Taherkhani, V. Bharadi and T. M. McGinnity, "Fingerprint classification using a deep convolutional neural network," *2018 4th International Conference on Information Management (ICIM)*, Oxford, UK, 2018, pp. 86-91, doi: 10.1109/INFOMAN.2018.8392815.
- Kaplesh, P., Gupta, A., Bansal, D. *et al.* Vision transformer for contactless fingerprint classification. *Multimed Tools Appl* **84**, 31239–31259 (2025). https://doi.org/10.1007/s11042-024-20396-4
- *Sif Eddine Boudjellal, Naceur-Eddine Boukezzoula, Abdelwahhab Boudjelal. Hybrid convolutional transformer-based network model for finger vein identification. Journal of Electronic Imaging, 2023, ff10.1117/1.JEI.32.5.053005ff. ffhal-04980806f*

CONTACT

Ketsiakemkuini@gmail.com