Bi-Encoder Contrastive Learning for Fingerprint and Iris Biometrics

Matthew So
Department of Computer Science
Columbia University
ms5513@columbia.edu

Judah Goldfeder
Department of Computer Science
Columbia University
jag2396@columbia.edu

Mark Lis
College of Medicine
SUNY Downstate Health Sciences University
Mark.Lis@downstate.edu

Hod Lipson
Department of Mechanical Engineering
Columbia University
hod.lipson@columbia.edu

Abstract

There has been a historic assumption that the biometrics of an individual are statistically uncorrelated. We test this assumption by training Bi-Encoder networks on three verification tasks, including fingerprint-to-fingerprint matching, iris-to-iris matching, and cross-modal fingerprint-to-iris matching using 274 subjects with $\sim$100k fingerprints and 7k iris images. We trained ResNet-50 and Vision Transformer backbones in Bi-Encoder architectures such that the contrastive loss between images sampled from the same individual is minimized. The iris ResNet architecture reaches 91 ROC AUC score for iris-to-iris matching, providing clear evidence that the left and right irises of an individual are correlated. Fingerprint models reproduce the positive intra-subject suggested by prior work in this space. This is the first work attempting to use Vision Transformers for this matching. Cross-modal matching rises only slightly above chance, which suggests that more data and a more sophisticated pipeline is needed to obtain compelling results. These findings continue challenge independence assumptions of biometrics and we plan to extend this work to other biometrics in the future. Code available: https://github.com/MatthewSo/bio_fingerprints_iris

1 Introduction

Biometric traits such as fingerprints, voice, iris texture, hand geometry, and face have a history of being assumed to be statistically independent when sampled from a single individual. Even modern security systems and forensic workflows explicitly encode this assumption in their decision rules and probabilistic models (Daugman, 2004; Hollingsworth et al., 2011; Risch and Devlin, 1992). Recent evidence, however, questions this assumption. Guo *et al.* reported statistically significant intra-subject patterns in fingerprints and proposed a ResNet-style architecture that predicts same subject fingerprint pairs with better than random chance (Guo et al., 2024). This result has generated substantial interest across the biometrics community (Starr, 2024; Baily, 2024).

A similiar independence assumption has persisted in the iris recognition space. Daugman's foundational work framed the left and right irises of an individual as fundamentally random and uncorrelated (Daugman and Downing, 2001), and Hollingsworth *et al.* reaffirmed that the left and right irises of the same person have uncorrelated irises when they assert, "it is generally accepted that the left and right irises of the same person have uncorrelated iris codes" (Hollingsworth et al., 2011). In (Fang et al., 2019), researchers experiment with matching genetically identical irises using grayscaled

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Imageomics.

versions of the images on different channels of a CNN. They were able to demonstrate the predictive power of CNNs to match the irises of matching vs non-matching individuals.

In this paper we extend the empirical investigation of intra subject biometric dependence in three key ways:

1. We attempt to replicate the fingerprint study of Guo *et al.* under a low-data regime, while also evaluating additional deep learning architectures.
2. We conduct an analysis of potential correlations between an individual's two irises using contrastive loss with bi-encoder networks.
3. We examine cross-biometric correlation, and propse a framework to learn the extent to which patterns in fingerprints and irises can indicate a shared origin.

Together, these efforts are aimed toward the development of a more nuanced understanding of biometric uniqueness. The work here is a jumping-off point for further research that will have direct implications for the design and security analysis of next generation biometric systems.

2 Problem Formulation and Approach

Fundamentally, the problem we are trying to solve is determining if two images (two fingerprints, two irises, one iris/one fingerprint) come from the same individual or come from two different individuals. Formally, we treat biometric verification as a binary decision task. Given two images $x_1 \in \mathcal{X}_1$, $x_2 \in \mathcal{X}_2$ ($\mathcal{X}_1$ may be the same, in the case of iris-to-iris matching, or different, in the case of fingerprint-to-iris matching, from $\mathcal{X}_2$), the goal is to estimate

$$f : \mathcal{X}_1 \times \mathcal{X}_2 \longrightarrow \{0, 1\},$$

where $f(x_1, x_2) = 1$ denotes a valid pair and 0 an invalid pair. Bi-Encoder (also known as Siamese networks) networks are excellent model candidates for this type of matching problem on images, and a similar approach was taken in Guo et al (Malhotra, 2023).

In classic Bi-Encoder architectures, we train a network that takes in two images (generally in the same representational space) and learns embeddings of the images (Koch et al., 2015). During training, one calculates the loss using a Contrastive Loss function that calculates a distance metric between the embeddings and then penalizes matching pairs with high distances and non-matching pairs with low distances. The contrastive loss in our formulation is calculated as:

$$\mathcal{L}(x_1, x_2, y) = (1 - y) \left\| g(z_1) - g(z_2) \right\|_2^2 + y \left[\max(0,\, m - \left\| g(z_1) - g(z_2) \right\|_2) \right]^2$$

where g is the embedding network, m is the margin and y is the true label for x_1 and x_2

This results in embeddings that should be similar for matching pairs and far apart for non-matching pairs.

In attempting to create networks to predict whether two random irises belong to the same individual (iris-to-iris) or two random fingerprints belong to the same individual (fingerprint-to-fingerprint), we used a traditional Bi-Encoder architecture. In tackling the one iris and one fingerprint problem, our first attempt involved a naive Bi-Encoder/Two Tower hybrid network approach that involved training two networks simultaneously, both of which had not previously seen fingerprint or iris data during pre-training. We also attempted to use the pretrained Bi-Encoder models from the iris-to-iris and fingerprint-to-fingerprint specific task models.

3 Methodology

3.1 Dataset

We use a de-identified multi-biometric dataset collected by West Virginia University and Clarkson University, which links records via subject identifiers and enables cross-trait analysis while preserving privacy (Crihalmeanu et al.). For this study we selected subjects with both fingerprint and iris data, yielding 274 individuals with $\sim$100k fingerprint images and 7k iris images.

Table 1: Principal training configs. All models use ImageNet-1K initialization; † indicates towers pretrained from i1/f1.

ID	Task	Backbone	Batch	GPUs	LR
i1	Iris–Iris	RN50	50	7	3e-4
i2	Iris–Iris	ViT-B/16	16	3	3e-4
i3	Iris–Iris	ViT-L/32	16	2	3e-4
f1	Fingerprint–Fingerprint	RN50	50	7	3e-4
f2	Fingerprint–Fingerprint	ViT-B/16	16	3	3e-4
c1	Iris–Fingerprint	2×RN50	50	7	3e-4
c2	Iris–Fingerprint	2×RN50†	50	7	5e-5

† Towers initialized from i1 (iris) and f1 (fingerprint).

3.2 Data Preparation

All images were grouped by subject and split at the subject level into **train/val/test** sets (80%/10%/10%), preventing identity leakage across splits. For each task we constructed balanced (1:1) positive/negative pair sets within each split:

- **Iris–iris:** all left–right iris combinations from the same subject (label 1); negatives were left–right pairs from different subjects in the same split (label 0).
- **Fingerprint–fingerprint:** pairs of different fingers from the same subject (label 1); negatives drawn from different subjects (label 0).
- **Iris–fingerprint:** each iris of a subject paired with each fingerprint of the same subject (label 1); negatives paired with fingerprints from randomly sampled other subjects (label 0).

Because image counts per subject vary, we normalized subject frequency by additional sampling of under-represented identities so that each individual contributes equally to training. When comparing multiple backbones, the exact same subject lists for train/val/test were reused across models to avoid cross-model leakage.

3.3 Training

We train Bi-Encoder models with ResNet-50 and Vision Transformer backbones (He et al., 2015; Dosovitskiy et al., 2020). The final classification layer is replaced by a projection head (fully connected + batch normalization) producing the embedding. Models are optimized with Adam (initial LR 3×10^{-4}) and a step-decay scheduler (step size $= 5$ epochs, $\gamma = 0.1$). We apply light augmentations (color jitter, small rotations, normalization) during training and disable augmentation at test time. Early stopping is based on validation performance, and decision thresholds for binary classification are selected on the validation set.

4 Results and Discussion

We evaluate three verification tasks using ROC AUC as the primary metric and report complementary accuracy, precision, and recall (Fig.1,3,4,5).

4.1 ROC AUC Performance

Figure 1 aggregates our ROC AUC scores alongside two fingerprint baselines from Guo *et al.* (Guo et al., 2024). Overall, iris-iris achieves the strongest separability; fingerprint-fingerprint reproduces the positive intra-subject signal from prior work; cross-modal iris-fingerprint remains weak.

4.2 Iris-to-Iris

The ResNet-50 bi-encoder attains ROC AUC 0.91, indicating that contralateral irises are not statistically independent. Qualitative inspection of errors suggests many false negatives stem from poor image quality, eyelid/eyelash occlusion, or specular reflections. Vision Transformers (ViT-B/16,

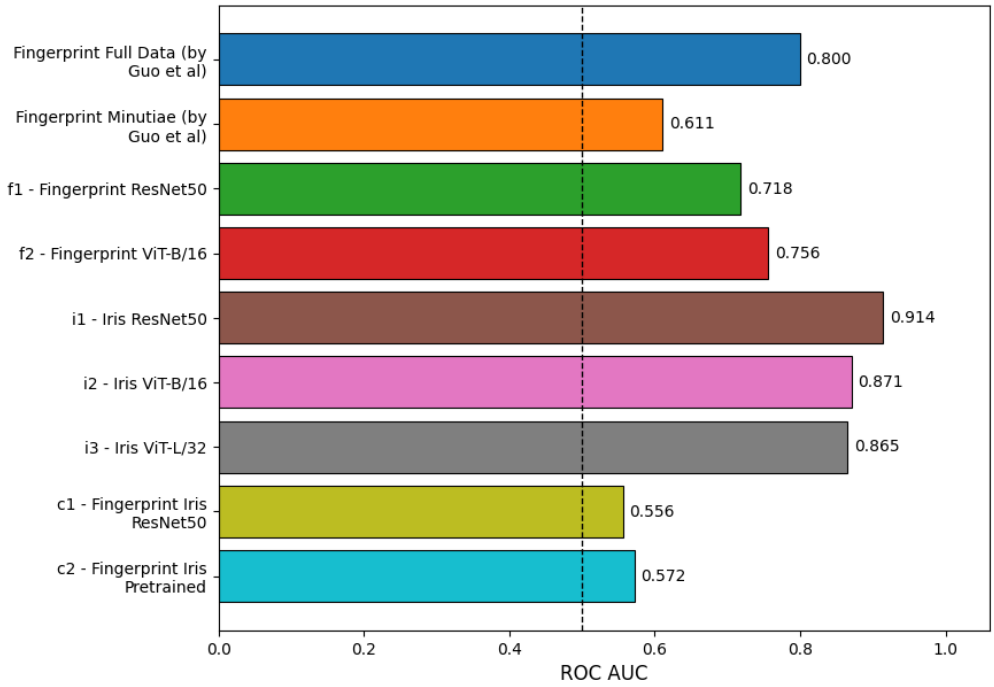


Figure 1: ROC AUC (including Guo *et al.* results).

ViT-L/32) exceed chance but plateau in the mid-to-high 0.80s; with limited data, CNN inductive biases appear advantageous relative to ViTs trained from scratch.

4.3 Fingerprint-to-Fingerprint

We reproduce the qualitative conclusion of Guo *et al.* that fingerprints from the same subject exhibit intra-subject signal, though our ROC AUC does not match their best report. Likely factors include fewer participants ($\sim 3\times$ fewer), less extensive pretraining (e.g., synthetic data), and architectural/training differences. Notably, ViT-B/16 outperforms ResNet-50 in our setting, suggesting transformer backbones can be competitive when enough fingerprint data are available.

4.4 Fingerprint-to-Iris (Cross-Modal)

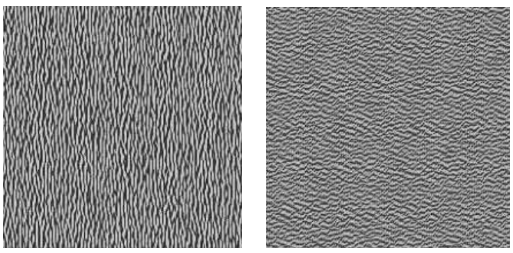
Cross-modal verification yields only slight gains over chance, even with our strongest two-tower configurations. This suggests a weaker measurable correlation between fingerprint and iris traits under our data and training regime; larger datasets, cross-modal pretraining, or stronger alignment objectives may be required to uncover robust signal.

5 Interpretability

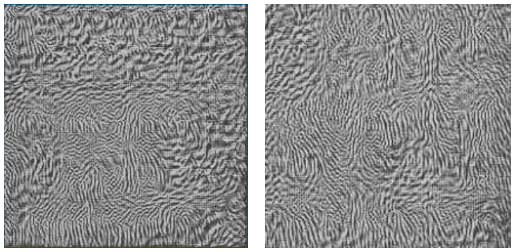
To explore the internal representations learned by the networks, we adopted a DeepDream-style open-source CNN visualisation code of Ozbulak (Ozbulak, 2019) as was done by Guo *et al.*. For select CNN layers, we perform gradient ascent on an initial image comprised of noise to see what features in the input image the layer is responding to. This gives intuitive insight into the visual patterns that affect the results.

Early-layer features. Figures 2a and 2c present the resultant images for the first convolutional block of the fingerprint and iris ResNet50 architectures, respectively. The networks for both tasks seem to be learning the core structures of fingerprints and irises.

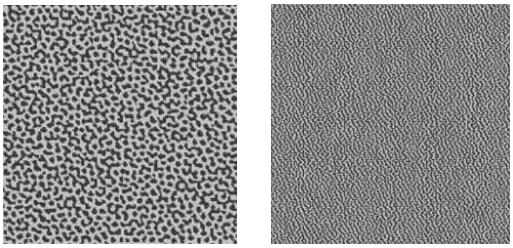
Deep-layer features. In deeper layers (Figures 2b and 2d), the images contain more complex representations of the patterns found on fingerprints and irises. For fingerprints, the network appears to focus on minutiae-like patterns, including swirls, ridges and bifurcations, that may be used to identify an individual. The iris network, on the other hand, seems to demonstrate patterns of the stoma of the iris. The notable differences between early and late layers suggest that the Bi-Encoder architecture learns a sequential hierarchy from coarse structural features to complex patterns.



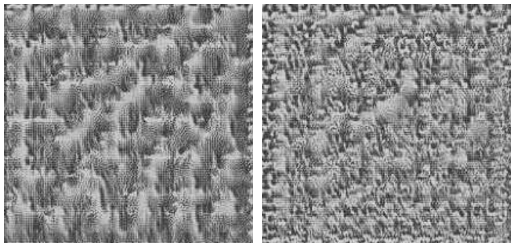
(a) DeepDream visualisations from **Layer 1** of the fingerprint ResNet-50. The results resemble course fingerprint patterns.



(b) DeepDream visualisations from **Layer 4** of the fingerprint ResNet-50. Later filters capture complex minutiae-like ridge patterns.



(c) DeepDream visualisations from **Layer 1** of the iris ResNet-50. Filters the core patterns of irises.



(d) DeepDream visualisations from **Layer 4** of the iris ResNet-50. Later filters seem to respond to complex patterns found in irises.

6 Future Work

The main areas of future work are: (i) systematic hyperparameter and architecture search with scaling studies; (ii) automated image-quality scoring and filtering of low-quality samples, complemented by targeted augmentations; (iii) domain-specific pretraining (e.g., synthetic biometrics, self-supervised objectives) and exploration of alternative contrastive losses/margins and alignment objectives; and (iv) broader cross-modal studies on larger, more diverse datasets, including voice, vein patterns, and hand geometry.

7 Conclusion

This work revisits the independence assumption for biometric traits using bi-encoder contrastive learning. Under a tighter data budget, we replicate fingerprint intra-subject effects and, in a large left-right iris study, observe substantial signal (ROC AUC 0.91 with ResNet-50). Vision Transformers are competitive for fingerprints but lag on irises in this data regime, and cross-modal iris–fingerprint verification remains only slightly above chance. Feature visualizations indicate the models capture meaningful ridge and iris-texture patterns from low- to high-level representations. These findings motivate re-examining multi-trait fusion methods that assume independence and scaling cross-modal studies with larger datasets.

References

Skyla Baily. Ai discovers intra-person fingerprint similarity, 2024.

Simona Crihalmeanu, Wv Morgantown, Arun Ross, and Lawrence Hornak. A protocol for multibiometric data acquisition, storage and dissemination.

J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14 (1):21–30, 2004.

John Daugman and Cathryn Downing. Epigenetic randomness, complexity and singularity of human iris patterns. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1477):1737–1740, 2001.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *CoRR*, abs/2010.11929, 2020.

Beihua Fang, Yuanfu Lu, Zhisheng Zhou, Zhihui Li, Yuwen Yan, Linfeng Yang, Guohua Jiao, and Guangyuan Li. Classification of genetically identical left and right irises using a convolutional neural network. *Electronics*, 8 (10), 2019.

Gabe Guo, Aniv Ray, Miles Izydorczak, Judah Goldfeder, Hod Lipson, and Wenyao Xu. Unveiling intra-person fingerprint similarity via deep contrastive learning. *Science Advances*, 10(2):eadi0329, 2024.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

Karen Hollingsworth, Kevin W. Bowyer, Stephen Lagree, Samuel P. Fenker, and Patrick J. Flynn. Genetically identical irises have texture similarity that is not detected by iris biometrics. *Computer Vision and Image Understanding*, 115(11):1493–1502, 2011.

Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. 2015.

Abhiraj Malhotra. Single-shot image recognition using siamese neural networks. In *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pages 2550–2553, 2023.

Utku Ozbulak. Pytorch cnn visualizations. *GitHub repository*, 2019.

Neil J. Risch and B. Devlin. On the probability of matching dna fingerprints. *Science*, 255(5045):717–720, 1992.

Michelle Starr. Groundbreaking study reveals your fingerprints aren't as unique as we thought, 2024.

A Appendix

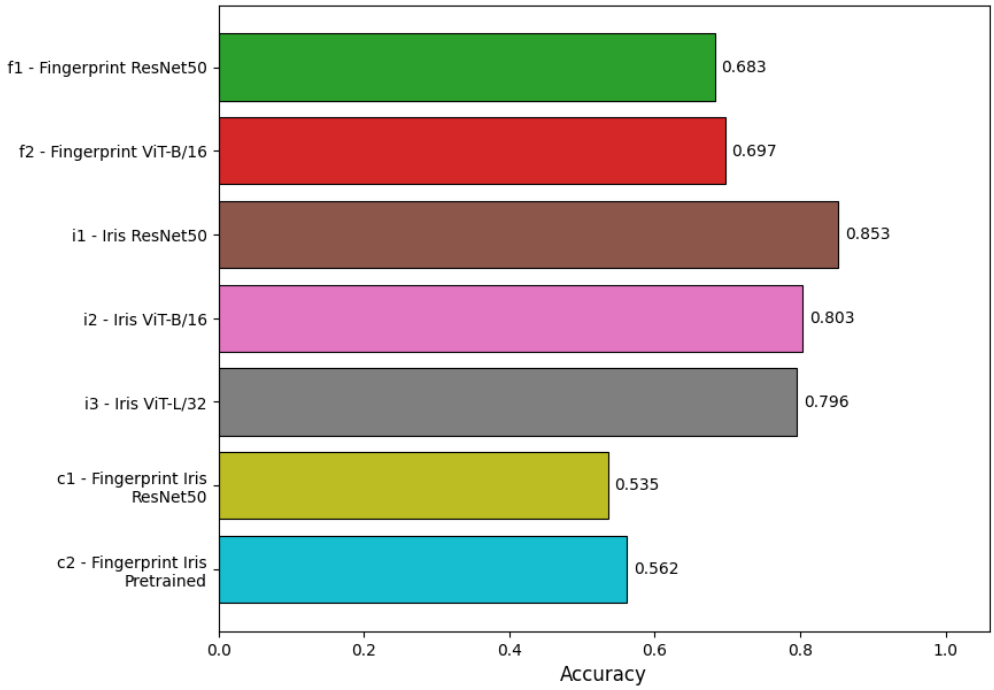


Figure 3: Accuracy on the held-out test set.

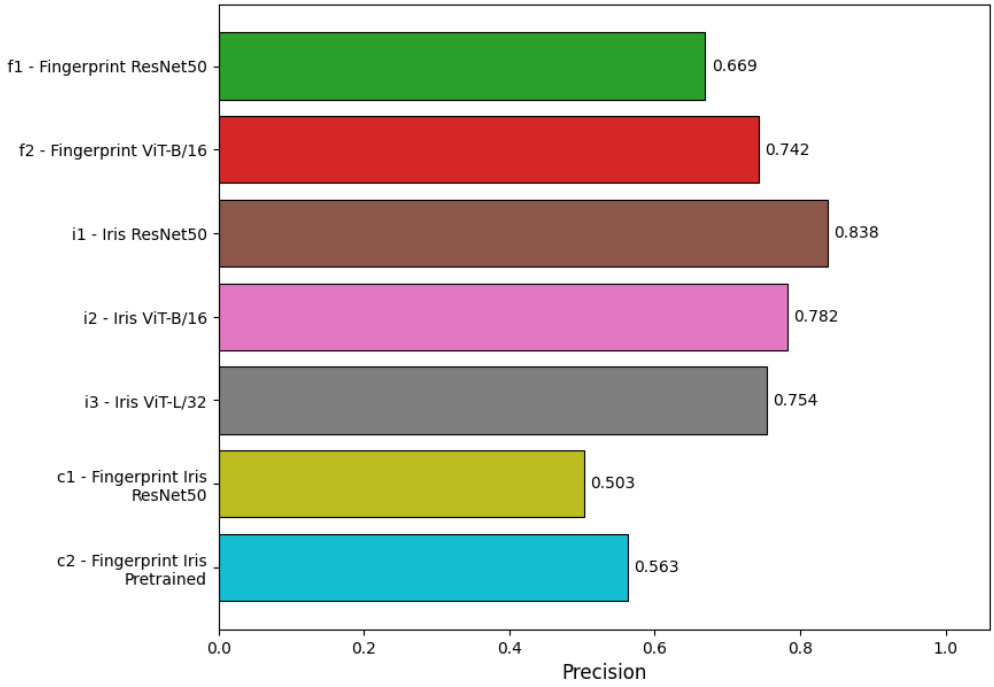


Figure 4: Precision on the held-out test set.

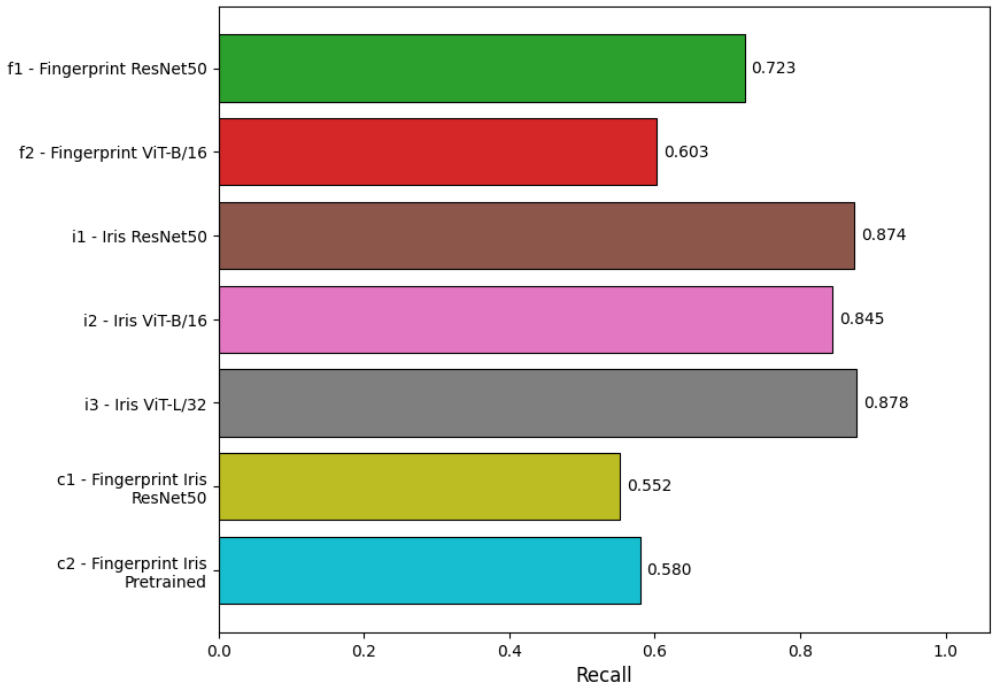


Figure 5: Recall on the held-out test set.