VeinNet: CNN Palm Vein Identification System

Lucian Dorin Crainic
Department of Computer Science
crainic.1938430@studenti.uniroma1.it

La Sapienza University of Rome

Abstract

VeinNet is a Convolutional Neural Network (CNN)–based system for palm vein identification. The approach begins by extracting the palm region of interest (ROI) from a dataset of multispectral hand images, focusing on the vein patterns within this ROI as input to the CNN. Trained on these hand images, VeinNet demonstrates the ability to effectively recognize unique vein patterns in the human palm. The system is evaluated using three distinct types of evaluations: identification with a closed set, identification with an open set, and verification. The dataset is split differently for each evaluation setup, ensuring that the system is tested under various conditions. The results demonstrate the effectiveness of VeinNet in identifying individuals based on their palm vein patterns, with promising performance metrics across all evaluation setups.

1 Introduction

In recent years, hand biometrics has become an increasingly popular modality in biometric recognition systems due to its accessibility and rich discriminatory features. Traditionally, hand-based systems have relied on contact-based devices equipped with pegs or plates for image acquisition. While effective, these systems often raise hygiene concerns and reduce user acceptance. In response, there has been a shift toward contact-free systems that eliminate the need for physical contact during data capture [Xiong et al., 2005b, Jiang et al., 2007, Xiong et al., 2005a]. However, increased freedom of hand movement in contact-free setups often leads to reduced recognition accuracy. The integration of multispectral imaging techniques has been successfully applied to improve recognition performance in other biometric domains, such as face recognition [Kong et al., 2007, Singh et al., 2008]. Similarly, for hand biometrics, [Wang and Leedham, 2006] demonstrated that passive infrared imaging is inadequate for extracting vein patterns from the palm. This limitation has led to the exploration of active multispectral imaging across visible to near-infrared wavelengths. Previous studies, such as [Wang et al., 2007], have illustrated the potential of combining palmprint and palm vein images using fusion techniques applied at the image level. However, these approaches often rely on semi-touchless acquisition systems or frequency-division hardware, which may limit scalability and increase costs.

This report proposes a method for palm vein pattern identification using a convolutional neural network (CNN) architecture. Unlike traditional approaches relying on pixel-level fusion and feature-level registration techniques, the proposed method uses the power of CNNs to automatically learn and extract distinctive features from palm vein patterns. By focusing on the rich discriminatory information in vein structures. This approach represents a significant step forward in developing efficient, hygienic, and user-friendly biometric systems.

2 Dataset

This section provides an overview of the CASIA Multi-Spectral Palmprint Image Database, including data acquisition details and a description of the dataset.

2.1 Data Aquisition

The self-designed imaging device for acquiring hand images [Hao et al., 2008, Hao et al., 2007] is shown in Figure 1. The device operates in a contact-free environment, The imaging process involves the following

steps:

1. **Illumination Setup**: The device uses six groups of LEDs (violet to near-infrared) activated sequentially, employing a time-division strategy to acquire multispectral images under varying illumination, capturing different skin layers through light absorption and scattering.

2. **Reflective Imaging**: Images are captured reflectively in a sheltered environment with consistent illumination, while circularly arranged LED groups, diffused with ground glass, ensure even lighting across the hand.

3. **Contact-Free Operation**: Subjects are instructed to naturally stretch their hands, palms facing the camera, without any physical contact with a tangible surface or plate.

4. **Sequential Image Capture**: A single camera is used to sequentially capture images under each spectral light. This time-division strategy improves scalability and offers a better performance-to-cost ratio compared to frequency-division methods requiring multiple cameras.

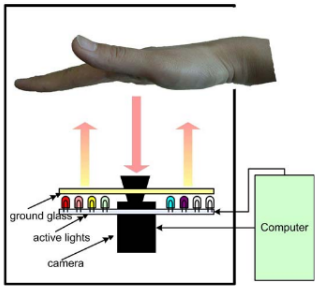


Figure 1: Device architecture for acquiring multispectral palm images.

2.2 Dataset Description

The CASIA Multi-Spectral Palmprint Image Database consists of 7,200 palm images captured from 100 individuals using a self-designed multispectral imaging device. The images are 8-bit gray-level JPEG files. For each hand, two sessions of palm images were captured, with a time interval of more than one month between the sessions to simulate real-world conditions and introduce natural variability. Each session includes three samples, with each sample containing six palm images captured simultaneously under six different electromagnetic spectrums, corresponding to wavelengths of 460 nm, 630 nm, 700 nm, 850 nm, 940 nm, and white light shown in Figure 2. Variations in hand postures were allowed between the two sessions to increase the diversity of intra-class samples, thereby simulating practical usage scenarios and enhancing the robustness of biometric recognition systems trained on this dataset.

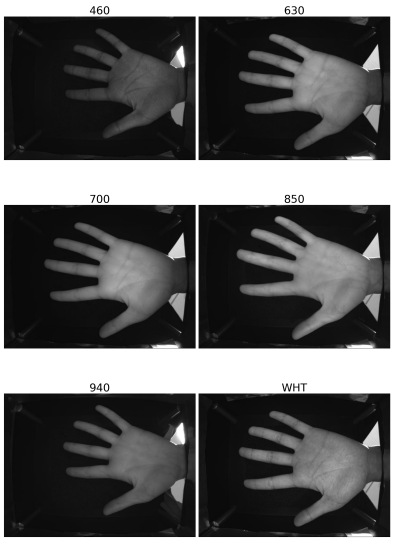


Figure 2: Palmprint images from the CASIA Multi-Spectral Palmprint Image Database with the six spectral bands. Starting from the top-left corner and moving clockwise: 460 nm, 630 nm, 700 nm, 850 nm, 940 nm, and white light.

3 Methodology

3.1 Data Preprocessing

1. **Image Loading and Cropping** The palm vein image is loaded in grayscale mode using OpenCV imread function described in Equation 1.

$$I(x, y) = \text{imread}(f, \text{GRAYSCALE}) \tag{1}$$

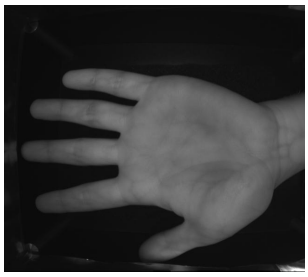
where $I(x, y)$ represents the pixel intensity at coordinates (x, y). To isolate the palm region, the rightmost part of the image was cropped using Equation 2.

$$I_{\text{cropped}}(x, y) = I(x, y), \quad \forall\, x \in [0, W - 120] \tag{2}$$

where W is the original width of the image.

Figure 3a shows the input image selected for preprocessing, while Figure 3b displays the cropped image.

(a) Input Image selected for preprocessing.



(b) Cropped Image after removing the rightmost part.

2. **Blurring and Thresholding** A Gaussian blur was applied to smooth the image and reduce noise and Thresholding was then applied to convert the image to binary format:

$$I_{\text{thresholded}}(x,y) = \begin{cases} 255 & \text{if } I_{\text{blur}}(x,y) \geq T, \\ 0 & \text{if } I_{\text{blur}}(x,y) < T \end{cases} \quad (3)$$

where $T = 50$.

Figure 4a shows the thresholded image after applying Gaussian blur.

3. **Contour Detection** Contours were extracted from the binary image. The largest contour, corresponding to the hand region, was selected.

(a) Thresholded Image after Gaussian Blur.

(b) Contour Image after extracting contours.

Figure 4b shows the contour image after extracting contours.

4. **Convexity Defects Analysis** The convex hull of the largest contour was calculated using the convexHull function from OpenCV described in Equation 4 where C_{largest} represents the largest contour.

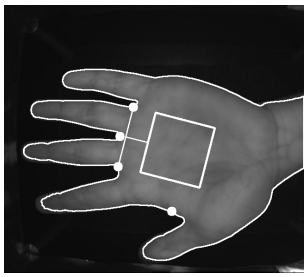
$$H = \text{convexHull}(C_{\text{largest}}) \quad (4)$$

Convexity defects, representing regions where the contour deviates inward from the hull, were identified using OpenCV's convexityDefects function described in Equation 5.

$$D = \text{convexityDefects}(C_{\text{largest}}, H) \quad (5)$$

5. **Region of Interest (ROI) Extraction** Key points from the convexity defects were used to define the palm's region of interest. The midpoint of a line connecting two key points was calculated to determine the center of the palm using Equation 6. Once the midpoint was identified as the hand center,

$$M = \left(\frac{x_1 + x_3}{2}, \frac{y_1 + y_3}{2} \right) \quad (6)$$



(a) FAR vs. FRR

(b) Confusion Matrix

Figure 5: TODO

6. **Feature Enhancement**
Histogram equalization was applied to improve contrast:

$$I_{\text{equalized}} = \text{HE}(I_{\text{rectified}})$$

A Gabor filter was then used to emphasize vein patterns:

$$G(x,y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2} \right) \cos\left(2\pi \frac{x'}{\lambda} + \psi \right)$$

7. **CLAHE**
Contrast Limited Adaptive Histogram Equalization (CLAHE) was performed iteratively to refine local contrasts:

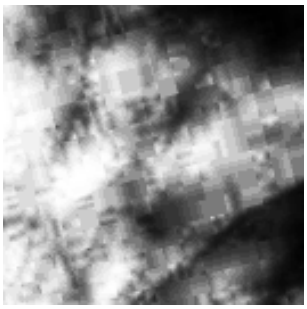
$$I_{\text{CLAHE}} = \text{CLAHE}(I_{\text{filtered}}, \text{clipLimit}, \text{tileGridSize})$$

8. **Binary Thresholding**
The processed image was binarized to isolate vein patterns:

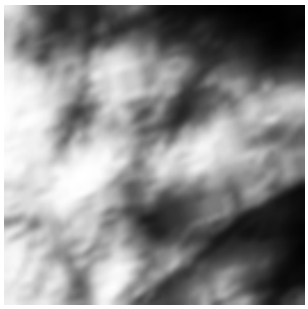
$$I_{\text{binary}}(x,y) = \begin{cases} 255 & \text{if } I_{\text{CLAHE}}(x,y) \geq T_{\text{final}}, \\ 0 & \text{otherwise} \end{cases}$$

9. **Resize and Save** Each image after the binary thresholding step was resized to 128×128 pixels to standardize the input size for the neural network model.

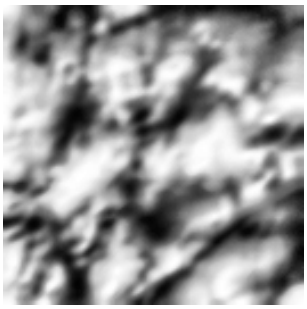
Figure 7b shows the final output of the preprocessing pipeline.



(a) Image after applying Histogram Equalization, Perspective Transformation, and Warp Perspective.



(b) Image after applying Gabor Filter.



(a) Enhanced image after Using CLAHE and Gaussian Blur.

(b) Binary Thresholding of the image.

3.2 Model Architecture

Convolutional Neural Networks (CNNs) are a class of deep learning models specifically designed to process and analyze grid-like data such as images. They use layers of convolutional filters to automatically detect and learn hierarchical patterns in the data, making them particularly effective for image recognition and classification tasks [Lecun et al., 1998].

The structure of the neural network models developed for the palm print identification system is detailed here. These models share a similar core structure, with differences in specific components to address their respective objectives: closed-set identification, open-set identification, and verification tasks.

Core Structure: Both models rely on the same convolutional backbone for feature extraction. This shared structure consists of four convolutional layers that extract features from the input images. Each convolutional layer increases the number of feature maps while reducing spatial dimensions using ReLU activation followed by max pooling. After the convolutional layers, the flattened feature maps are passed through a linear layer to produce a 128-dimensional representation of the image features. This linear layer, acting as a dense fully connected layer, is essential for transforming the high-dimensional convolutional features into a compact representation suitable for further processing. This feature extraction backbone forms the foundation of both models.

$$f(x) = \max(0, x) \tag{7}$$

$$y_{i,j} = \max_{(m,n)\in R_{i,j}} x_{m,n} \tag{8}$$

This model is designed for classifying palm print images into predefined classes, such as 100 patients. Following the shared backbone, the model includes a linear layer for reducing the dimensionality, a dropout layer to prevent overfitting, and a final classification layer with a softmax activation function. The output provides class probabilities, enabling closed-set identification.

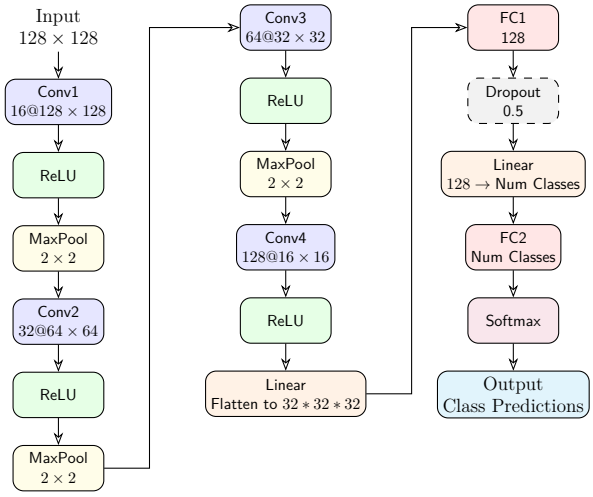


Figure 8: Forward pass of a Convolutional Neural Network (CNN) divided into three columns, including Input and Output. The output provides class predictions.

This model extends functionality to handle open-set identification and verification tasks. After the shared backbone, the model uses a linear layer to extract a compact feature representation, followed by label embeddings to capture label-specific information. It concatenates the 128-dimensional image features with the label embeddings to form a combined 256-dimensional vector. This fused representation is passed through a fully connected layer and sigmoid activation to output a verification score between 0 and 1, indicating the likelihood of a valid image-label pair.

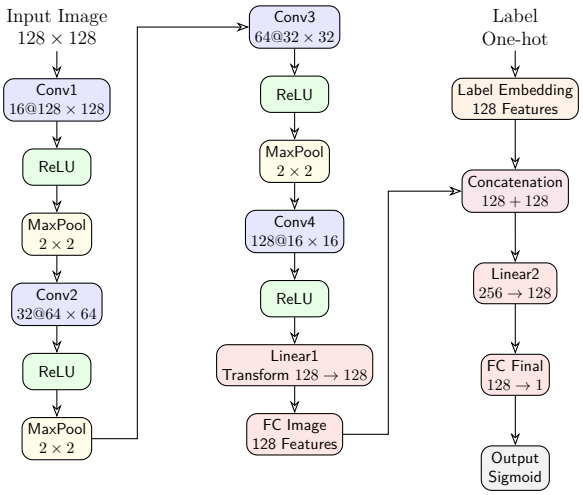


Figure 9: Neural network layout with added Linear1 (image features), Linear2 (concatenated features), FC Final, and Sigmoid layers for output.

4 Experimental

This section describes the experimental setup introducing the 3 evaluation setups used to assess the performance of the proposed biometric system and the data splitting strategy used for each evaluation setup.

4.1 Evaluation Setups

The evaluation was performed using three distinct types of evaluations:

- **Identification with a Closed Set**: involves including all enrolled patients in the dataset. Each image is classified into one of the known classes (patients) based on its extracted features. The system is trained and tested with the same predefined set of enrolled patients, ensuring no unknown users are present in the dataset.
- **Identification with an Open Set**: excludes a percentage of the enrolled patients from the dataset during training. These excluded patients represent unknown users during the evaluation phase. The model is tested on both known and unknown classes, where the unknown classes are expected to be classified as unknown to simulate open-set identification.
- **Verification**: tests the system's ability to verify the identity of users. Genuine samples consist of images correctly matched to their claimed identities, while imposter samples are created by associating images with incorrect user identities to simulate attempts to mislead the system.

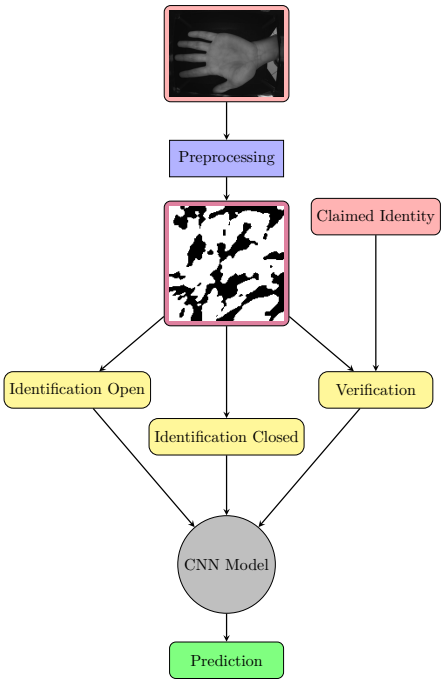


Figure 10: Diagram scaled

4.2 Data Splitting

The dataset used in this study was split differently for the three evaluation setups:

Identification with a Closed Set includes all patients in the dataset, and their images are divided into training and test sets. The first four images per patient are used for training and the remaining images are used for testing. This setup ensures that all enrolled patients contribute images to the training and testing phases, facilitating evaluation in a controlled, closed-set scenario.

Identification with an Open Set For open-set identification, 70% of patients are randomly selected as known, with their images split into training (first four images) and testing (remaining images). The remaining 30% serve as unknown patients, with their images used solely for testing to assess the system's ability to handle unenrolled users, simulating open-set scenarios.

Verification includes genuine and imposter samples for evaluation. For each patient, the first four images are used for training, and the remaining images are used for testing. Genuine samples consist of matching the correct image to the claimed identity, while imposter samples are created by pairing images from different patients, simulating attempts to impersonate other users.

To ensure consistency and reproducibility across all

splits, a fixed random seed was used during shuffling. This guarantees that the data partitioning remains consistent across different runs and experimentsss

5 Evaluation Metrics

This chapter describes the evaluation metrics used to assess the performance of the biometric recognition system.

- **False Acceptance Rate (FAR)**: The proportion of impostor samples incorrectly accepted by the system.

$$\text{FAR} = \frac{\text{Number of False Acceptances}}{\text{Total Number of Impostor Attempts}}$$

  A lower FAR indicates better system security.

- **False Rejection Rate (FRR)**: The proportion of genuine samples incorrectly rejected by the system.

$$\text{FRR} = \frac{\text{Number of False Rejections}}{\text{Total Number of Genuine Attempts}}$$

  A lower FRR indicates better system usability.

- **Equal Error Rate (EER)**: The point where the FAR and FRR are equal, providing a single measure of system performance.

$$\text{EER} = \text{FAR} = \text{FRR at Threshold}$$

  A lower EER signifies a more balanced system.

- **Receiver Operating Characteristic (ROC) Curve**: Plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various thresholds.

$$\text{TPR} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{FPR} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}}$$

  The Area Under the Curve (AUC) summarizes the ROC curve, where higher AUC values indicate better performance.

- **Detection Error Tradeoff (DET) Curve**: A variation of the ROC curve that plots the FRR against the FAR on a logarithmic scale, providing insight into trade-offs between errors.

- **Cumulative Match Characteristic (CMC) Curve**: Represents the probability of correct identification as a function of rank in a closed-set scenario.

$$\text{Rank-k Identification Rate} = \frac{\text{Number of Correct Identifications}}{\text{Total Number of Queries}} \quad (9)$$

  Higher Rank-1 and rapid convergence to 1.0 indicate strong identification performance.

- **Confusion Matrix**: Summarizes the system's performance in terms of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total Samples}}$$

  Provides a comprehensive view of the system's classification abilities.

6 Results

6.1 Identification with a Closed Set

The biometric system's performance for closed-set identification was evaluated using the Cumulative Match Characteristic (CMC) curve, which measures the probability of correct identification at various ranks.

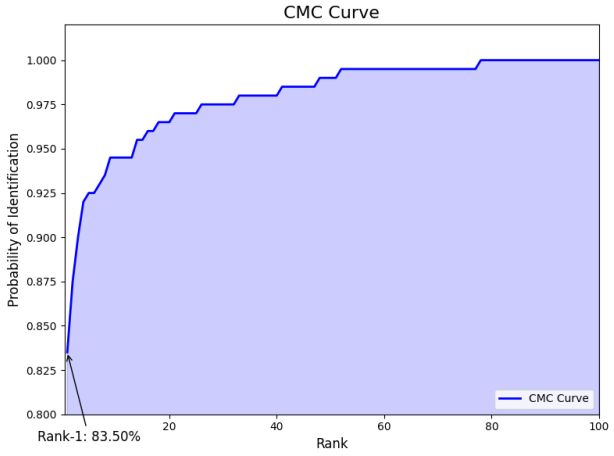


Figure 11: CMC Curve demonstrating the probability of identification across ranks for closed-set identification.

CMC Curve: The CMC curve indicates that the system achieves a Rank-1 identification rate of 83.50%, highlighting its ability to correctly identify subjects in the top rank most of the time. The curve rapidly approaches a probability of 1.0 as the rank increases, demonstrating the system's reliability in closed-set scenarios where all subjects belong to the known population.

Overall, the results from the CMC curve establish the system's strong identification performance in closed-set conditions, with high accuracy for identifying subjects within the top few ranks.

6.2 Identification with an Open Set

The biometric system's performance for open-set identification was evaluated using the Receiver Operating Characteristic (ROC) curve, plotting the Detection and Identification Rate (DIR) against the False Alarm Rate (FAR).

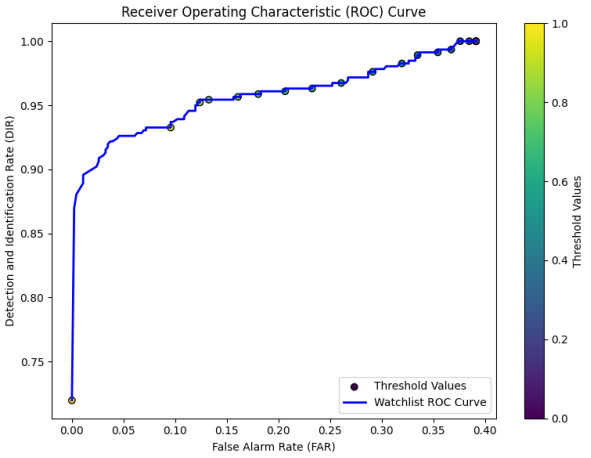


Figure 12: Watchlist ROC Curve demonstrating the DIR vs. FAR for open-set identification.

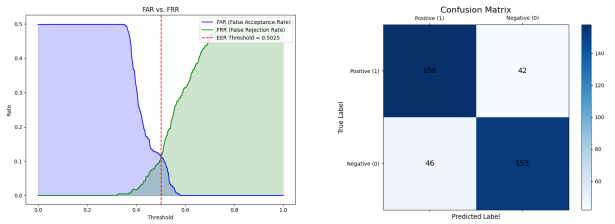
Watchlist ROC Curve: The ROC curve demonstrates the system's effectiveness in maintaining a high Detection and Identification Rate (DIR) while minimizing the False Alarm Rate (FAR). The curve achieves strong performance, as evidenced by a consistent increase in DIR with lower FAR values. Threshold markers along the curve provide insights into the system's operating characteristics and trade-offs between detection and false alarms.

Overall, the evaluation results indicate that the system is capable of handling open-set identification scenarios effectively, distinguishing between known and unknown subjects with high reliability.

6.3 Verification

The biometric verification system was evaluated using False Acceptance Rate (FAR), False Rejection Rate (FRR), Equal Error Rate (EER), and Receiver Operating Characteristic (ROC) metrics.

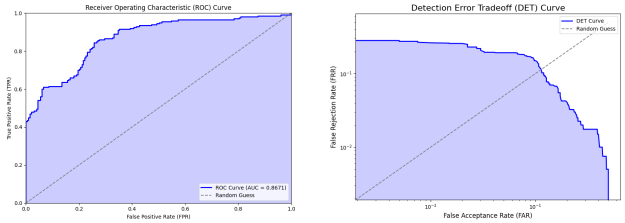
FAR vs. FRR Analysis: The FAR and FRR curves intersect at a threshold of 0.5025, yielding an EER of 0.5025. This indicates a balanced trade-off between the acceptance of impostors and rejection of genuine users, critical for assessing verification performance.



(a) FAR vs. FRR (b) Confusion Matrix

Figure 13: FAR vs. FRR curve with the EER threshold and the Confusion Matrix.

Confusion Matrix: The system achieved a good balance between true positives (158) and true negatives (153) while maintaining acceptable rates of false positives (42) and false negatives (46). This supports the system's robustness in distinguishing between genuine and impostor samples.



(a) ROC Curve (b) DET Curve

Figure 14: ROC and DET curves highlighting the system's performance.

ROC Curve: With an Area Under the Curve (AUC) of 0.8671, the ROC curve demonstrates strong discriminatory power between genuine and impostor samples, showcasing the system's reliability in verification tasks.

Detection Error Tradeoff (DET) Curve: The DET curve highlights the system's error rates over a range of operating conditions. The downward trend reflects the system's ability to minimize errors as the threshold is optimized.

Overall, the evaluation metrics indicate that the system achieves a reliable performance in biometric verification, effectively balancing FAR and FRR while providing a solid classification accuracy. These results establish the system's viability for biometric recognition applications.

7 Conclusion and Future Work

7.1 Conclusion

In this work, a straightforward yet highly effective CNN-based recognition system was developed for the purpose of hand vein recognition. Despite the architecture's simplicity, it demonostrated remarkable efficiency by achieving rapid training convergence and maintaining low computational requirements. The results indicate that the model achieves an accuracy of **xx.x%** on the training set, traslating to a robust perfomance when deployed on the test set, where it obtained an accuracy of **xx.x%**. Such a perfomance level is significant for a recognition task, as it ensures reliable identification while keeping resource consumption and development complexity to a minimum. Overall, this implementation highlights that even a relatively uncomplicated CNN framework can deliver strong accuracy and efficiency, making it an appealing solution for practical vein-based recognition systems.

7.2 Future Work

An interesting extension to this project for future exploration involves expanding the focus beyond the current palm-based Region of Interest (ROI). Instead of restricting the system solely to the palm vein pattern, it would be beneficial to consider additional hand characteristics such as overall hand geometry, finger length and spacing, and other distinctive biometric features of the hand. Using these complementary features would allow us to develop multiple specialized models, one dedicated to palm veins (as implemented in this project), another concentrating on hand geometry, another on finger structure, and so on. By utilizing an ensemble learning strategy (see Figure 15), where each specialized model contributes its prediction, we could combine these outputs into a final, more robust decision. This multi-model, ensemble-based approach has the potential to significantly improve the overall accuracy and reliability of the system.

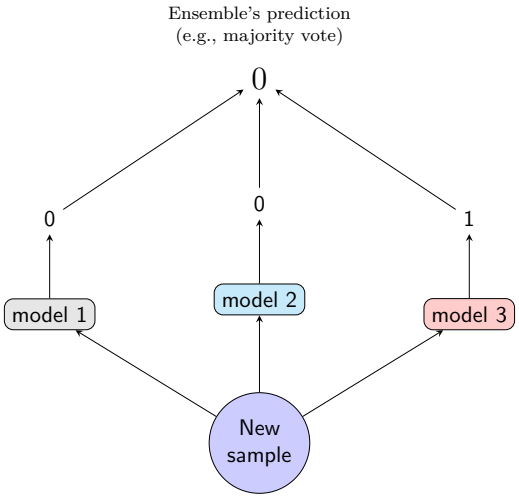


Figure 15: Illustration of an ensemble learning strategy. Each model processes the same input and produces a prediction. These predictions are combined (e.g., through majority voting) to create the ensemble's final output.

References

[Hao et al., 2007] Hao, Y., Sun, Z., and Tan, T. (2007). Comparative studies on multispectral palm image fusion for biometrics. In *Asian conference on computer vision*, pages 12–21. Springer.

[Hao et al., 2008] Hao, Y., Sun, Z., Tan, T., and Ren, C. (2008). Multispectral palm image fusion for accurate contact-free palmprint recognition. In *2008 15th IEEE International Conference on Image Processing*, pages 281–284. IEEE.

[Jiang et al., 2007] Jiang, X., Xu, W., Sweeney, L., Li, Y., Gross, R., and Yurovsky, D. (2007). New directions in contact free hand recognition. In *2007 IEEE International Conference on Image Processing*, volume 2, pages II–389. IEEE.

[Kong et al., 2007] Kong, S. G., Heo, J., Boughorbel, F., Zheng, Y., Abidi, B. R., Koschan, A., Yi, M., and Abidi, M. A. (2007). Multiscale fusion of visible and thermal ir images for illumination-invariant face recognition. *International Journal of Computer Vision*, 71:215–233.

[Lecun et al., 1998] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

[Singh et al., 2008] Singh, R., Vatsa, M., and Noore, A. (2008). Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. *Pattern Recognition*, 41(3):880–893.

[Wang et al., 2007] Wang, J.-G., Yau, W.-Y., Suwandy, A., and Sung, E. (2007). Fusion of palmprint and palm vein images for person recognition based on" laplacianpalm" feature. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE.

[Wang and Leedham, 2006] Wang, L. and Leedham, G. (2006). Near-and far-infrared imaging for vein pattern biometrics. In *2006 IEEE International*

Conference on Video and Signal Based Surveillance, pages 52–52. IEEE.

[Xiong et al., 2005a] Xiong, W., Toh, K.-A., Yau, W.-Y., and Jiang, X. (2005a). Model-guided deformable hand shape recognition without positioning aids. *Pattern recognition*, 38(10):1651–1664.

[Xiong et al., 2005b] Xiong, W., Xu, C., and Ong, S. H. (2005b). Peg-free human hand shape analysis and recognition. In *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 2, pages ii–77. IEEE.