SVM for Hand Based Identification

Classifier for identification based on palmar and dorsal images of hands

Alessio Taruffi 1940484, Patrizio Renelli 1937842, Nicolo Candita 2216597, Matteo Rocco 2218397

Summary

1. Introduction

Biometric hand recognition is an authentication technique increasingly used in security and personal identification contexts. This project aims to develop and test a biometric recognition system based on the analysis of hand images (palm and dorsal) using Support Vector Machines (SVM) classifiers.

This work builds on a previous project that uses convolutional neural networks (CNN) to distinguish the gender (male/female) of people through the analysis of images of the palm and dorsal of the hand. However, while the previous approach exploited CNNs for automatic feature extraction, in this project we chose to adopt different extraction techniques such as Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG), combining with an SVM classifier.

This report describes in detail the entire project workflow, from data preprocessing to the implementation of the recognition system, up to performance evaluation using specific biometric metrics. The final goal is the creation of an automatic supervised recognition system.

2. Dataset Description

The dataset used for this project is a public dataset called "11k Hands", containing hand images from various subjects, each accompanied by associated metadata. It includes a total of 11,076 hand images with a resolution of 1600x1200 pixels. The images were captured from a base of 190 subjects ranging in age from 18 to 75 years old. The dataset features photos of both sides of the hand (palm and dorsal) and provides multiple images for each subject.

The available metadatas for each picture are:

- SUbject ID;
- Subject gender;
- Subject age;
- Hand skin color;
- Hand side;
- Right or left hand;
- Presence of accessories or nail polish

Such metadatas will be essential for image sampling, in the CNN training phase and during the testing phase to evaluate the correctness of the predictions obtained.

The following classifications can be made using the available metadata:

Hand side	Left	Right
Dorsal	2786	2895
Palmar	2585	2810

Table 1: Division by Hand Side

Presence of accessories	Number
No	7865
Yes	3211

Table 2: Division by Accessories

Color	Number
Dark	758
Fair	3493
Medium	6495
Very fair	330

Table 3: Division by Skin color

Gender	Number
Female	7109
Male	3967

Table 4: Division by Gender

From the classifications reported above, the dataset would seem to be correctly balanced with regard to the global subdivision of the data based on the side (palm/back) and the side (right/left).

The main problem of the dataset concerns the subdivision of the images by subject, by gender and, within the individual samples, the distribution relative to the analyzed side (palm/back).
If we were to divide the images according to the subject belonging to it, we can notice a non-homogeneous distribution by subject, as the weight of some, within the dataset, is much greater than the others.

Subject	Subject's number %	Associated images	Associated imgs %
Top 19	10.00%	2318	20.93%
Top 30	15.79%	3429	30.96%
Bottom 30	15.79%	733	6.62%

By analyzing the images for each subject in detail, we find a non-homogeneous subdivision with regards to the number of images for each side.

Subject	Number of subjects	Percentage of subjects
With min threshold at 40%	33	34.78%
With min threshold at 50%	178	93.68%

Taking into account all the above analyses, a sampling process was created that allows the model to be trained correctly.

2.1 Sampling

In addition to the problems indicated above in the Problem section, in the sampling phase, to avoid the creation of bias, it was decided to exclude from the training phase the images containing accessories.
Furthermore, by using a single dataset for both the training and test phases, it was decided to set a policy to avoid using the images already used in the training phase also for the test phase.

The implemented sampling takes some of its characteristics from two different sampling techniques:

1. **Random sampling**: Each element of the dataset has the same probability of being selected, ensuring an unbiased sample;
2. **Oversampling and Undersampling**: Used specifically to balance unbalanced datasets, by adding or removing elements of some classes.

We can have a more in depth look at the pipeline implemented for sampling. After having scanned and loaded into memory the file containing the metadata of the images and subjects contained within the dataset, we went on to build a custom data structure (python dictionary) that allows us to manipulate and select the data. As we have structured our model, the sampling phase is of cardinal importance since each selection has a particular criterion to respect.

2.1.1 Subdivision of images between training and closed test sets

Initially we will extract n subjects randomly where the number of subjects will be equal to the value of the "num_sub" variable, which will be used, in the case of closed sets, to extract the images for both training and testing.
For each subject we will randomly extract n pairs of images (palm/dorsal), inserting them into a data structure that allows us to save the person's identity, indicated by the "person_id" label.

The number of images to extract will be calculated as follows:

- **70%** of the value passed into the “num_img” variable (rounded down) dedicated to the training phase.
- **30%** of the value passed in the “num_img” variable (rounded down) dedicated to the test phase.

Example

num sub = 10
num_img = 100

- We will extract 70 pairs of images (palm & dorsal) for each subject, to be dedicated to training, with possible repetitions.

- We will extract 30 pairs of images (palm & back) for each subject, to be dedicated to the test, with possible repetitions, including the possibility of fishing for some of the images collected in the training.

2.1.2 Subdivision of images between training and open test sets

If you want to perform recognition in open set mode, the extraction methods of users and images will change.
In fact, in the testing phase, n - num of impostor subjects will be extracted from among the subjects extracted for the creation of the gallery. To indicate that a person is an impostor, fundamental for the subsequent evaluation phase of the biometric system, the identity of the analyzed subject will be set to -1.

2.2 Image preprocessing

To increase the performance relating to the recognition of the side of the palm, it was decided to concentrate the extraction of the features only on the part containing the palmar creases. To select only this part of the image, a one-off function has been implemented, called "create_palm_cut_dataset", which uses the OpenCV and MediaPipe libraries.

The function executes the following flow:
- uploading images
- identification of Hand landmarks through the specific MediaPipe functions
- analysis of the extracted points and extraction of landmarks [0, 2, 5, 17] which allow identifying the part of the palm containing the palmar creases
- correction of any errors in landmark extraction
- extraction of the vertices of a box containing the area of interest
- cut of the palm images and creation of the new dataset with the specific images for the palm processed and the original ones for the dorsal

2.2.1 Preprocessing single image

Before being passed to the functions that deal with feature extraction, the images were subjected to specific preprocessing.

Custom transformations have been implemented such as:

1. **CustomLBPCannyTransform**: It is applied to all palm images that must be subjected to the feature extraction process through LBP. In sequence it executes:
    - a resize, bringing the image to be analyzed to a size of 150x150 pixels
    - converts the image to grayscale
    - increases the contrast

    - applies the canny filter that helps reduce noise and bring out the most important information
2. **CustomLBPTransform**: It is applied to all dorsal images that must be subjected to the feature extraction process through LBP. Performs the same steps as the transformation indicated above except for the application of the canny filter
3. **CustomHOGTransform**: It is applied to all images of both the dorsal and the palm that must be subjected to the feature extraction process through HOG. In sequence it executes:
    - converts the image to RGB
    - applies a Gaussian filter to the image to reduce noise
    - a resize, bringing the image to be analyzed to a size of 1024x1024 pixels (dorsal) or 150x150 pixels (palm)

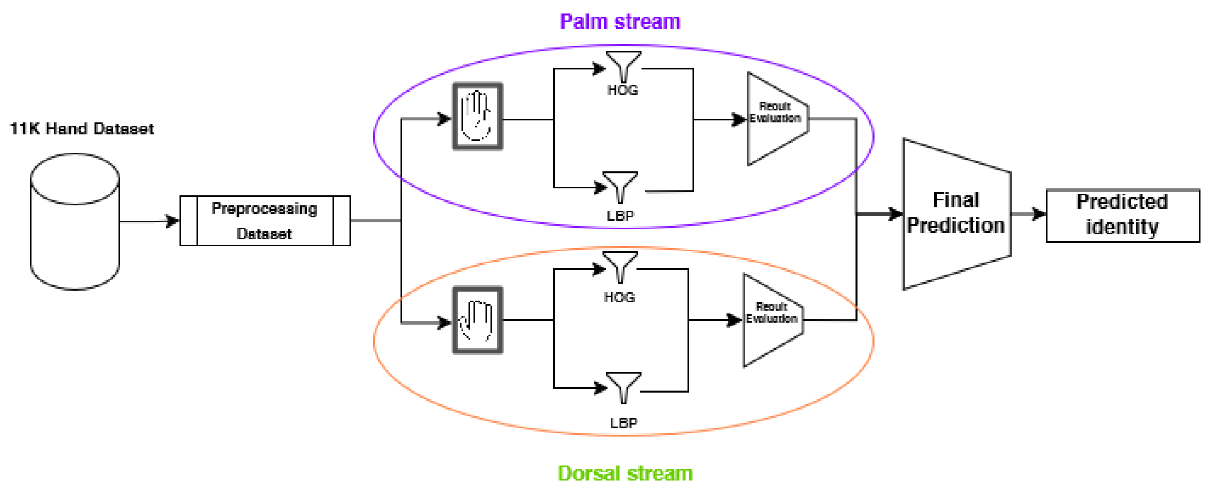
3. System implementation

After seeing how image preprocessing is carried out, we can move on to discussing the actual implementation of the biometric recognition system. This system follows a pipeline structured in several phases, ranging from image acquisition and preprocessing to final classification using Support Vector Machines (SVM).

3.1 Biometric Recognition Pipeline

Leaving aside the already discussed preprocessing, the entire system can be divided into the following key steps:

1. Feature Extraction (LBP and HOG)
2. Training of the SVM Model
3. Fusion of Predictions
4. Performance Evaluation



3.2 Feature Extraction

In the previous CNN-based project, features were automatically extracted from the neural network during training. However, in the present project we chose to use Support Vector Machines (SVM), a classifier that is not capable of independently learning features from raw images.

Consequently, feature extraction was performed using two well-established techniques: Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG). These methodologies allow obtaining a numerical representation of the images, which is then used by the SVM classifier to distinguish the subjects.

3.2.1 Local Binary Pattern (LBP)

Local Binary Pattern (LBP) is a technique used to describe the texture of an image based on the intensity values of surrounding pixels. The method involves the following steps:

1. Split of the image into cells of fixed size (in our case 3×3 pixels).
2. Comparison between the value of each pixel and its neighbors:
    - If the adjacent pixel has a value greater than or equal to the central pixel, 1 is assigned.
    - If it is less, 0 is assigned.
3. Conversion the binary pattern to a decimal number.
4. Construction of a histogram of LBP values, which represents the texture distribution in the image.

This technique is particularly useful in texture recognition, as it is robust to variations in lighting and allows you to capture distinctive details of the skin and lines of the palm.

We used LBP to capture local information about the skin texture and lines of the palm. For each pixel of the image, we compare the intensity value with the surrounding pixels and generate a binary map. The map is then converted in frequency histograms, which represent the image in a compact way.

3.2.2 Histogram of Oriented Gradients (HOG)

The Histogram of Oriented Gradients (HOG) is a feature extraction technique that analyzes the distribution of intensity gradients in the image, thus capturing the shape and contours of objects. The main steps are:

1. Calculation of pixel gradients (intensity variation between adjacent pixels).
2. Subdivision of the image into cells (4x4 pixels in our case).
3. Calculation of the dominant direction of the gradients in each cell.
4. Creation of an histogram of the predominant directions.
5. Union of histograms to obtain a feature vector of the entire image.

HOG is widely used in the recognition of objects and people, as it allows to capture the global shape of a hand more effectively than methods based on pixels alone. For the dorsal and palm it is used to extract the patterns of the grooves of the lines present in the hand.

3.2.3 Application in the Project

Both of these techniques have been applied in combination to improve the quality of biometric representation:

- LBP was used to analyze the texture of hand images, allowing details such as lines and skin roughness to be captured.
- HOG was used to detect the shape and contour of the hand, allowing the analyzed hand to be effectively localized.

4. Classification with Support Vector Machine

The extracted features, as described above, are fundamental for the implementation of Support Vector Machines (SVM) classifiers. SVMs are supervised machine learning models, particularly effective in contexts where the features have high dimensionality, but the training dataset is relatively small.

One of the main advantages of SVMs is their robustness to overfitting, thanks to the use of regularization techniques. Furthermore, compared to CNNs, SVMs are computationally lighter, both in the training and inference phases, making them particularly suitable for our project.

The main goal of a Support Vector Machine is to find an optimal hyperplane that separates data belonging to different classes in the best possible way.

To do this, the model:

1. Transforms data into higher dimensional space, if necessary, through the use of kernel functions.
2. Finds the hyperplane with the maximum margin between classes, i.e. the maximum distance between the closest points belonging to different classes (support vectors).
3. Assigns new instances to the corresponding class, based on their position relative to the optimal hyperplane.

Thanks to this methodology, SVM proves to be an effective model for the classification of complex data, such as that resulting from biometric images of hands.

4.1 Fusion of Predictions

In our biometric recognition system, we use images of both the dorsal and the dorsal of the hand to improve the classification accuracy. The idea behind prediction fusion is to combine the information extracted from both views to achieve more robust and reliable recognition.

This choice was adopted for several reasons:

- The palm contains clear details on the palmar lines and skin texture.
- The dorsal offers information on the shape of the hand and the distribution of the veins.

- Some people may have features that are barely distinguishable on one view, but more pronounced on the other.
- Fusion reduces the risk of errors due to environmental variations or in the image capture phase.

After extracting features from the palm and dorsal images, we train four separate SVM models:

- One SVM model for the palmar side, trained with LBP.
- One SVM model for the palmar side, trained with HOG.
- One SVM model for the dorsal side, trained with LBP.
- One SVM model for the dorsal side, trained with HOG.

Each model returns a classification probability, called score, for each subject. Fusion is performed by combining these scores with a specific weight for each view. This approach allows us to choose and weigh the most accurate views more heavily.

This fusion strategy increases the reliability of the system, allowing it to compensate for any deficiencies in one of the two views and improving the overall accuracy of the recognition.

5. Performance evaluation

5.1 Evaluation metrics used

After obtaining the combined prediction through the fusion of palm and back probabilities, we evaluated the effectiveness of the system using several key metrics to measure recognition performance.

More specifically, the metrics adopted are:

1. Accuracy:
    - It represents the percentage of correct identifications out of the total tests carried out.
    - Calculated as $\frac{TP + TN}{TP + TN + FP + FN}$
    - Where:
        - TP (True Positive) = Correct identifications of authorized users.
        - TN (True Negative) = Correct rejections by unauthorized users.
        - FP (False Positive) = Misidentifications of unauthorized users.
        - FN (False Negative) = Incorrect rejections by authorized users.
2. Confusion Matrix
    - It provides a detailed view of the model's performance, showing how predictions are distributed between correct and incorrect classes.
    - It allows to identify the main causes of recognition errors.
3. FAR (False Acceptance Rate)

- Indicates the probability that a system will erroneously accept an unauthorized user, i.e. a false positive.
- It is particularly important in security contexts, where high FAR can lead to unauthorized intrusions.
- Calculated as $\frac{FP}{FP + TN}$

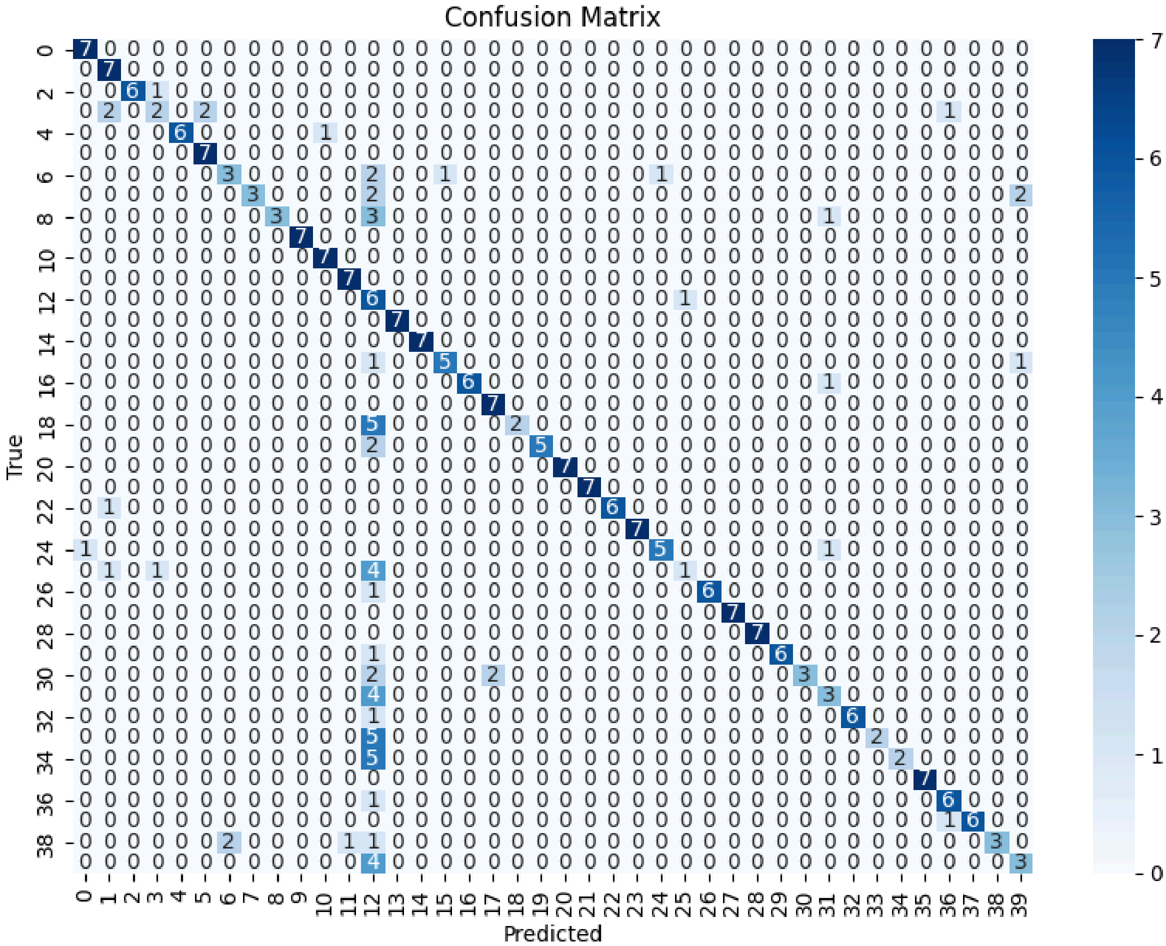
4. FRR (False Rejection Rate)
    - Indicates the probability that the system will erroneously reject an authorized user, i.e. a false negative.
    - A high FRR compromises the usability of the system, causing difficulty in accessing authorized users.
    - Calculated as $\frac{FN}{FN + TP}$
5. CMC (Cumulative Match Characteristic)
    - It measures the cumulative probability that an individual's correct identity is found within the first k positions in an ordered list of candidates.
    - This parameter is useful in biometric systems where an individual must be identified within a database.
    - A CMC curve with high values indicates a system with high identification reliability.

5.2 Results obtained and performance analysis

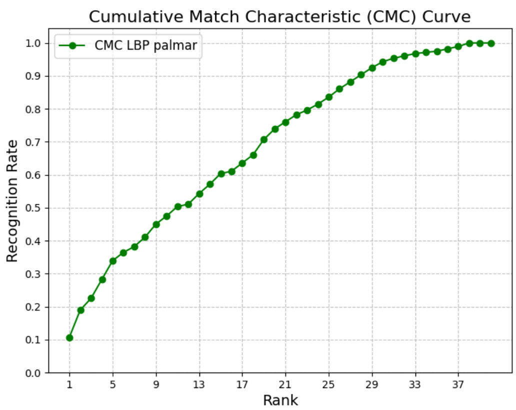
The implemented identification model works for both open set and closed set identifications. The scenario in which it is assumed that each individual to be identified is already present in the database is called closed set identification. In this scenario the system compares the biometric characteristic with those stored and always returns a match. However, the scenario in which the system does not assume that every user is present in the database is called open set identification. The recognition can therefore include unknown people, and the system must decide whether to accept or reject a user. In closed-set systems, the main risk is to misidentify a person, in open-set systems, it is crucial to avoid recognizing individuals not present in the database, which leads to challenges such as similarity thresholding.
The results for both types of identification are reported and analyzed below.

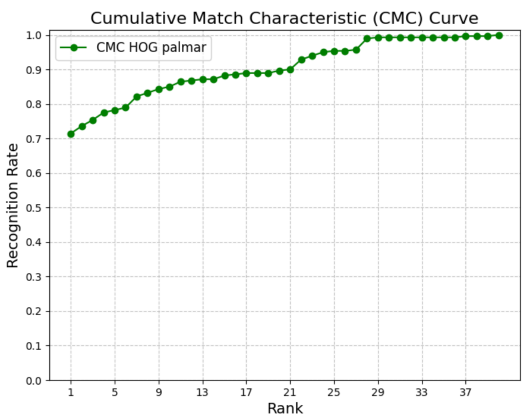
5.2.1 Closed Set Identification

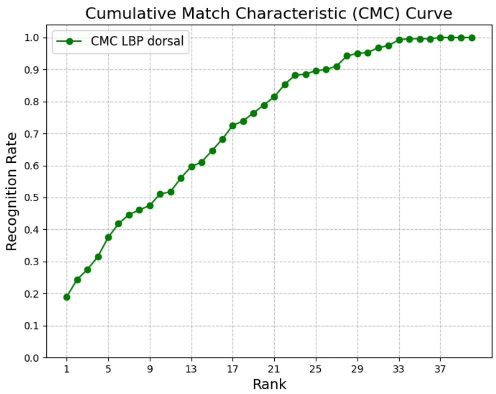
The first analyzed metric in closed set identification is the confusion matrix. This is useful for identifying any imbalances in the used dataset, identifiable as a certain identity that is predicted several times. A perfect confusion matrix has recognitions only on the main diagonal. The following is the confusion matrix obtained with a set of 40 subjects and 25 images:

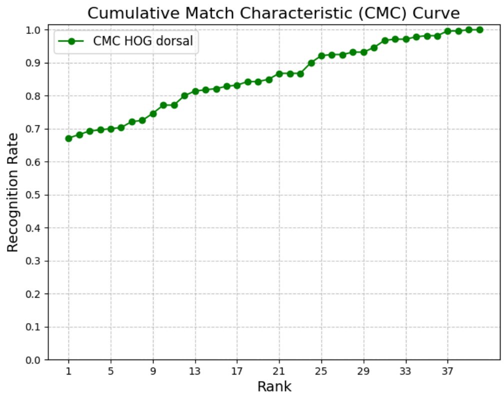


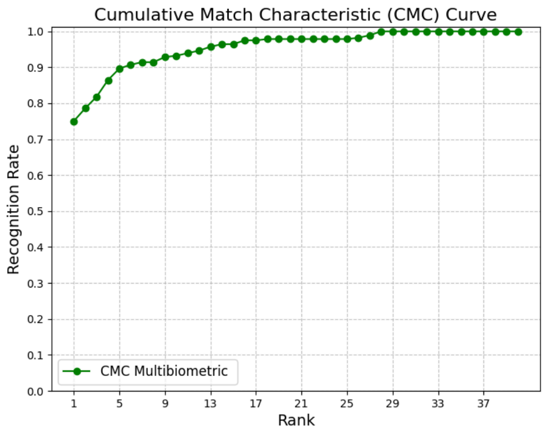
It is interesting to note that despite having a fairly homogeneous matrix on the diagonal (indicating a high number of correct recognitions) there is one identity predicted much more often than the others, number 8, predicted a total of 59 times, a sign of a strong bias which is most likely the result of an imbalance in the dataset.
The second metric used in closed set identification is the CMC curve, i.e. the cumulative graph of the probability that the correct identity of an individual is among the first k matches returned by the biometric system. In our case the CMC curve was calculated for each of the 4 identifications made via SVC and for the final multibiometric recognition. The results are as follows:











The CMC curve shows how identification is highly inaccurate for most test cases. While the performance of the system is good in the case of features extracted via HOG on the palm and dorsal sides of the hand (with a rank 5 accuracy of almost 80% in the first case), it drastically worsens in all other cases, with the worst case being that of identification via LBP features on the palm, where the model reaches 100% correct identifications among the first k predictions with k equal to 38 starting from an accuracy of 10% and increasing very slowly. Finally we can observe how the system behaves relatively well in multi-biometric recognition, with an accuracy of 90% at rank 10.

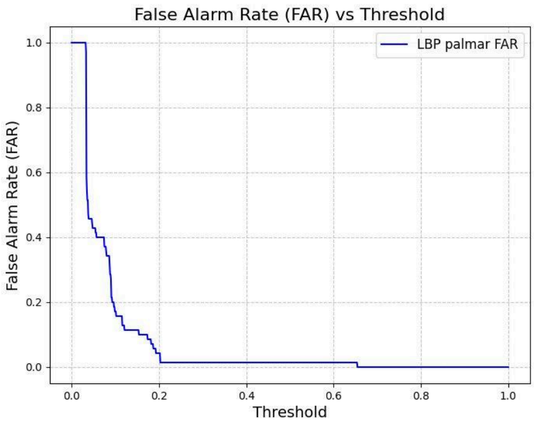
Finally, the last metric used for the evaluation of the biometric system is the accuracy, calculated on each of the 4 test cases and on the final multi biometric prediction. The results obtained are shown in the following table:

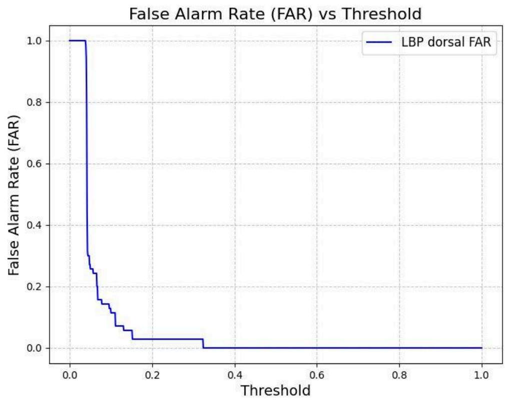
Test Case	Accuracy
LBP Palmar	10.2%
LBP Dorsal	18.7%
HOG Palmar	71.8%
HOG Dorsal	67.3%
Multi Biometric System	75.0%

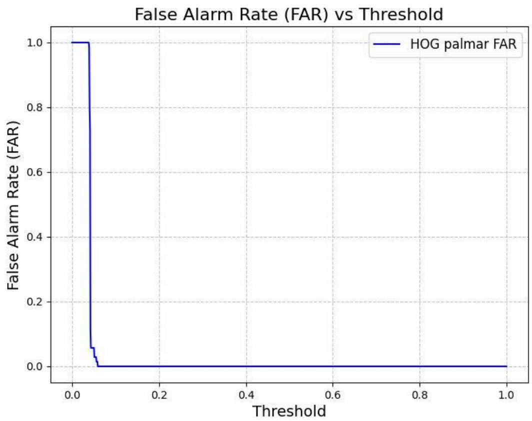
These results clearly highlight the usefulness of multi-biometric identification, which manages to achieve an accuracy of 75% even in the face of terrible accuracy of some of the models on which it is based.

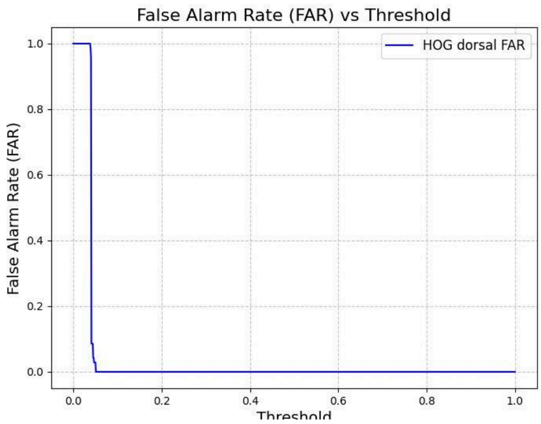
5.2.2 Open Set Identification

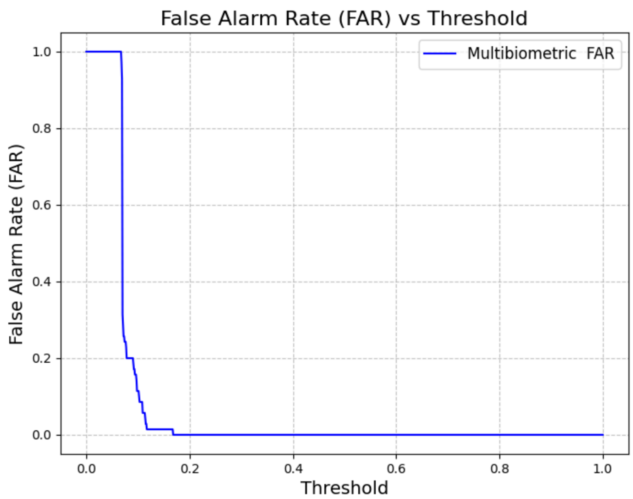
The first metric used in the analysis of the performance of open set identification is the FAR (False Alarm Rate) curve. It is related to the threshold used during the final prediction and describes the relationship between the acceptance threshold and the probability of generating false alarms. It is calculated for each of the 4 SVC identifications and for the final multi biometric identification and is as follows.



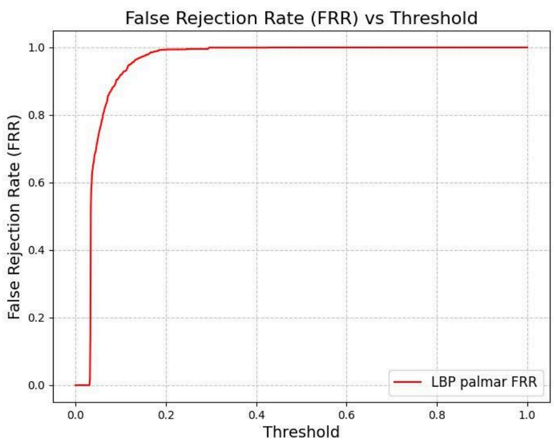


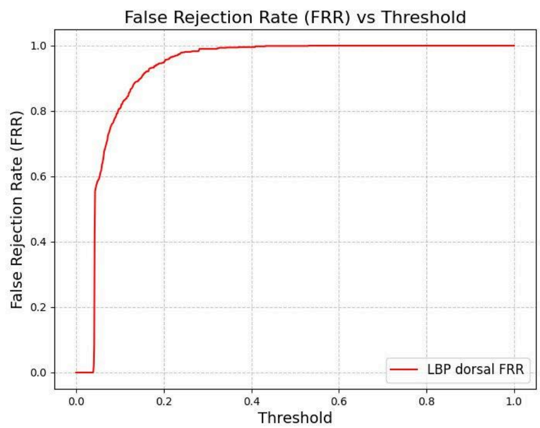


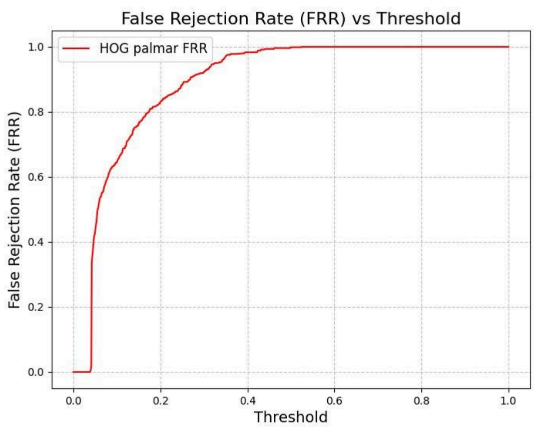


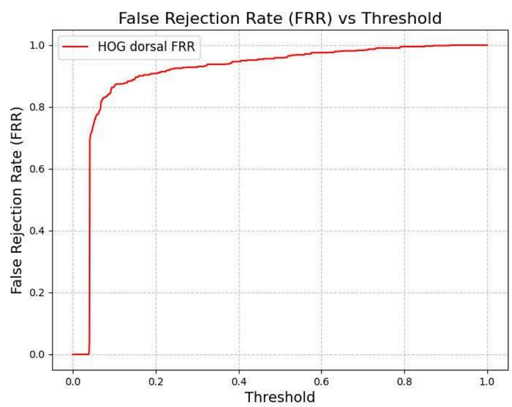


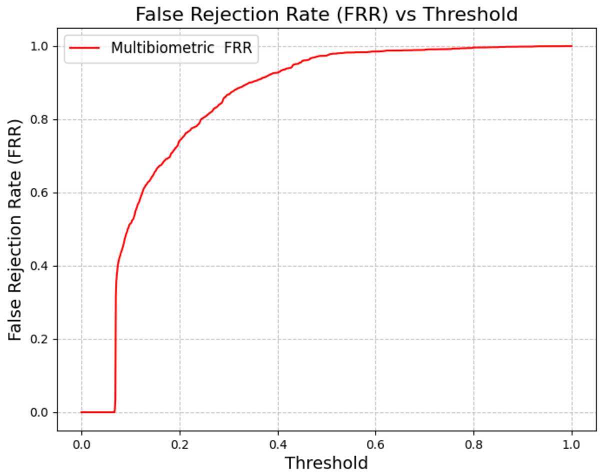
As can be seen from the graphs, the curve mostly has a trend characterized by sharp drops, indicating the fact that the model abruptly goes from being very permissive to being very restrictive. This may be due to a balance issue or dependency on a certain feature.
The second metric used to analyze the open set model is the curve formed by the relationship between FRR (False Rejection Rate) and Threshold, representing the relationship between the acceptance threshold and the probability that a legitimate user is rejected by the biometric system. The graphs returned by the model are as follows:



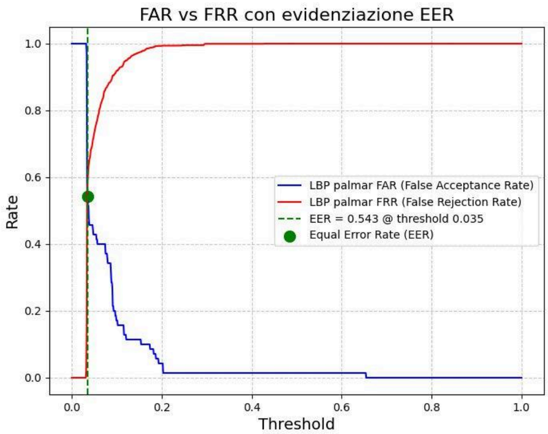


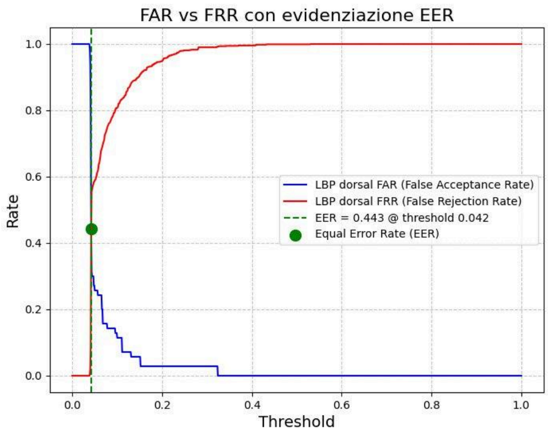


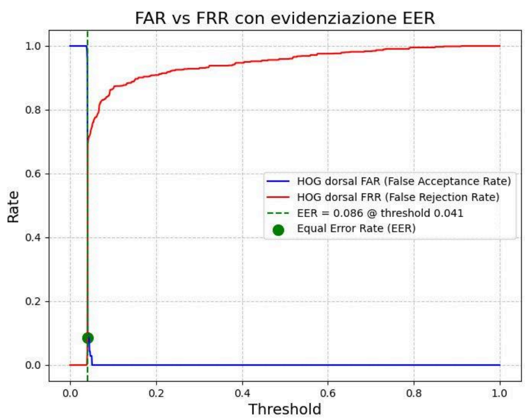


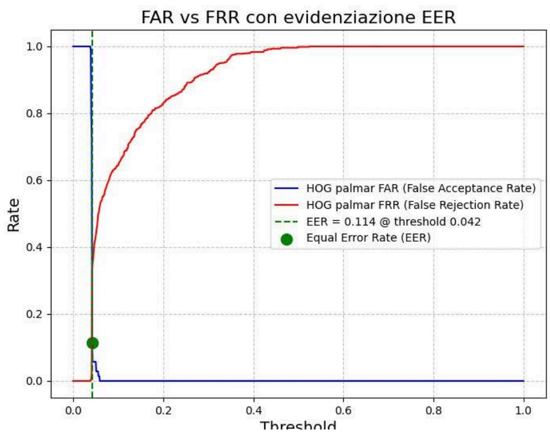


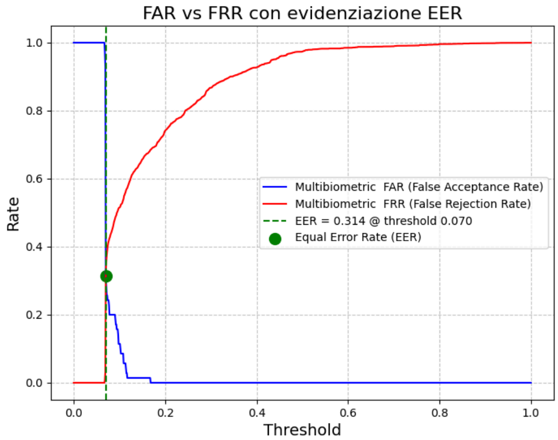
Here too, as for the previous metric, we note a mostly abrupt trend followed by a plateau. This indicates a sudden shift from a state in which all legitimate users are accepted to one in which almost none are accepted. This can indicate multiple problems, including poor model generalization or a poorly calibrated distribution of similarity scores. This trend is milder in the curve relating to multi-biometric recognition, demonstrating its potential effectiveness.
A further metric to evaluate the model is given to us by a combination of the previous ones, i.e. the relationship between FAR and FRR for each identification. These two metrics are inversely proportional, and the point where they meet (called EER - Equal Error Rate) is an excellent index of system balance: the lower the EER, the more precise the biometric recognition.











In cases where features are extracted via LBP the graphs present an average EER of around 0.5, indicating that the system is almost drawing at random regarding the decision between authentic users and impostors. This indicates that the model is not sufficiently capable of separating the two classes.
In cases where the features are extracted via HOG, however, the graphs present an average EER of 0.04, which indicates that the system is able to separate genuine users from impostors well with few errors.
Finally, in the case of multi-biometric recognition, the graph presents an EER of 0.314, i.e. halfway between the two previous cases.
Finally we can measure the accuracy between the various models:

Model	Accuracy
LBP Palmar	13.5%
LBP Dorsal	21.2%
HOG Palmar	66.1%
HOG Dorsal	31.3%
Multi Biometric System	68.9%

Also in this case we can notice a clear improvement in biometric recognition in the case of the multi-biometric model, proving its effectiveness compared to other models.

6. Possible Improvements and Future Developments

Significant improvement could come from optimizing the parameters of the SVM classifier, for example by experimenting with different kernel types to improve data separability. Furthermore, the integration of feature selection techniques could reduce the dimensionality of the data, while increasing computational efficiency.

The use of a second soft biometric trait such as gender, identified through CNN, in order to add a second verification threshold.

A soft biometric trait, such as gender, cannot be used for actual identification given its non-uniqueness but can be used to allow greater accuracy in the recognition phase.

Another promising development concerns the combined use of traditional approaches and neural networks. Manual feature extraction could be exploited to pre-process data, providing more structured input to CNNs and reducing the need for large datasets. This hybrid combination could improve accuracy while maintaining moderate computational complexity.

Finally, adapting the system for real-time applications and embedded devices represents an important direction. Hardware optimizations, such as the use of specialized GPUs, could make biometric recognition faster and more efficient.

7. Conclusions

The project demonstrated that a biometric recognition system based on hand-crafted features (LBP and HOG) combined with an SVM classifier can potentially provide acceptable performance, with advantages in terms of interpretability and reduced computational complexity. However, it has been widely highlighted that the suggested implementation suffers from a very serious dataset balancing problem which negatively impacts the quality of biometric recognition.

It was then demonstrated how a multi-biometric recognition system behaves in a decidedly superior manner compared to a single biometric trait system, constantly managing to outperform the best single biometric trait system even in cases of terrible performance of the other models on which it is based.

The developed system could potentially be used in several applications, including:

- Access control in secure environments.
- Biometrics-based authentication systems.
- Automated identification in forensics.

Thanks to its computational lightness, the system could be further developed to operate in embedded contexts, broadening its field of application in biometric hand recognition.

8. Sources

1. *11K Hands: Gender recognition and biometric identification using a large dataset of hand images - 2018 Mahmoud Afifi*
2. *Gender recognition and biometric identification using a large dataset of hand images - 2017 Mahmoud Afifi*

3. *HandNet: Identification Based on Hand Images Using Deep Learning Methods - 2021 Yimin Yuan et al*
4. *Identity Verification by Using Handprint - 2007 Ying Hao et al*