PROJECT OF DIGITAL SIGNAL AND IMAGE MANAGEMENT

Salvatore Rastelli 903949
Marco Sallustio 906149

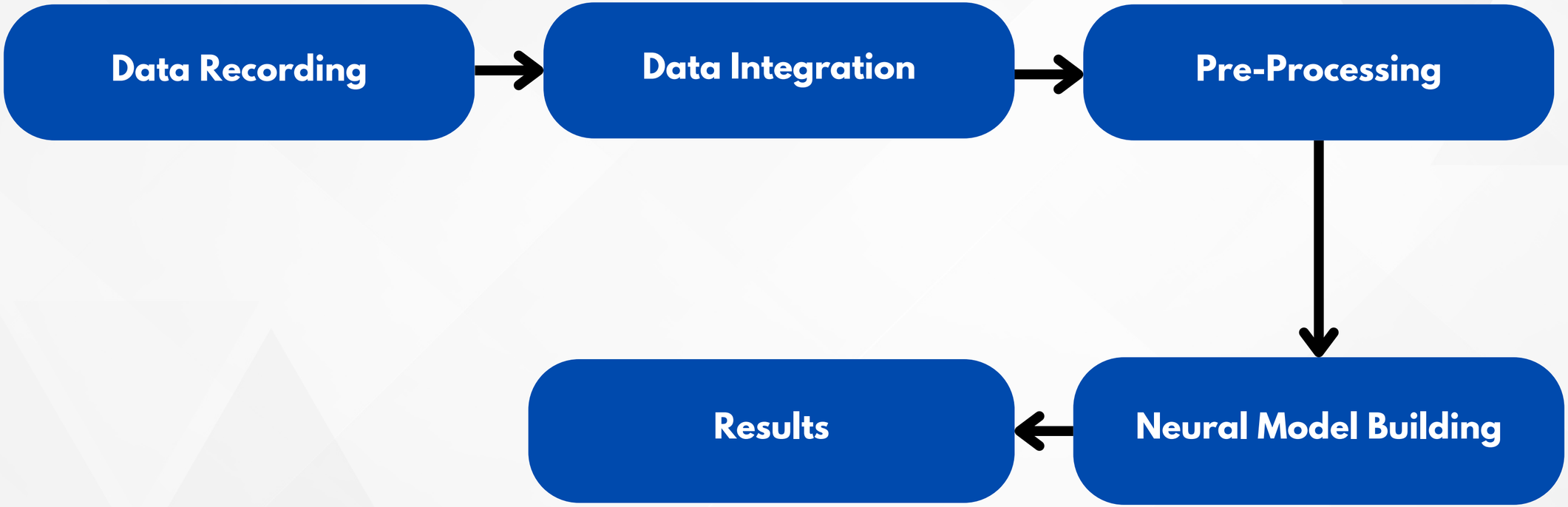
MAIN GOALS

MONO-DIMENSIONAL SIGNAL PROCESSING
Multi-Class Classification for Speaker & Digit Recognition

BI-DIMENSIONAL SIGNAL PROCESSING
Multi-Class Classification for Face Recognition

RETRIEVAL
Image Retrieval of the 10 most similar images given as input

MONO-DIMENSIONAL



DATA RECORDING AND INTEGRATION

- 50 recordings per group member of every digit from 0 to 9
- Trimmed every recording removing the silences before and after the pronunciation of the number
- Enriched using registrations from the Dataset **Free Spoken Digits**, labeling the speakers as “Unknown”

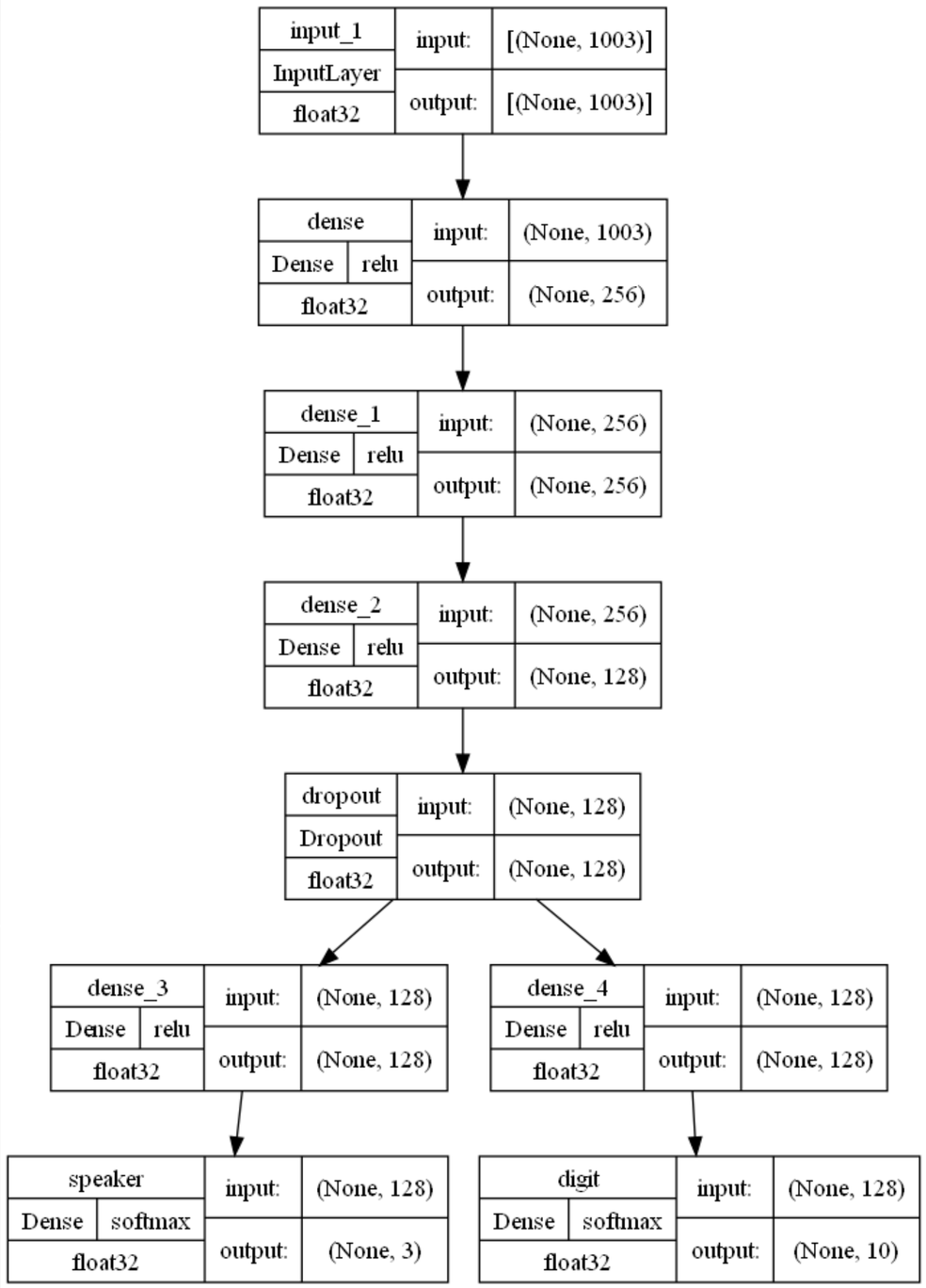
PRE-PROCESSING

We extracted differents feature to feed the neural networks builded:

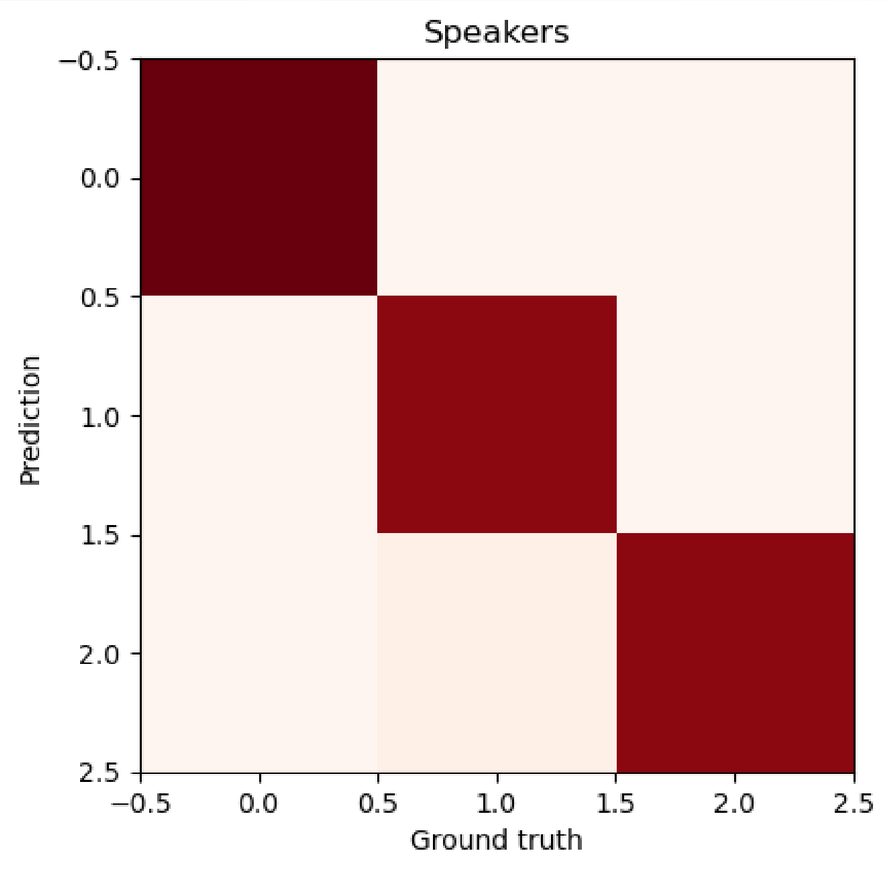
- For the **CNN**, we used a combination of *zcr, standard deviation, energy* and a *flattening of the Mfcc*.
- For the **RNN**, we used the complete *Mel Frequency Spectrogram*

CNN

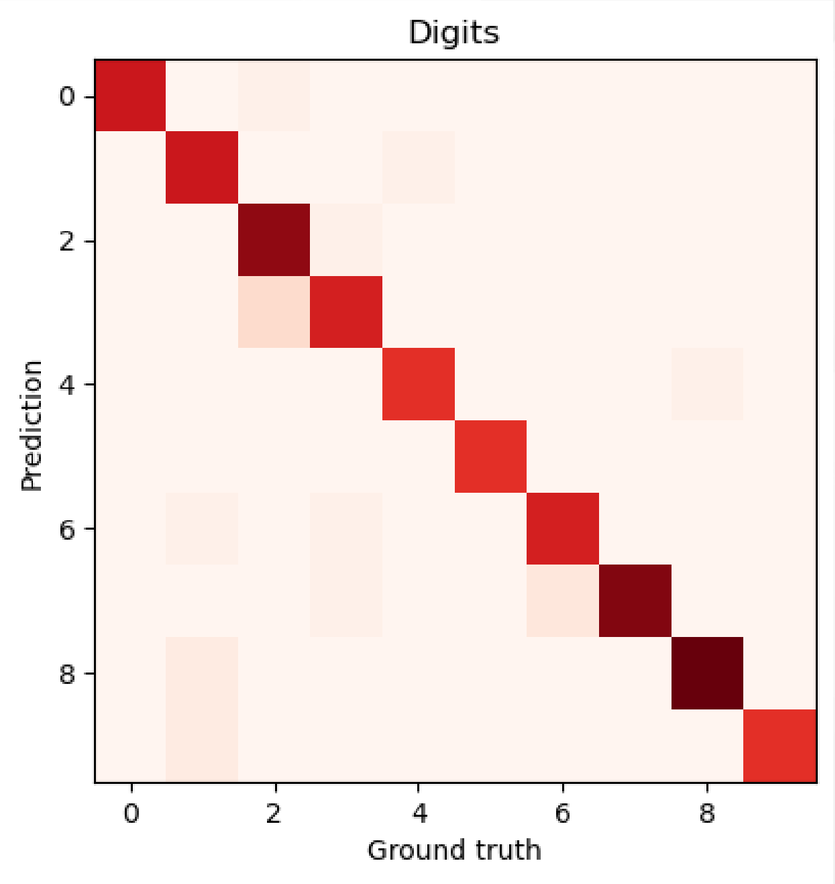
- **Optimizer**: Adam
- **Loss**:
  - *Speaker*: Categorical Cross Entropy
  - *Digits*: Sparse Categorical Cross Entropy
- **Metrics**:
  - *Speaker*: Categorical Accuracy
  - *Digits*: Accuracy



CNN: Results



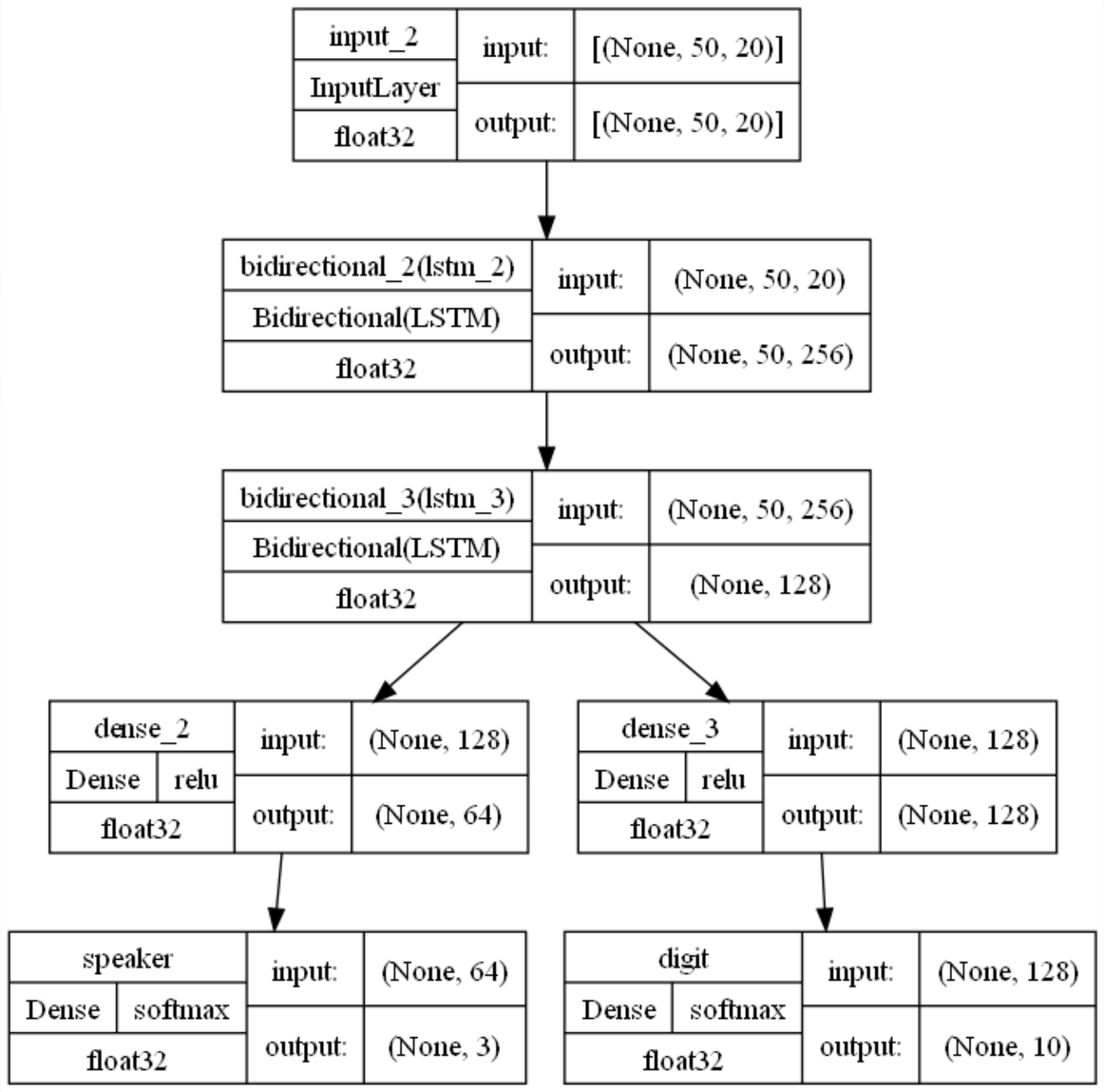
Accuracy: 0.987



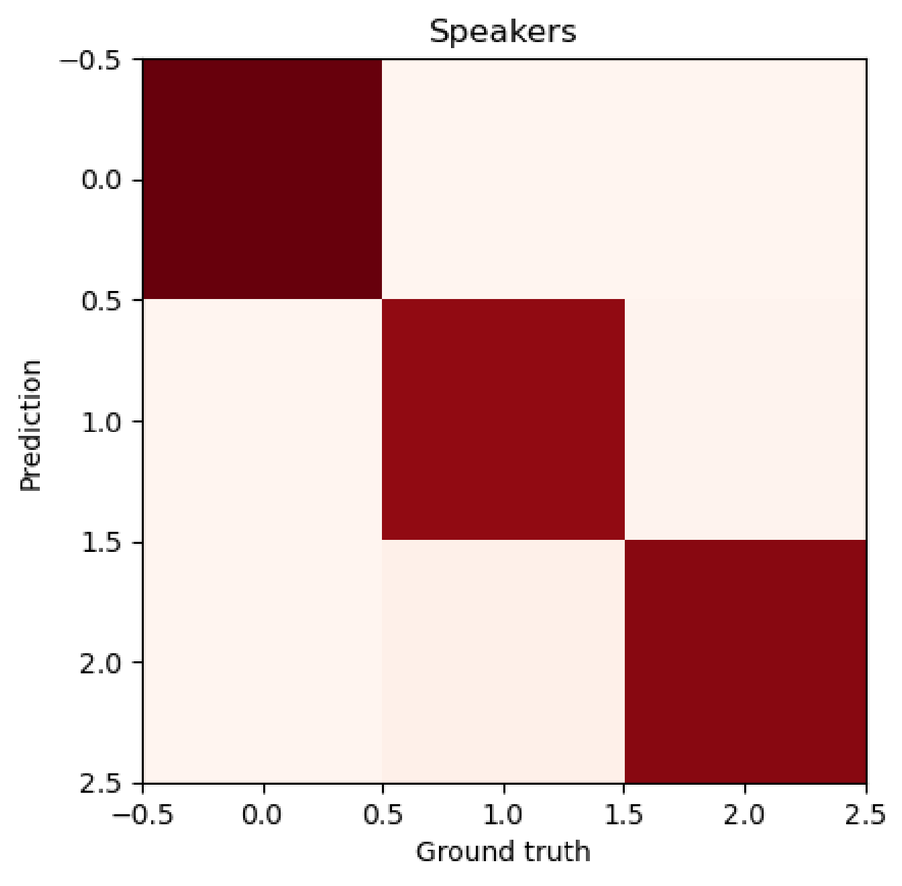
Accuracy: 0.937

RNN

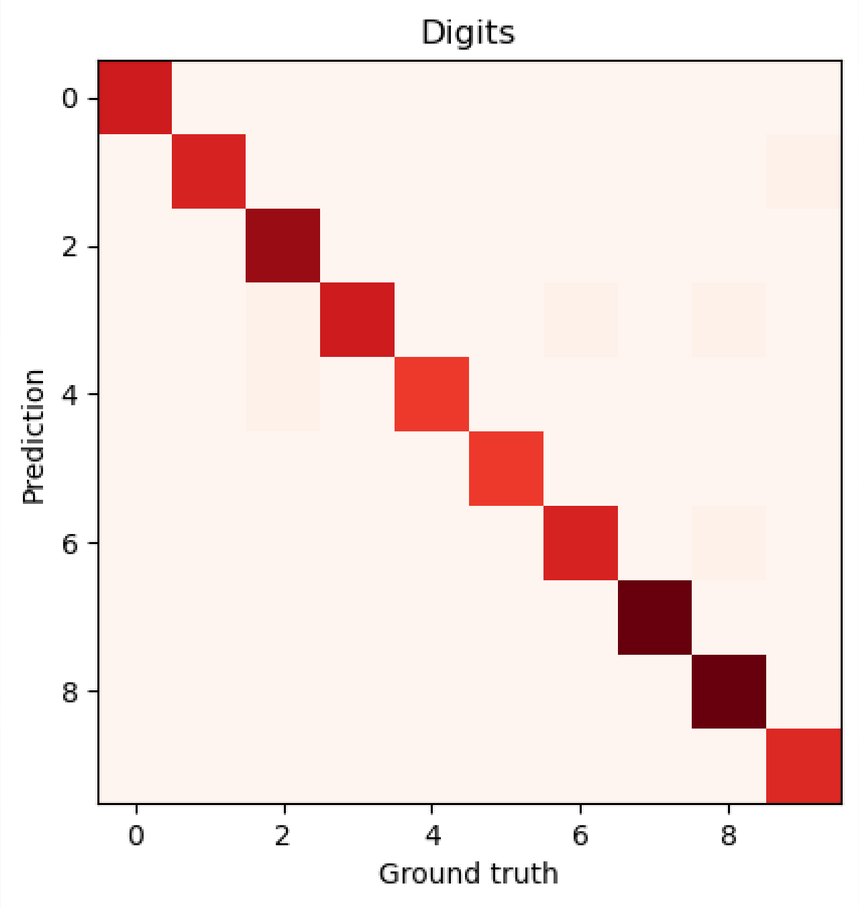
- **Optimizer**: Adam
- **Loss**:
  - *Speaker*: Categorical Cross Entropy
  - *Digits*: Sparse Categorical Cross Entropy
- **Metrics**:
  - *Speaker*: Categorical Accuracy
  - *Digits*: Accuracy



RNN: Results



Accuracy: 0.987



Accuracy: 0.980

BI-DIMENSIONAL

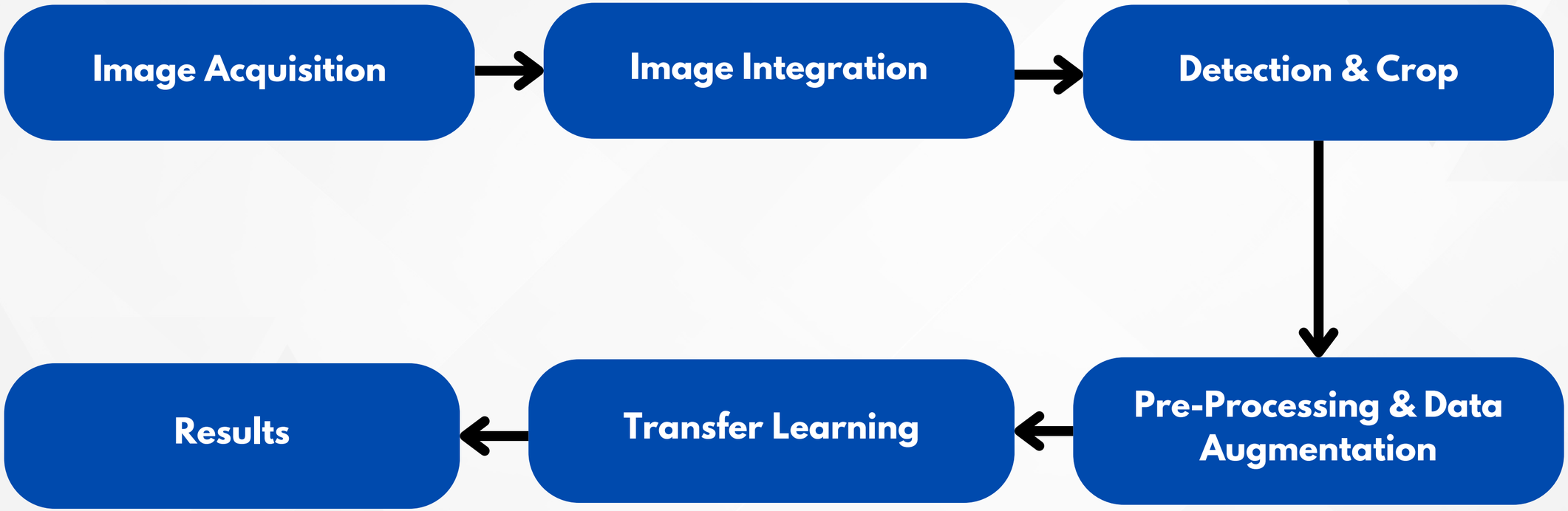
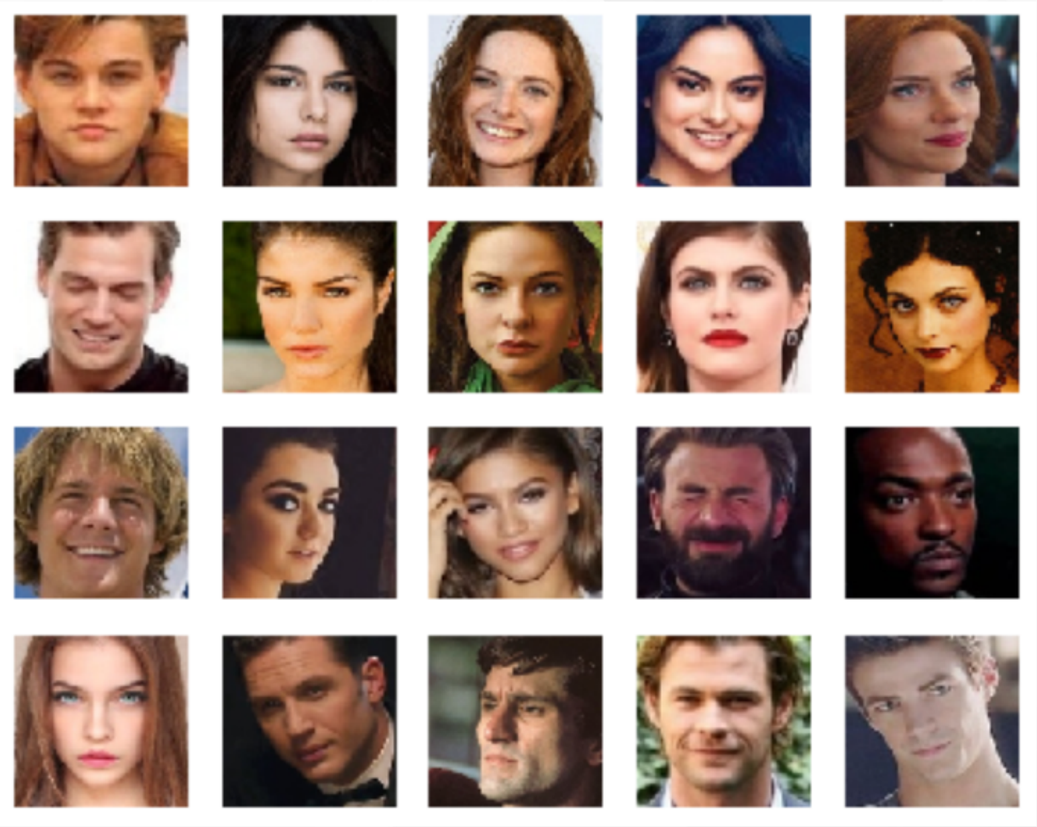


IMAGE ACQUISITION

- To tackle the task it was necessary to take some photos of the group members.
- Acquisition via Webcam every half second, using **OpenCV**
- Approach:
    1. *acquisition of 2000 photos per subject*
    2. *different light conditions*
    3. *different poses*
    4. *different environments*

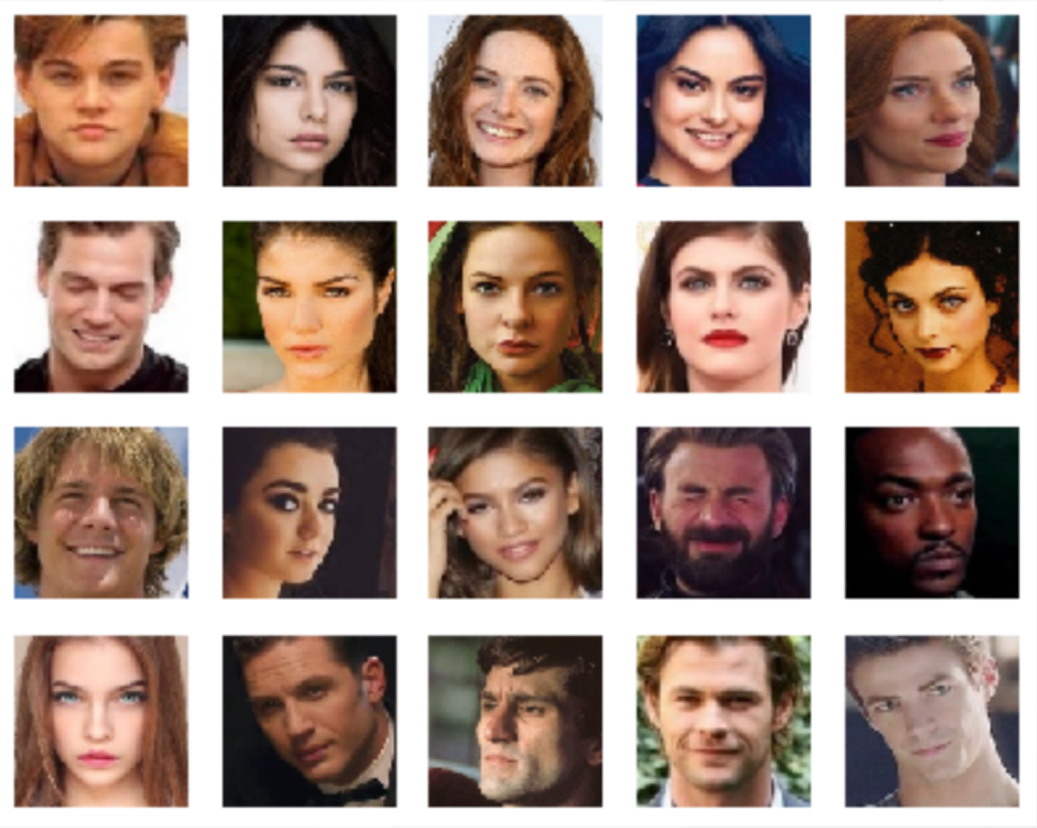
IMAGE INTEGRATION

- To prevent all the faces subjected to the model from being traced back to the two members of the group, we integrated a dataset containing photos of various well-known people extracted from Pinterest (**Cropped Pinterest Dataset**)
- We randomly chose some of these photos and assigned the label "**Unknown**"



FINAL DATASET

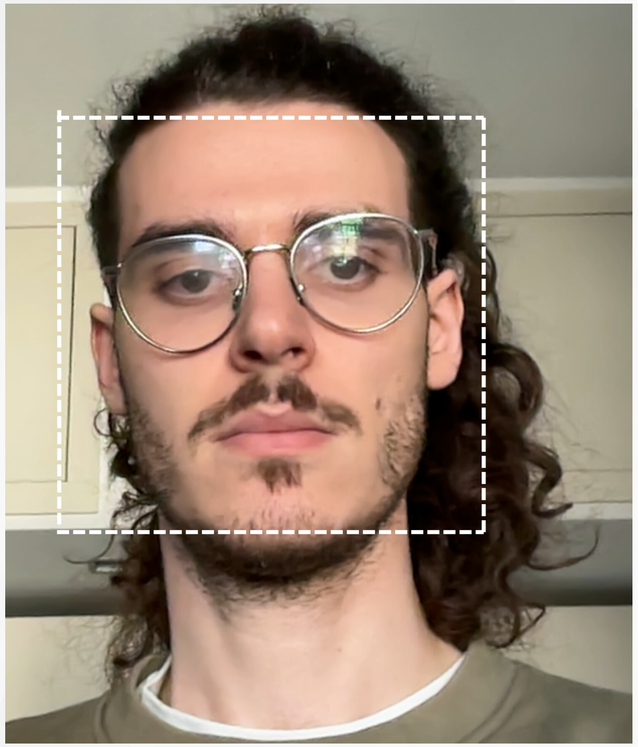
+



CROPPING

- We used the **MTCNN** model to detect the face within the captured image
- MTCNN (*Multi-Task Cascaded Convolutional Networks*) is a library used for detecting faces and facial features in images. It is a convolutional neural network (CNN) based approach that addresses 3 different steps of face detection in a sequential manner:
  - Face Detection
  - Facial Landmark Alignment
  - Face Detection Refinement
- After detecting the face, we cropped each image based on the bounding box coordinates extracted from MTCNN

CROPPING



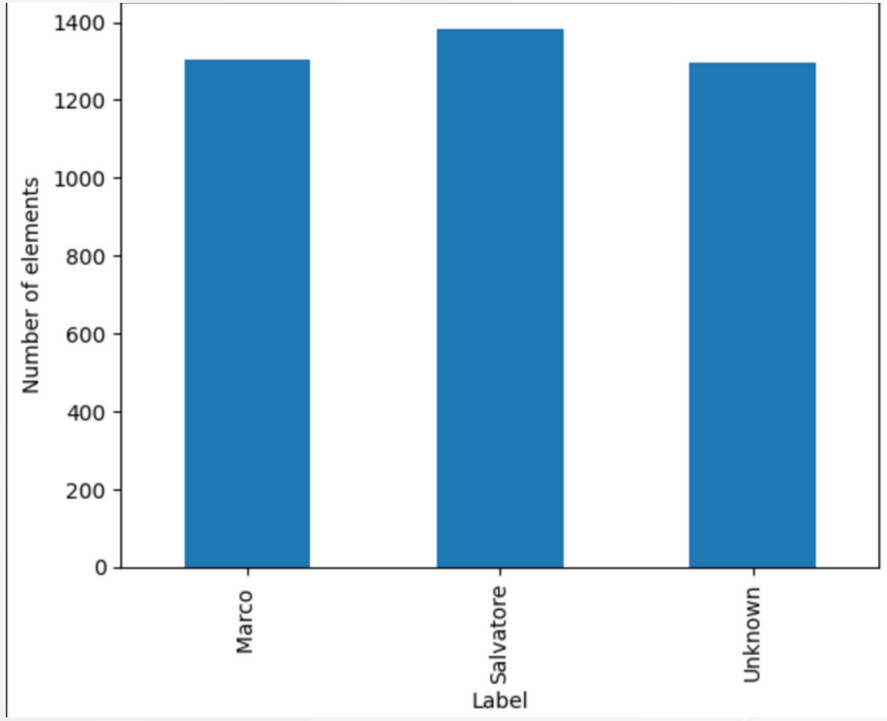
We created a function that automatically crops all the photos and discards when:

- no face was detected
- the size of the detected face is too small (width,height<50)

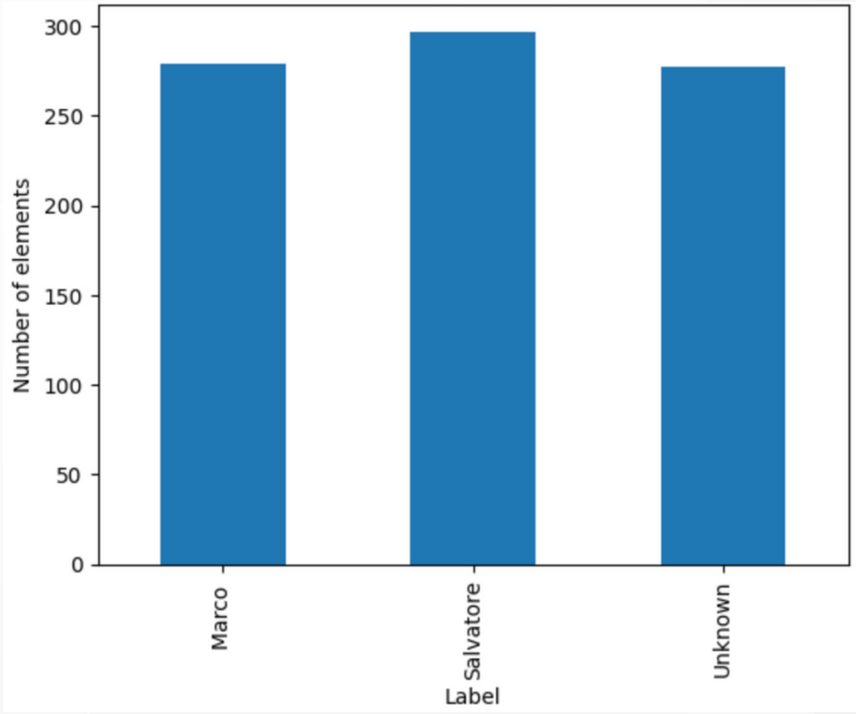
TRAIN/TEST/VALIDATION SPLIT

After the crop of the images, were obtained 5689 images which were divided into *Training, Test and Validation set*, with percentages of 70%, 15%, 15% respectively. The distribution of the classes in the three sets is shown below:

TRAIN:



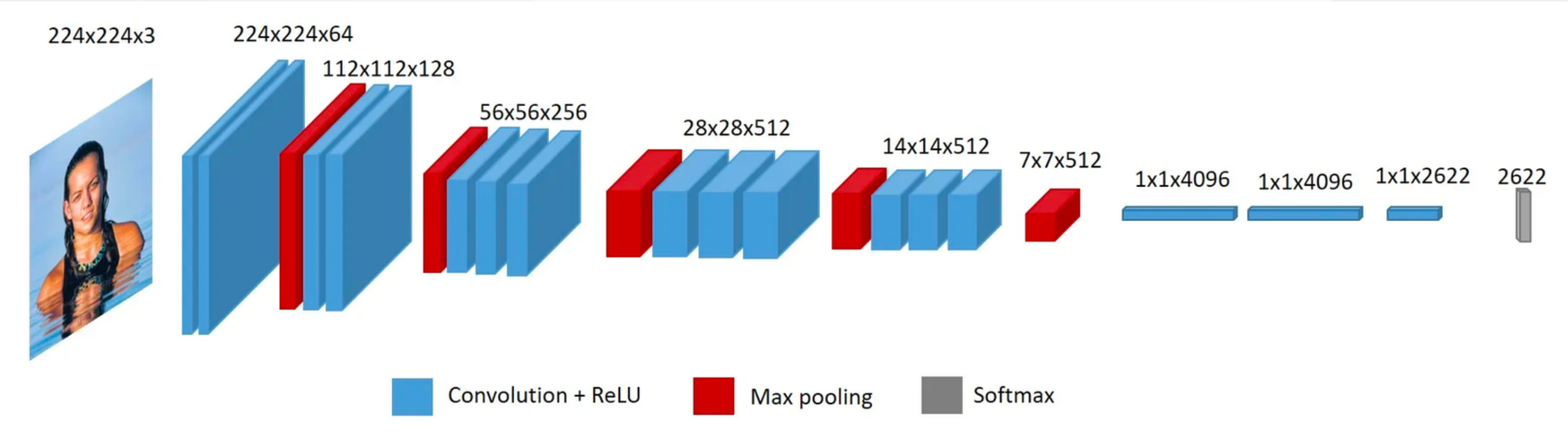
TEST & VAL:



DATA AUGMENTATION

- We have implemented a function designed to create data generators for training and validating neural network models in TensorFlow using **Keras' ImageDataGenerator module.**
- This module allows you to perform *data augmentation* during training, i.e. apply random transformations to images to improve the generalization of the model.
- A normalization is also applied to the pixel intensity of the image(1./255)
- The following transformations were performed:
  - Horizontal Flip
  - 15 degree rotation
  - Changing the brightness
  - Zoom

TRANSFER LEARNING



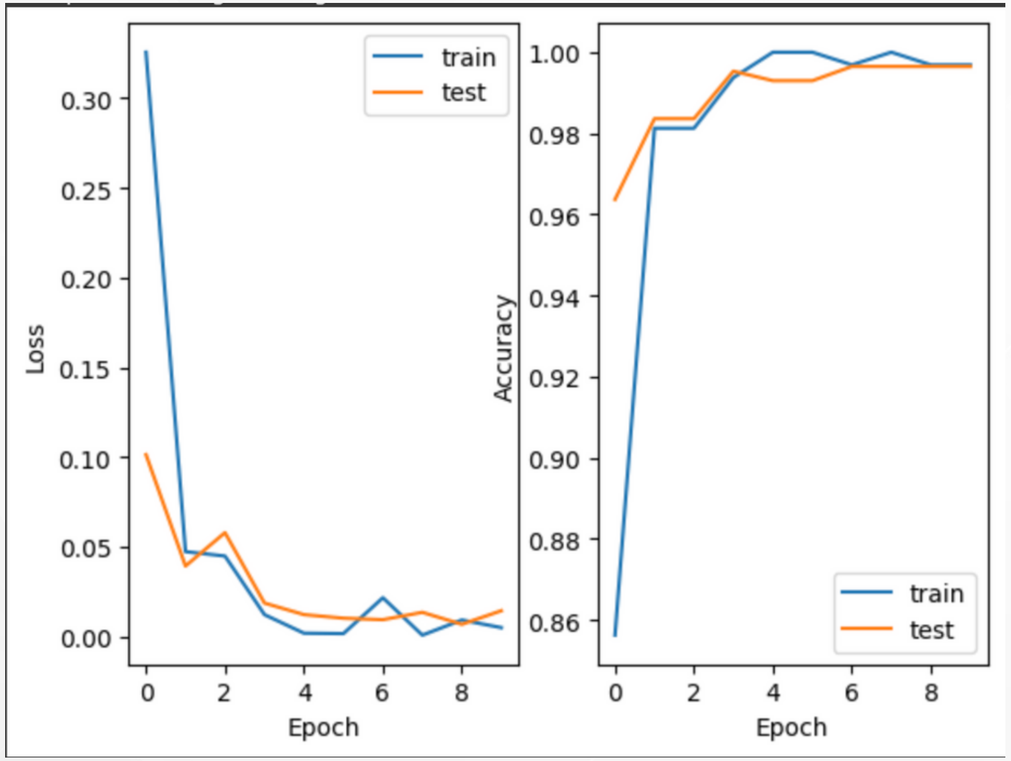
For the final classification we used the **VGGFace** network to which we added the following layers:

- Flatten layer to flatten the output of the pooling layer
- Two fully connected Dense layers with *ReLU* activation
- Dense layer with *softmax* activation for the output

RESULTS

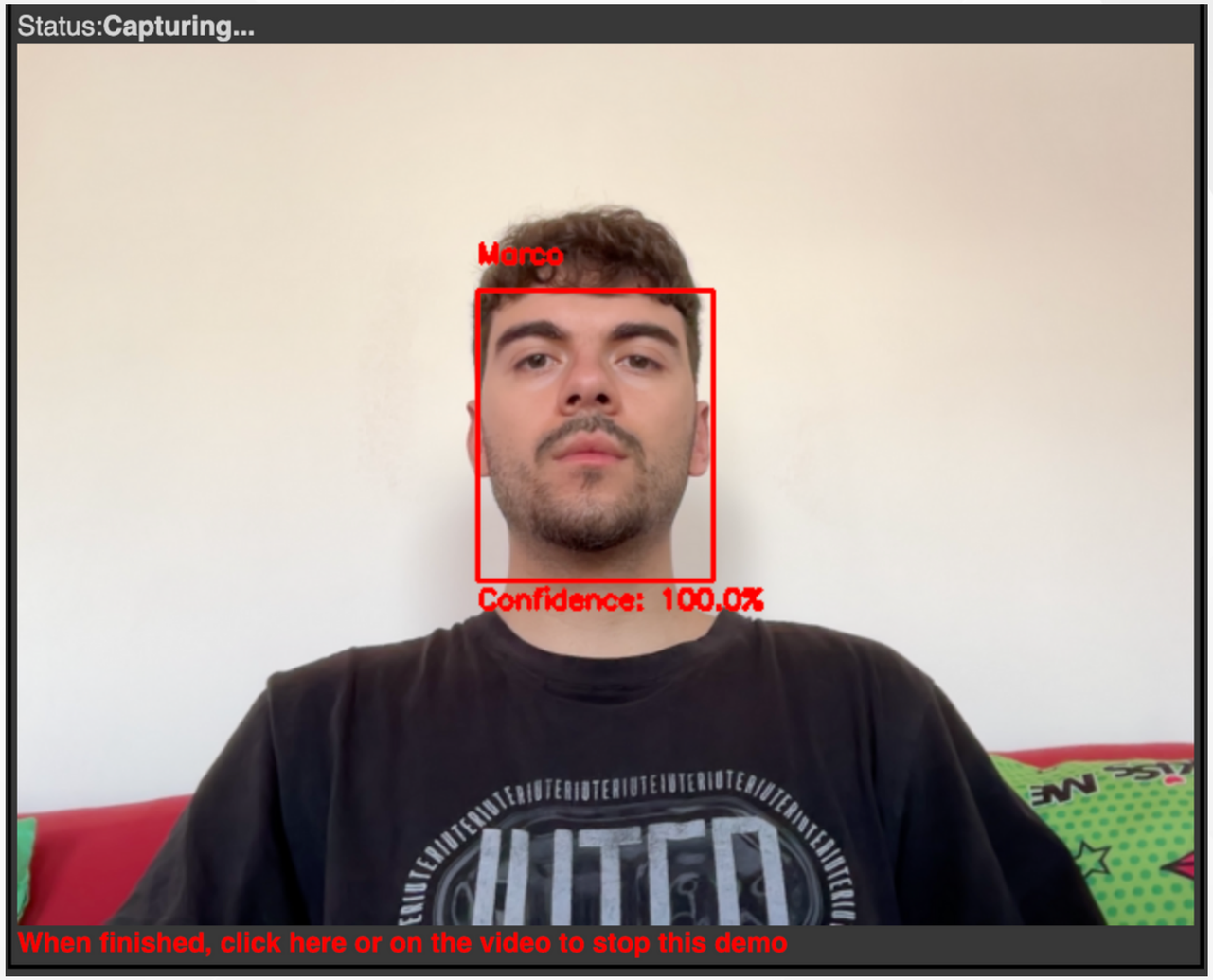
The network was trained with the following parameters:

- **Epochs**=10
- **Optimizer**=Adam
- **Loss**=Categorical Cross Entropy



	precision	recall	f1-score	support
Marco	0.99	1.00	1.00	279
Salvatore	1.00	0.99	1.00	297
Unknown	1.00	1.00	1.00	277
accuracy			1.00	853
macro avg	1.00	1.00	1.00	853
weighted avg	1.00	1.00	1.00	853

RESULTS


Status:Capturing...
Marco
Confidence: 100.0%
When finished, click here or on the video to stop this demo

FAILED ATTEMPT

Before using the VGGFace network we also used the **InceptionResNetV2 (IncRes)** network to which we added the following layers:

- Flatten layer to flatten the output of the IncRes
- Dropout layer to reduce overfitting
- Dense layer with softmax activation

This model also obtained excellent performances but the reason why it was not chosen was because we obtained predictions that were not always correct.
In particular, Marco's face was often classified as Unknown.

RETRIEVAL

Feature Extraction → **Query Matching** → **Precision Evaluation**

Database created from the **Cropped Pinterest Dataset**, picking 10 Images for each Vip →

Two model tested for the feature extraction:

- **MobileNet**
- **VGG-Face**

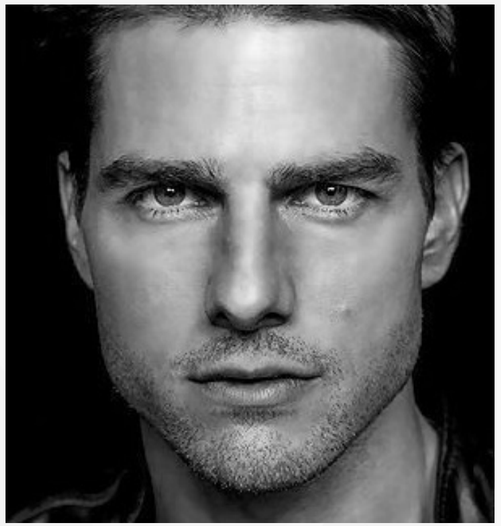
→

Implementation: **KDTree** to extract the 10 nearest neighbors of the image of input

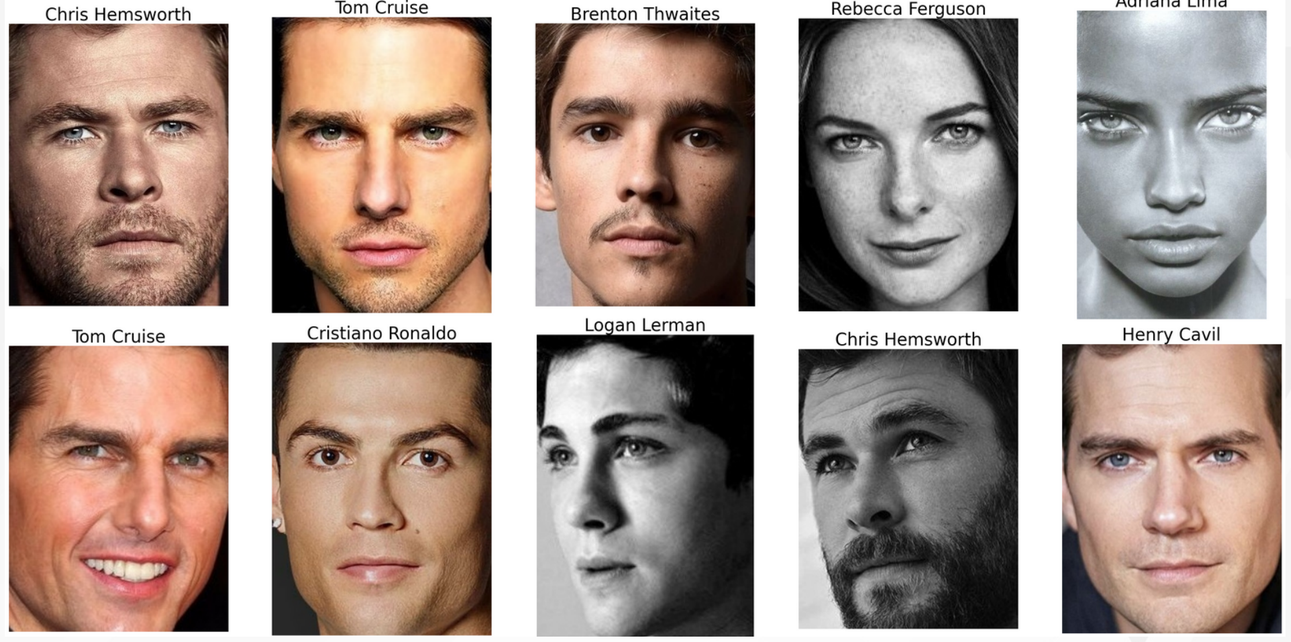
↓

Chosed the one with the best precision

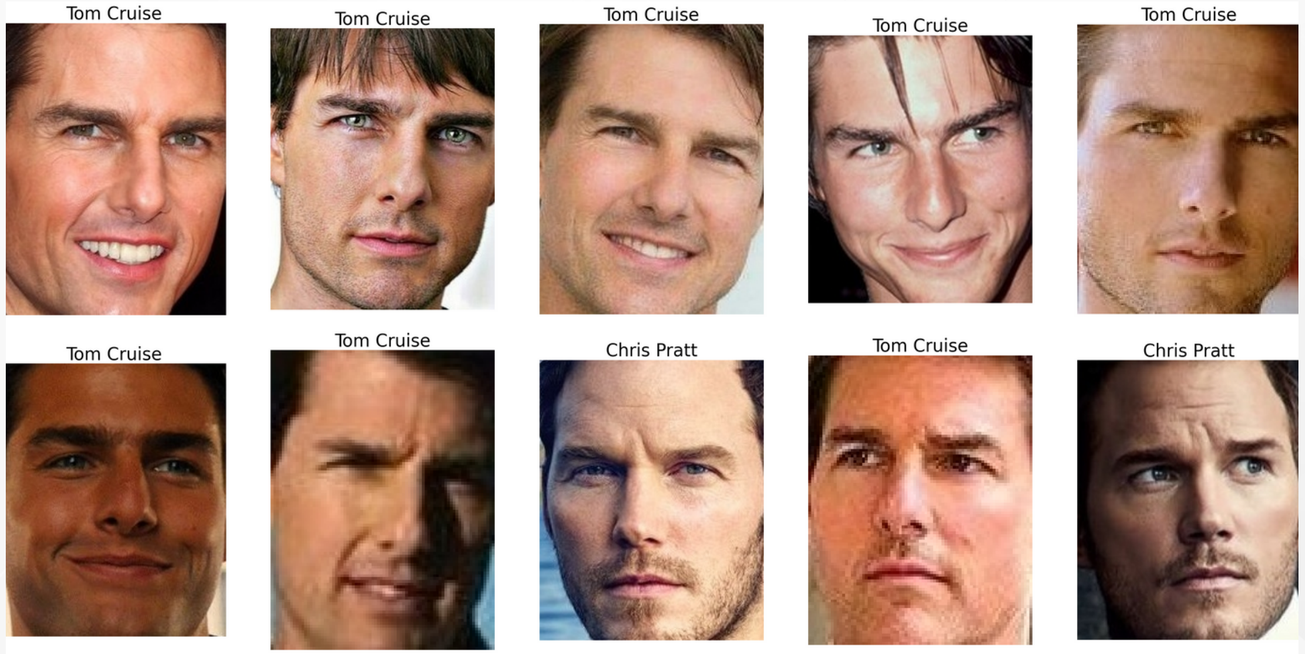
PRECISION



MobileNet



VGG-Face



Precision 2/10

Precision 8/10

PRECISION

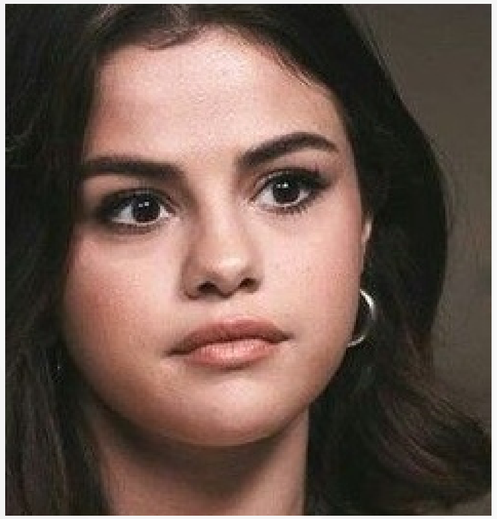
MobileNet

Precision 1/10

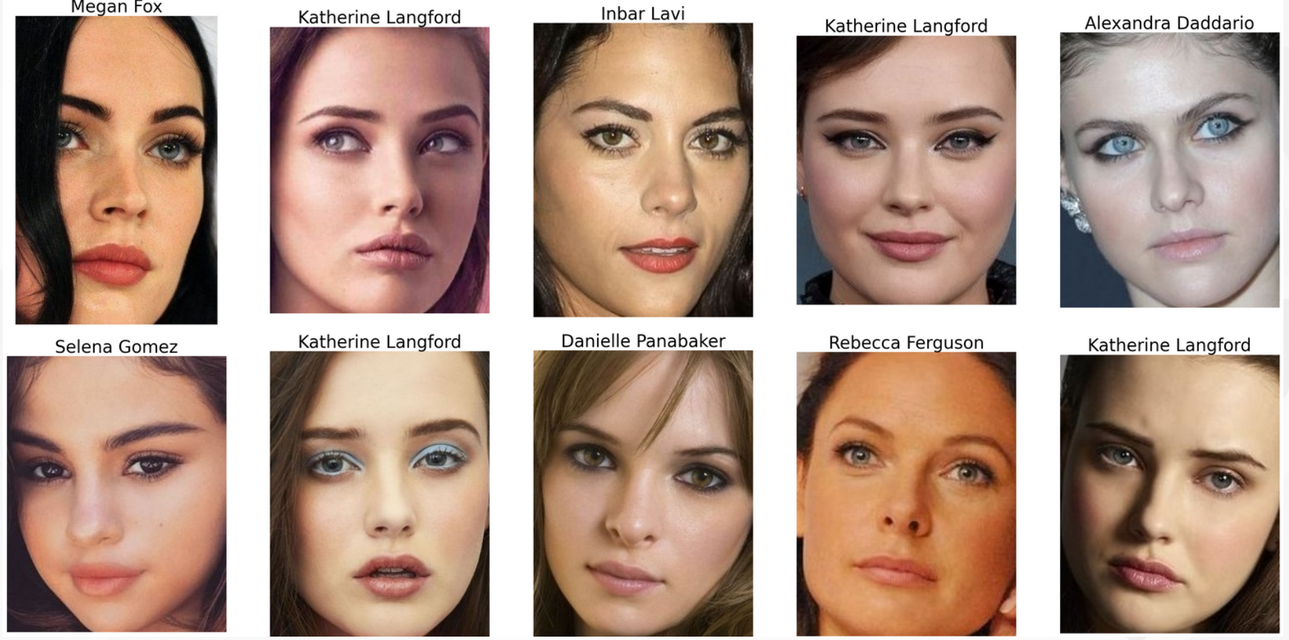
VGG-Face

Precision 9/10

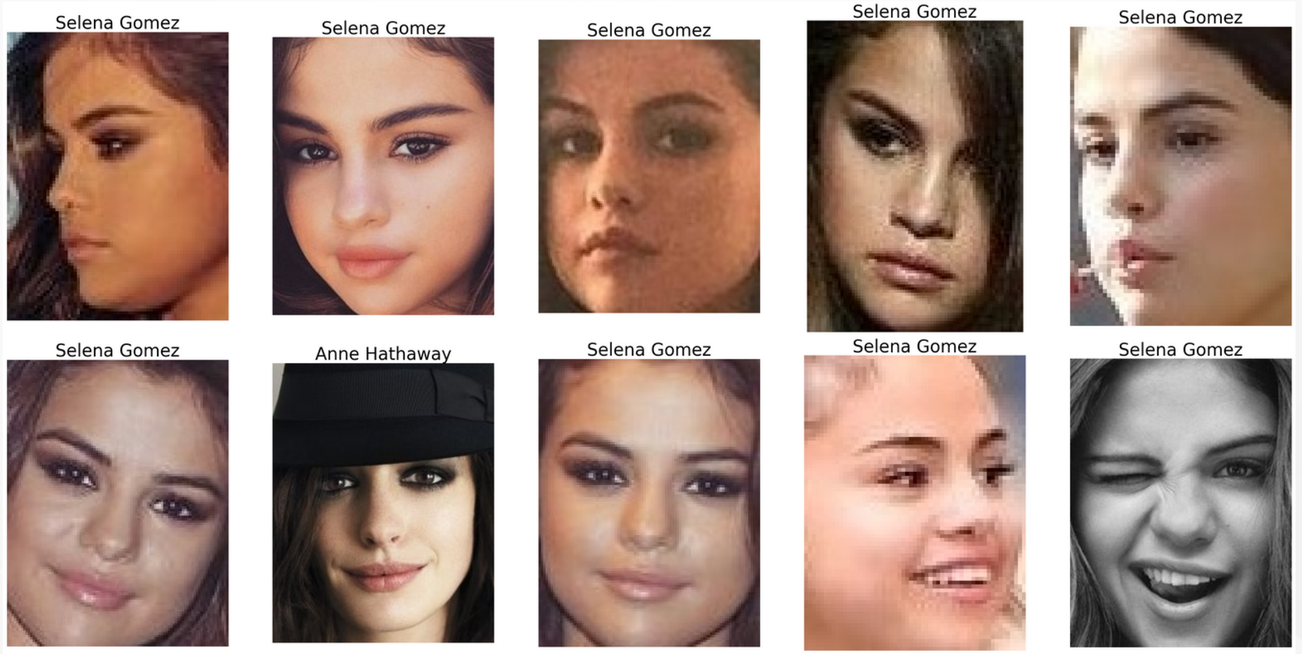
PRECISION



MobileNet



VGG-Face



Precision 1/10

Precision 9/10

RESULTS

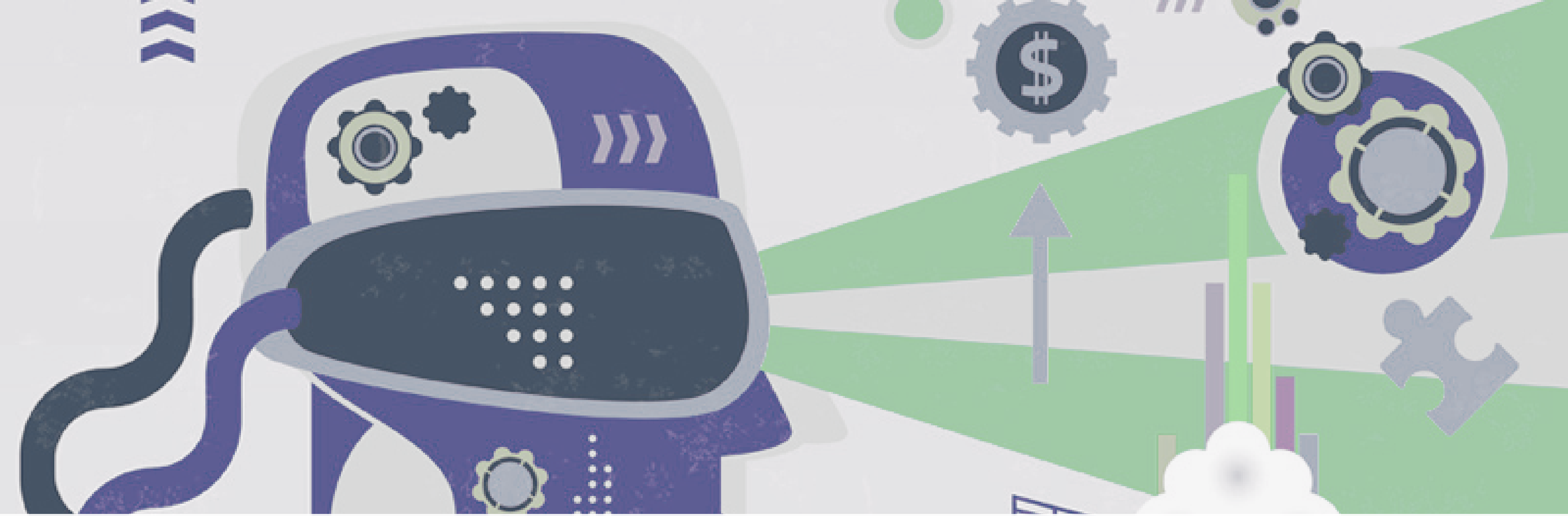
Vip	P@10(MobileNet)	P@10(VGGFace)
Hugh Jackman	0/10	8/10
Lionel Messi	1/10	9/10
Rami Malek	0/10	6/10
Selena Gomez	1/10	9/10
Tom Cruise	2/10	8/10
Tom Hardy	1/10	9/10
Tom Holland	0/10	8/10
Media	0.07	0,81

RESULTS


Logan Lerman
Logan Lerman
Zac Efron
Alvaro Morte
Henry Cavil
Logan Lerman
Mark Ruffalo
Alex Lawther
Chris Hemsworth
Leonardo DiCaprio

RESULTS


Christian Bale
Henry Cavil
Zac Efron
Pedro Alonso
Mark Ruffalo
Christian Bale
Henry Cavil
Alvaro Morte
Mark Ruffalo
Jake Mcdorman



THANK YOU

Università degli Studi Milano Bicocca