

# *Image Classification for Inspection Machines: a comprehensive review and application study*



*Università Cattolica del Sacro Cuore,  
Applied Data Science For Banking and Finance  
Elisa Salvi*

# *AGENDA*

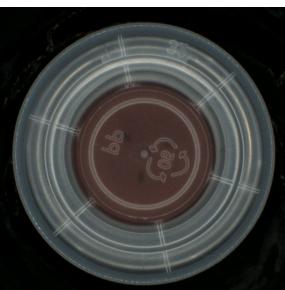
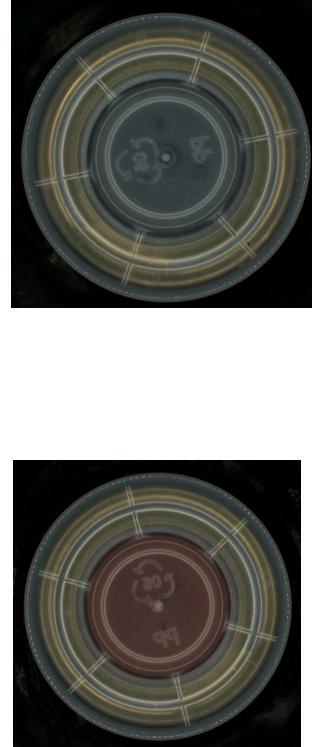


- ✓ Dataset with 9 types of Images
- ✓ Image Pre-processing:
  - Edge Detection
  - Circle Hough Transform
- ✓ Features Extraction and Selection:
  - RGB levels
  - SIFT
  - HOG
- ✓ Classification Algorithms for Computer Vision:
  - Decision Tree, Random Forest, SVM, KNN
  - Convolutional Neural Network
  - Vision Transformer
- ✓ Results
- ✓ GUI



## MOTIVATION

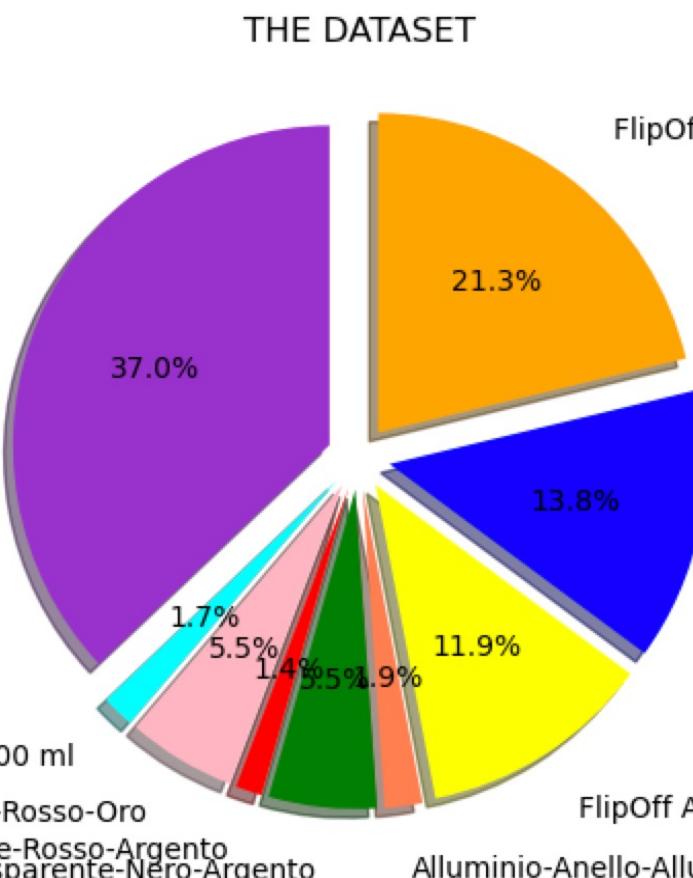
My internship project's main scope is to *classify* *Flip-off's images* acquired by the Inspection Machine department of *Antares Vision S.p.A.*



Trasparente-Nero  
Oro-250 ml

Trasparente-Nero-Oro-1000 ml  
Trasparente-Rosso-Oro  
Trasparente-Rosso-Argero  
Trasparente-Nero-Argero

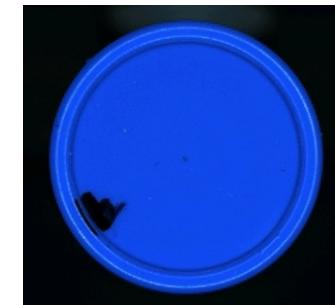
Count



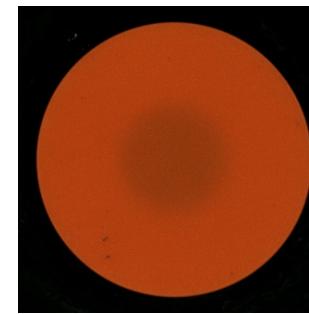
FlipOff Arancio con scritta



Flipoff Blu



FlipOff Arancio

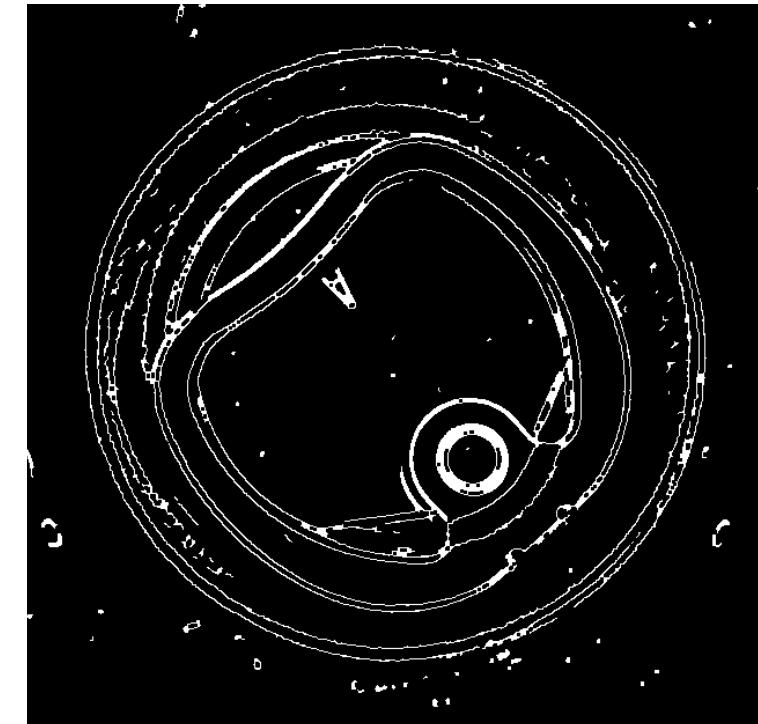


# *EDGE DETECTION*



Edge detection is a technique of image processing used to identify points in a digital image with discontinuities, such as sharp changes in the image brightness.

The points where the image brightness varies sharply are called the *edges* (or *boundaries*) of the image.



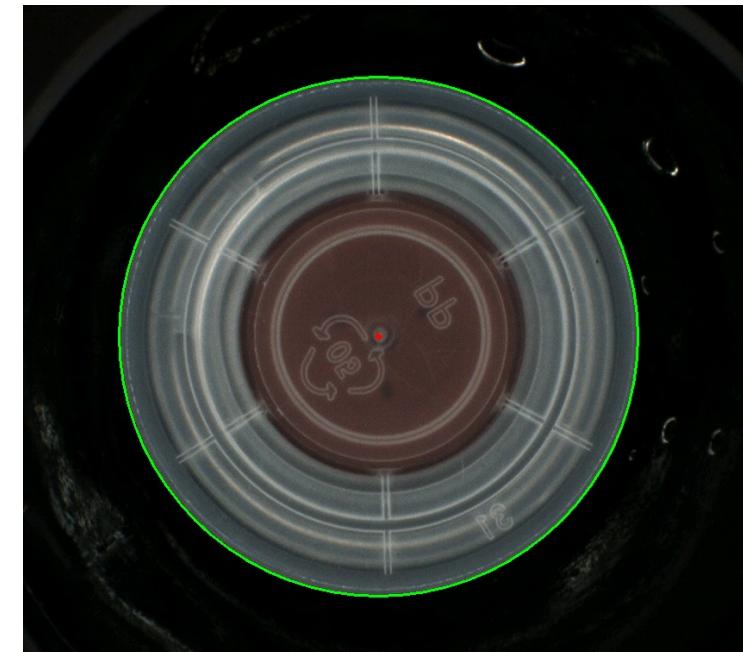
# **CIRCLE HOUGH TRANSFORM**



The ***Hough Transform*** is an extraction technique to find the shape class of objects in an image even if in boundaries are present some imperfections or gaps.

The ***Circle Hough Transform*** is a feature extraction technique used for detecting circles in imperfect images.

The circle candidates are produced by “voting” in the Hough parameter space and then selecting local maxima in an accumulator matrix.



# FEATURE EXTRACTION



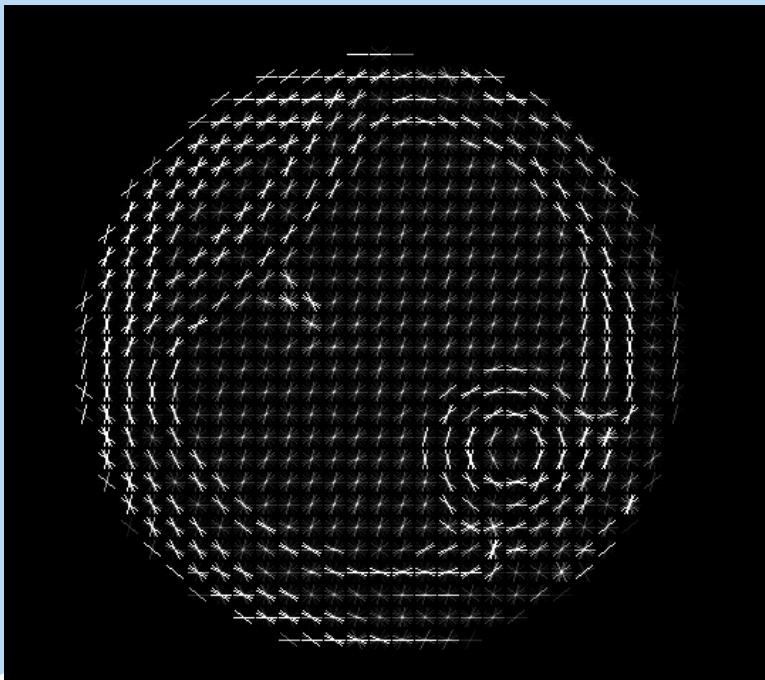
- ***RGB*** mean, variance and standard deviation
- ***HOG*** min, max, mean and standard deviation
- ***SIFT*** mean and standard deviation

The features consider the internal circle area (80% of the initial radius) in order to expunge from the images useless shades and the background.

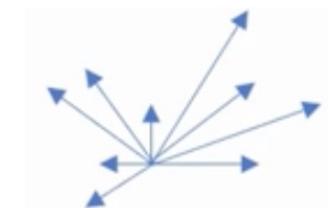
NOT ALL the features extracted are used for the models to avoid overfitting. Two feature selection algorithms facilitate this choice ranking the features importance: a XGBtree and a regression.

# HISTOGRAM OF ORIENTED GRADIENTS

---



- HOG calculates the *gradient* of every 8x8 pixels cell in the image.
- The magnitude is higher when there is a sharp change in pixels' intensity, such as around the edges.
- Using the gradients and the Pythagoras theorem, the *magnitude* and *direction* for each pixel value are calculated.
- With this two information, an *histogram* is generated for each cell. The 9 bins represent the direction ( $0^\circ$ - $180^\circ$ ) and the height of the bins is the magnitude of the gradient.
- Each histogram is then transformed in *a vector representation*.



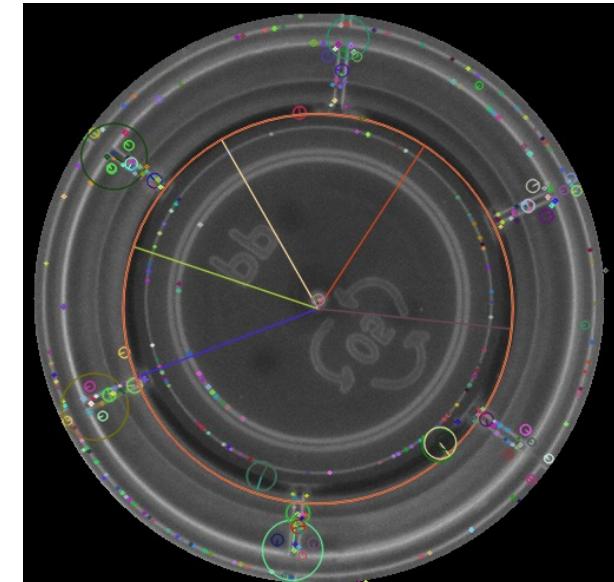
# **SCALE INVARIANT FEATURE TRANSFORM**

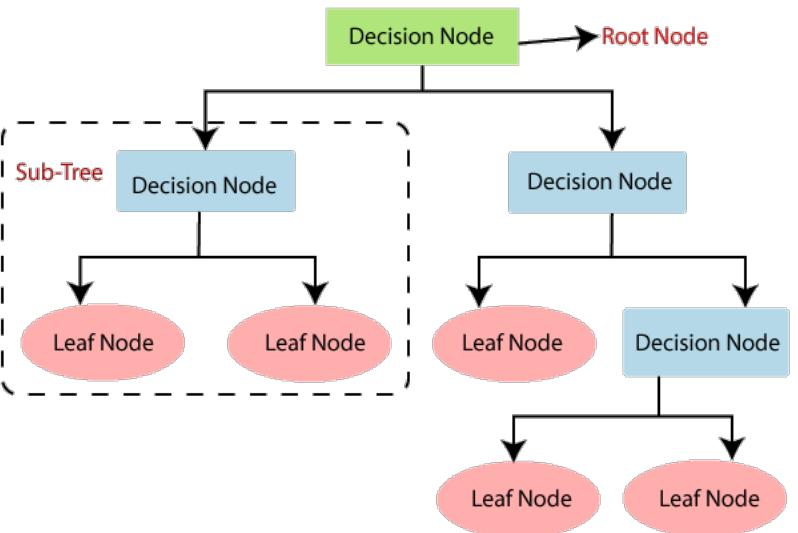
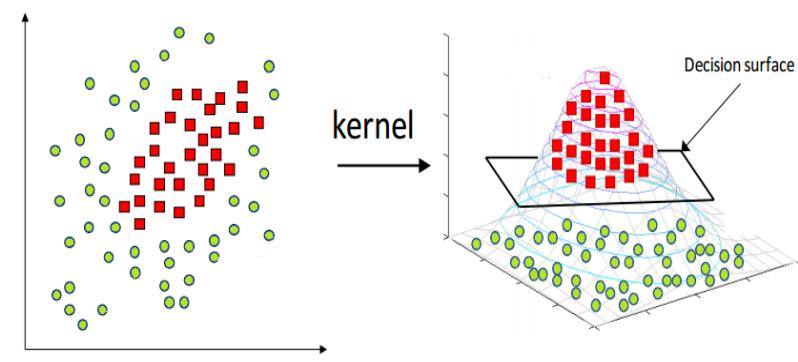


The **SIFT** is an algorithm used to detect and describe local features in digital images, invented by D. Lowe in 1999.

It locates certain **key points** and then describes them with quantitative information (so-called **descriptors**) which can be used for object recognition.

The descriptors are supposed to be **invariant** against various transformations (rotation, scaling, lightening, ...) which might make images look different, although they represent the same object.

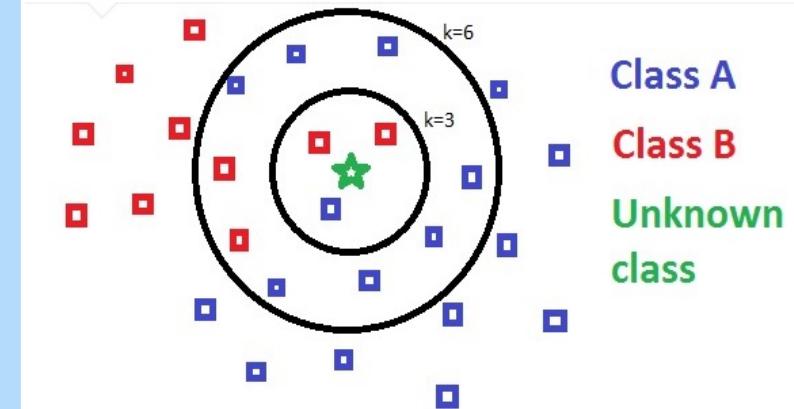
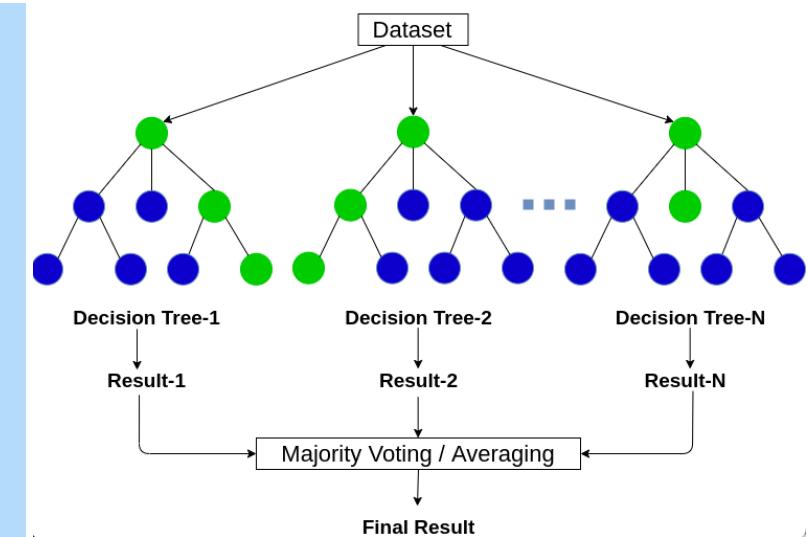




# CLASSIFICATION ALGORITHMS

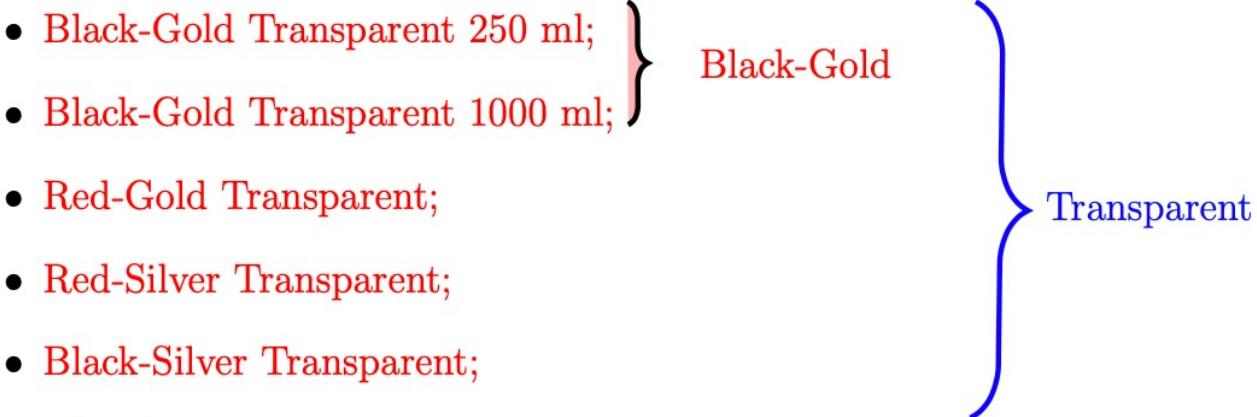


- *Decision Tree*
- *Random Forest*
- *Support Vector Machines*
- *K Nearest Neighbor*



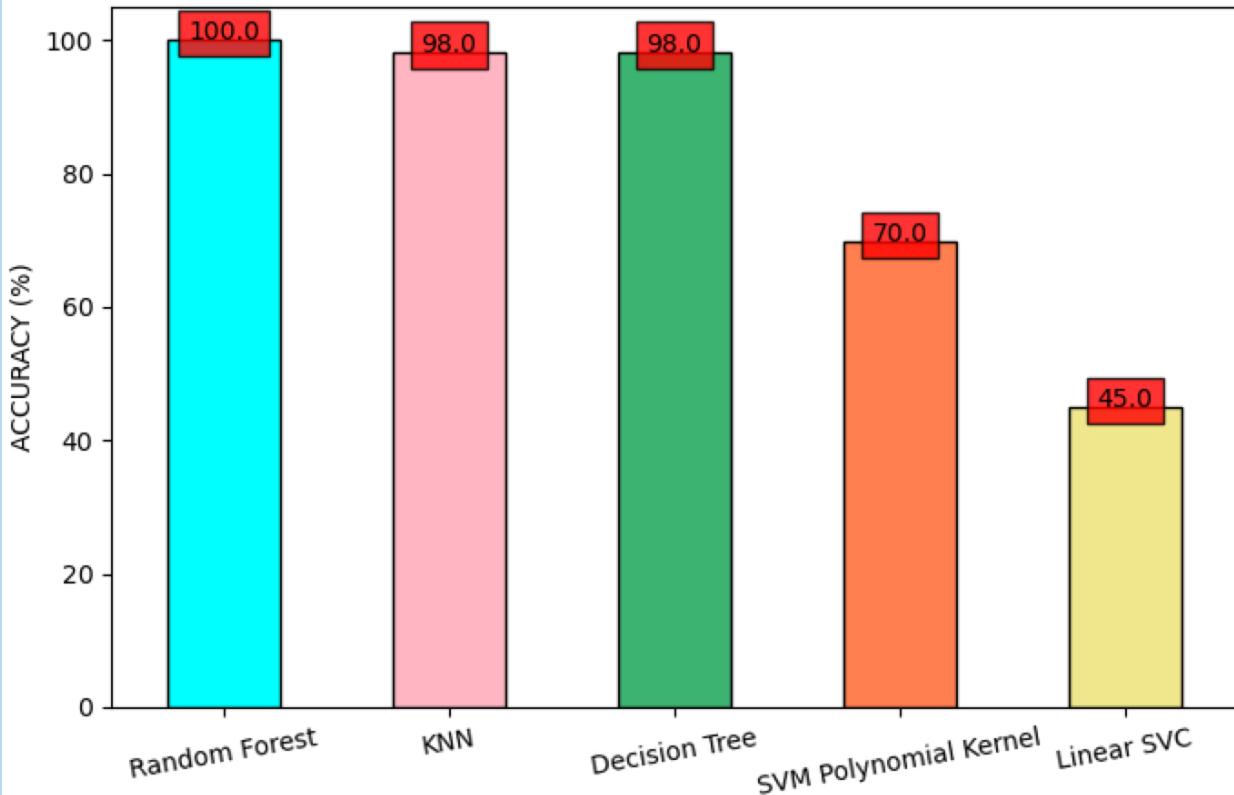
## *PROBLEM DIVIDED IN TWO SUB-PROBLEMS*



- Black-Gold Transparent 250 ml;
  - Black-Gold Transparent 1000 ml;
  - Red-Gold Transparent;
  - Red-Silver Transparent;
  - Black-Silver Transparent;
  - Aluminium;
  - Orange;
  - Blue;
  - Orange with text.
- 
- The list items are grouped into two main categories: 'Black-Gold' (in red) and 'Transparent' (in blue). The 'Black-Gold' group contains the first two items. The 'Transparent' group contains the remaining seven items. A red brace on the right side of the first two items spans them both. A blue brace on the right side of the remaining seven items spans all of them.

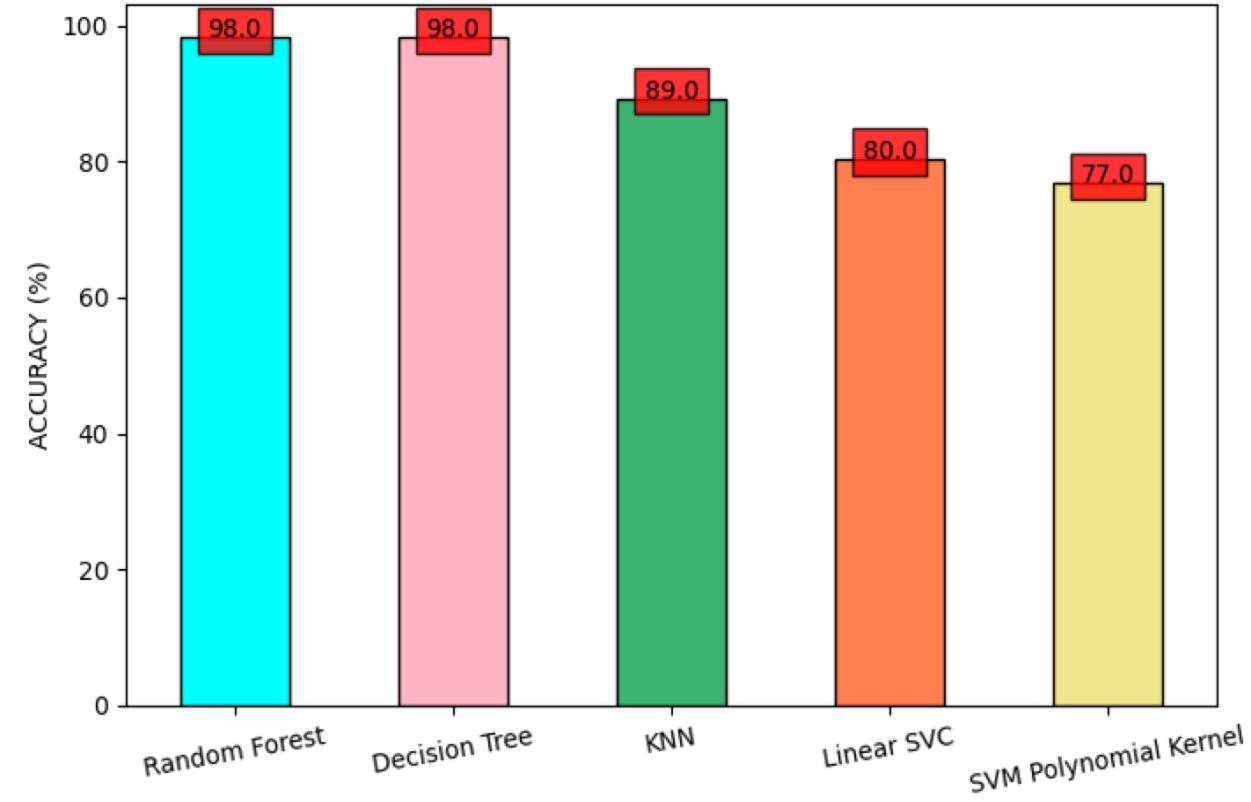
# *CLASSIFICATION GRUPPING TRANSPARENT FLIPOFF*

## **THE MODELS ACCURACY**

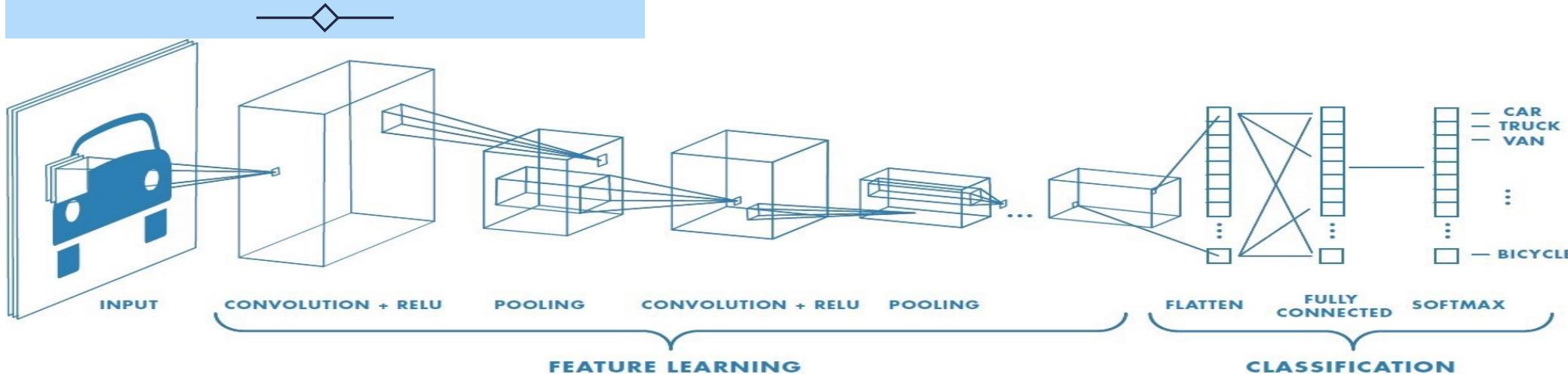


# *CLASSIFICATION TRANSPARENT FLIPOFF*

## **THE MODELS ACCURACY**



# CONVOLUTIONAL NEURAL NETWORK

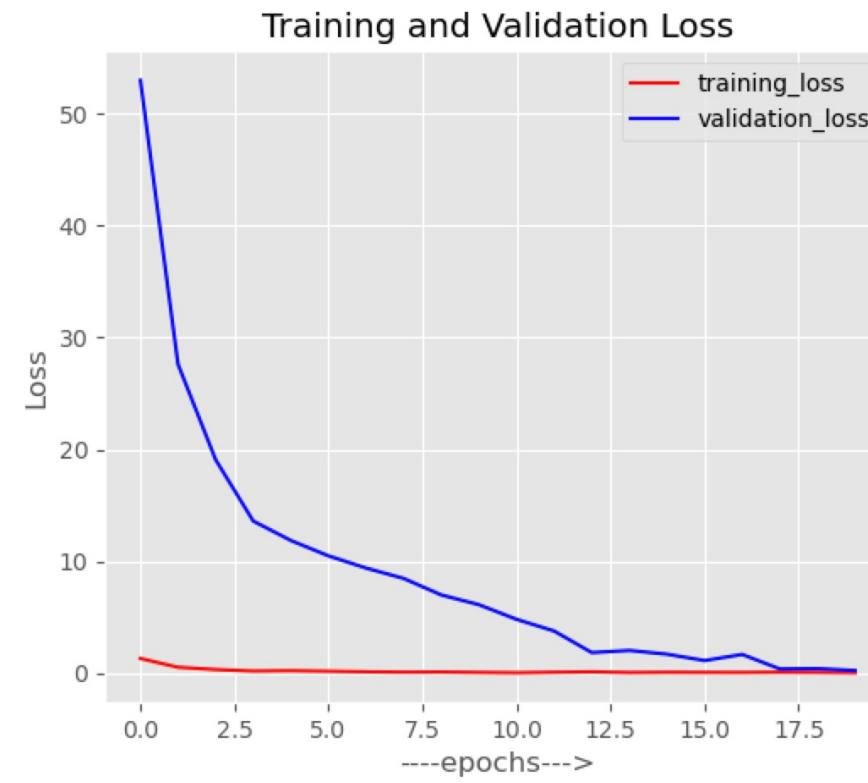
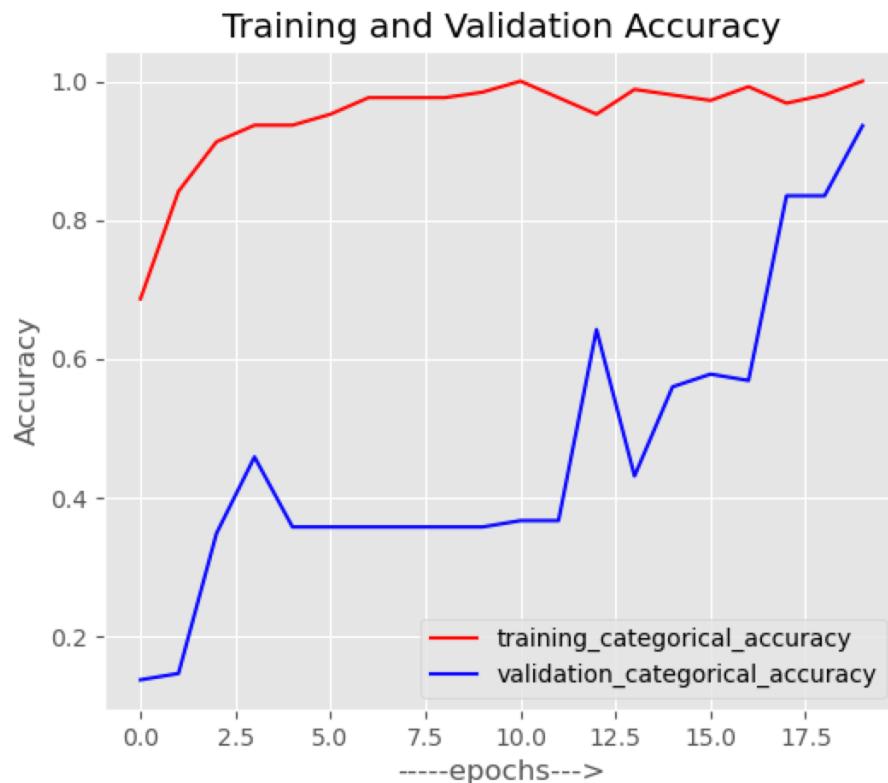


Multiple, sequential layers that are directly fed with raw data.

- **Convolutional** layers: apply Filters to extract features and discard irrelevant information with the ReLu.
- **Max Pooling** layer: reduces the dimensionality of the output.
- Layers are repeated to extract higher-level features.
- **Dense** layer: aggregates information before classification.
- The “**SoftMax**” activation function is used to convert the output into the membership probability.

# CNN CLASSIFICATION:

accuracy set from the GUI at 90%



Accuracy:	Number of epochs:	Time:
30%	4	35 sec.
50%	14	2 min.
60%	17	2 min. and 25 sec.
70%	13	2 min.
80%	16	2 min. and 20 sec.
90%	20	3 min.

**macOS Big Sur**

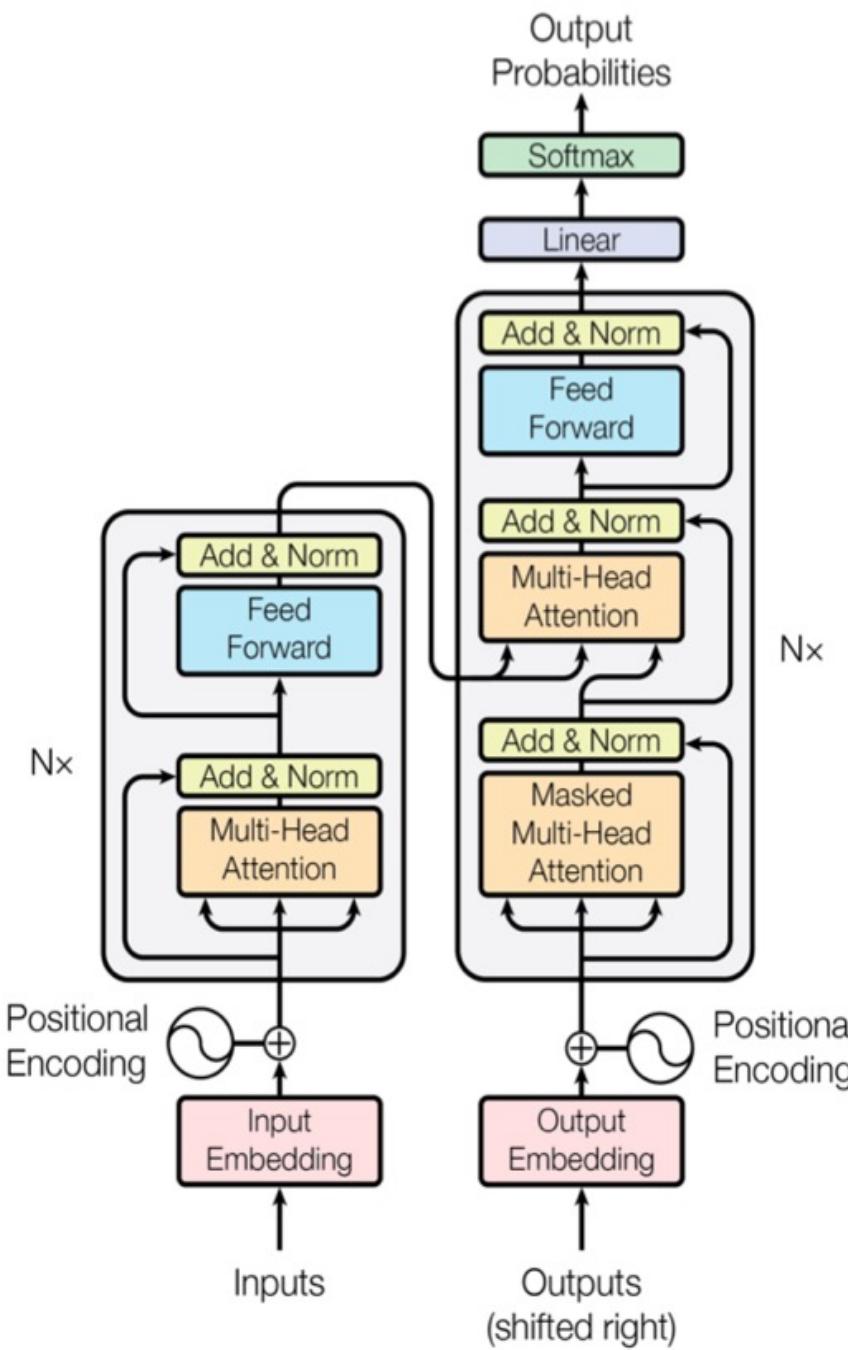
Versione 11.6.8

MacBook Pro (16-inch, 2019)

Processore 2,4 GHz Intel Core i9 8 core

Memoria 64 GB 2667 MHz DDR4

Scheda grafica Intel UHD Graphics 630 1536 MB



# TRANSFORMER

CNNs assume that neighbour pixels are more important than distant pixels.

“*Self-Attention*” mechanism captures long-range relations between pixels.

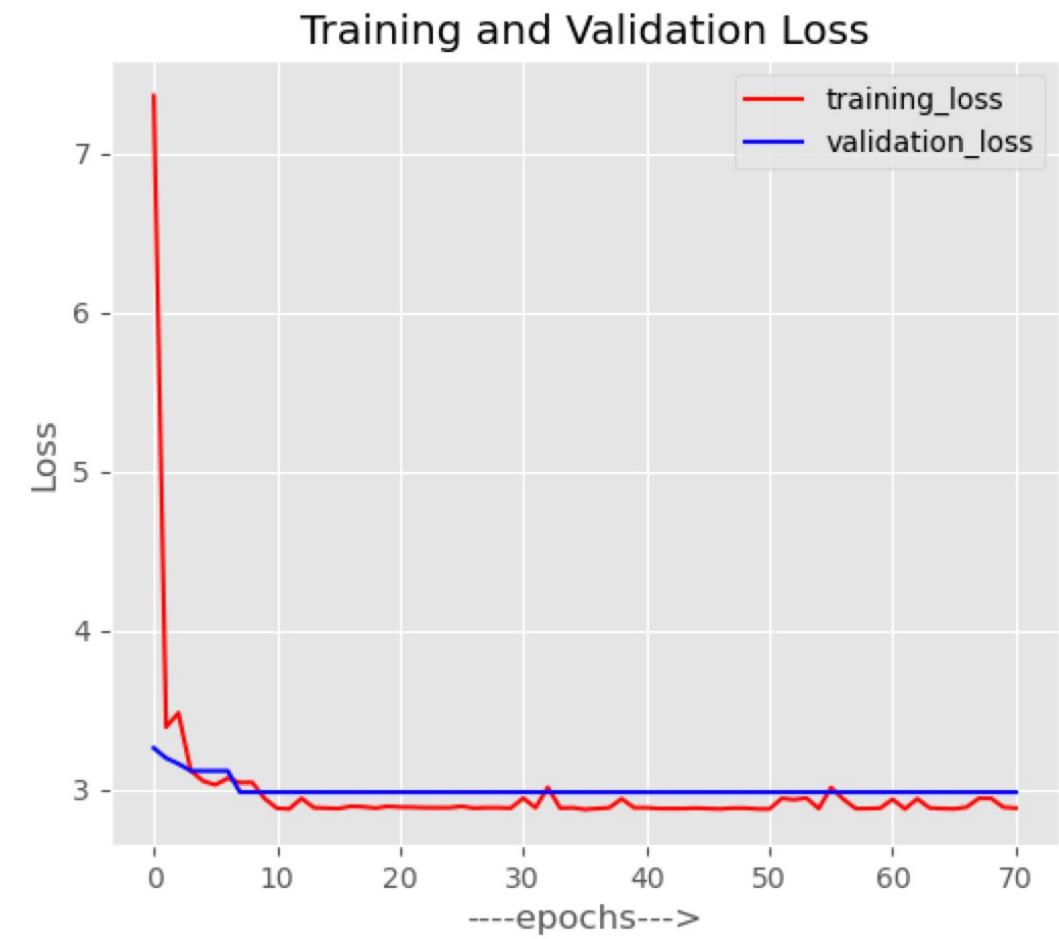
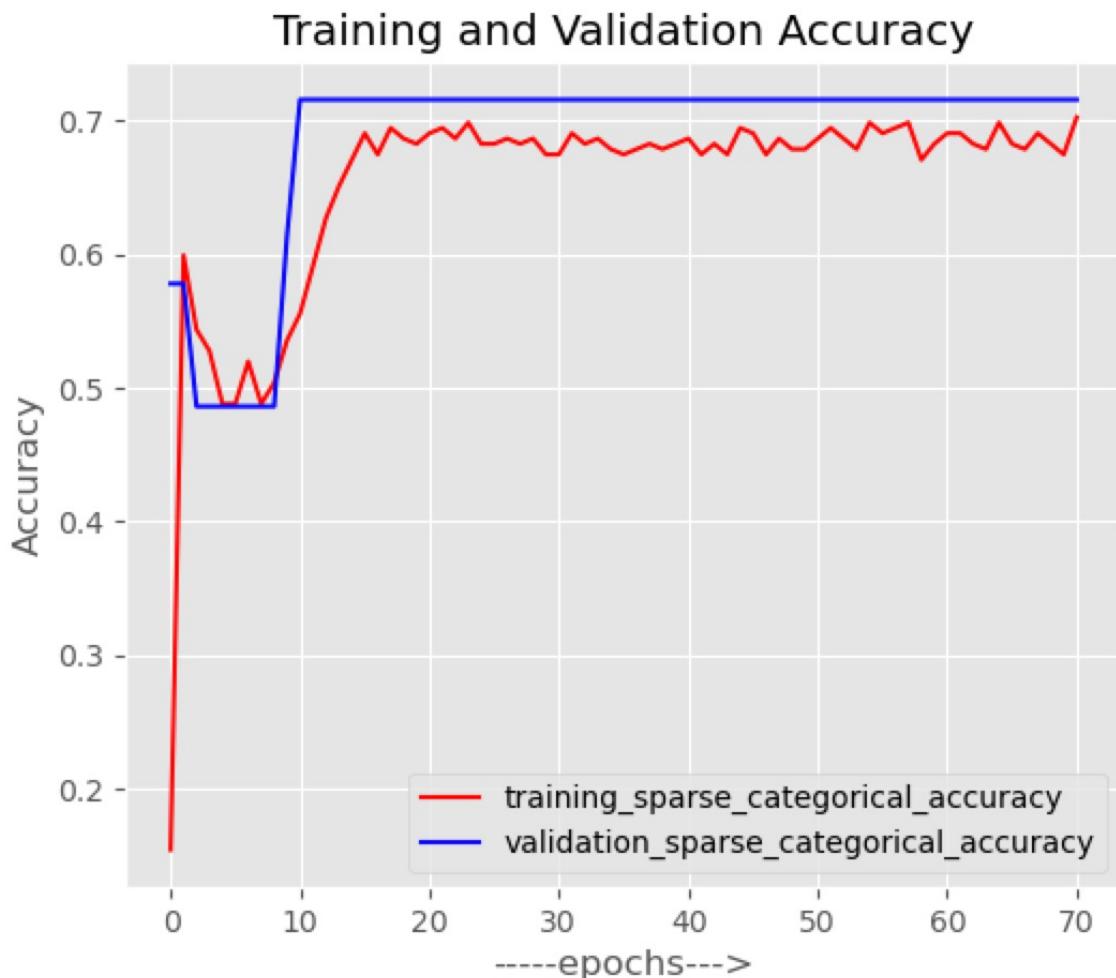
In **2021 Bazi** used a *ViTransformer* for image classification for the first time.

**INPUT EMBEDDING:** input is broken down in *tokens* that are represented by vectors, called “*embeddings*”. These vectors lie on a multi-dimensional space, called “*embedding space*”, that aims to capture relationship between the pixels based on their distance.

$$\text{Attention}(Q, K, V) = W \cdot V = \text{Softmax} \left( \frac{Q \cdot K^T}{\sqrt{d_{keys}}} \right) \cdot V$$

# VI-TRANSFORMER CLASSIFICATION

accuracy set from the GUI at 70%



Accuracy:	Number of epochs:	Time:
30%	2	18 sec.
50%	2	18 sec.
60%	9	1 min.
70%	70	8 min.

# ***GRAFICAL USER INTERFACE WITH “TKINTER”***



IMAGE ANALYSIS

Models Prediction

Models Prediction for Transparent Flipoff

Neural Network

Vision Transformer

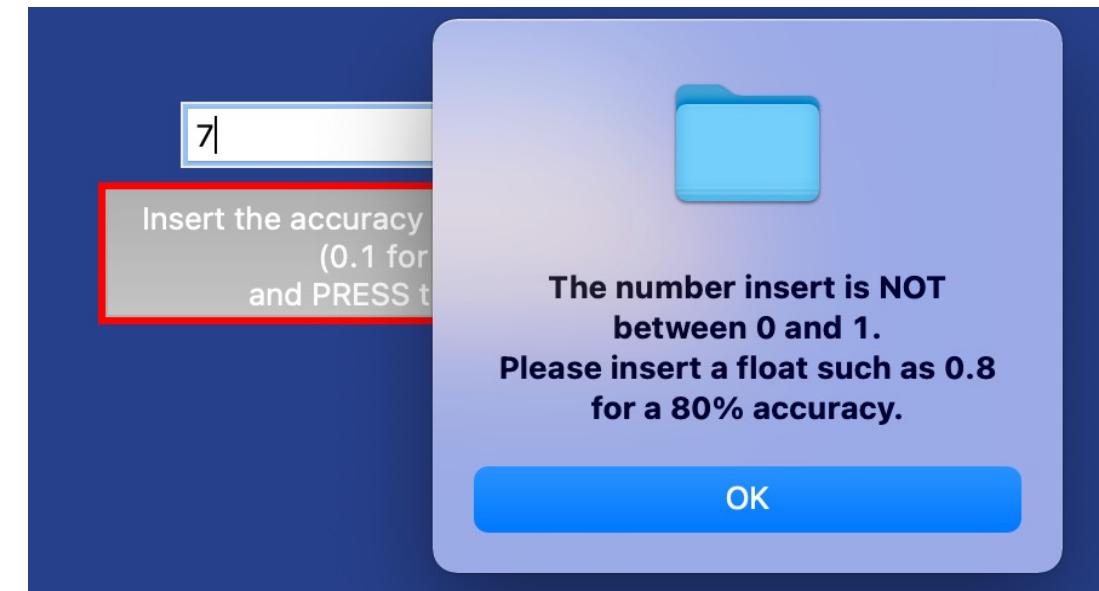
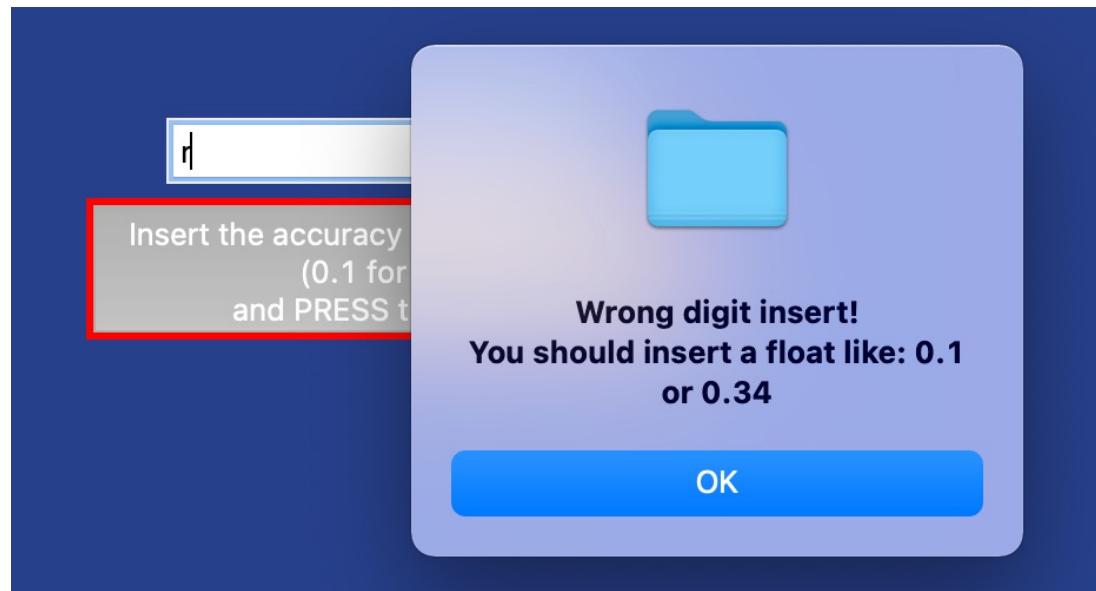
EXIT

# **TOOL ROBUSTNESS**

## *Exceptions/Warnings*

except ValueError:

```
tk.messagebox.showerror(title="ERROR", message="Wrong digit insert! \n You should insert a float like: 0.1 or 0.34")
```

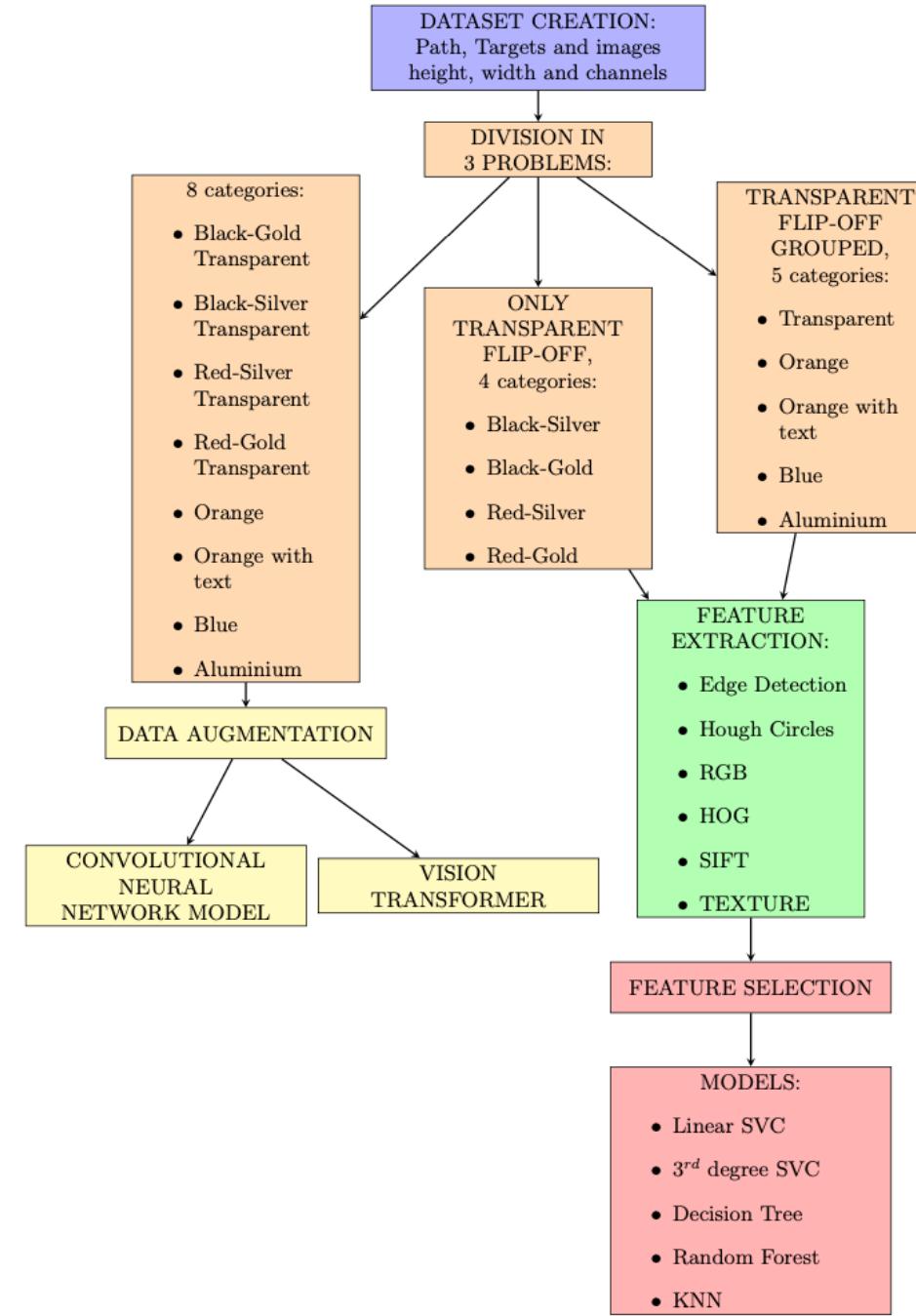


***THANK YOU!***



*Elisa Salvi*

**THESIS  
PROJECT  
FLOWCHART**



# IMAGE ANALYSIS

THE DATASET

EDGE DETECTION

HOUGH CIRCLE

80% MASK

HOG?

HISTOGRAM OF ORIENTED GRADIENTS

SIFT?

SCALE INVARIANT FEATURE TRANSFORM

FEATURE EXTRACTION

FEATURE IMPORTANCE lin\_reg

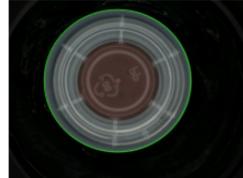
FEATURE IMPORTANCE XGBClassifier

BACK

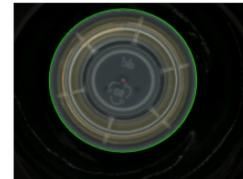
Trasparente-Nero  
Oro-250 ml



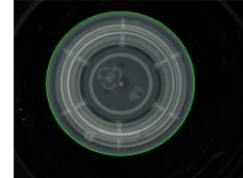
Trasparente-Rosso-Arancio



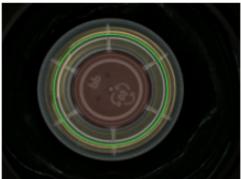
Trasparente-Nero-Oro-1000 ml



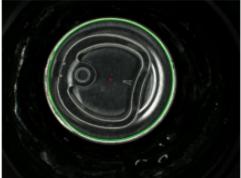
Trasparente-Nero-Arancio



Trasparente-Rosso-Oro



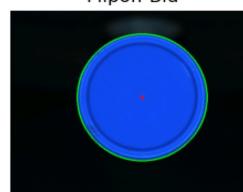
Alluminio-Anello-Alluminio



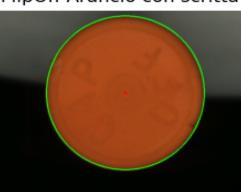
FlipOff Arancio



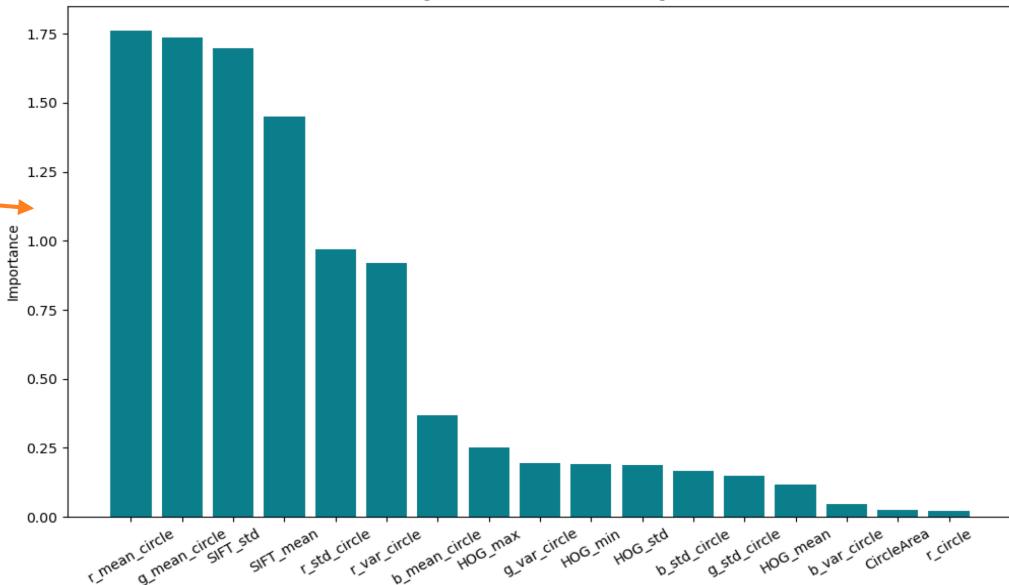
Flipoff Blu



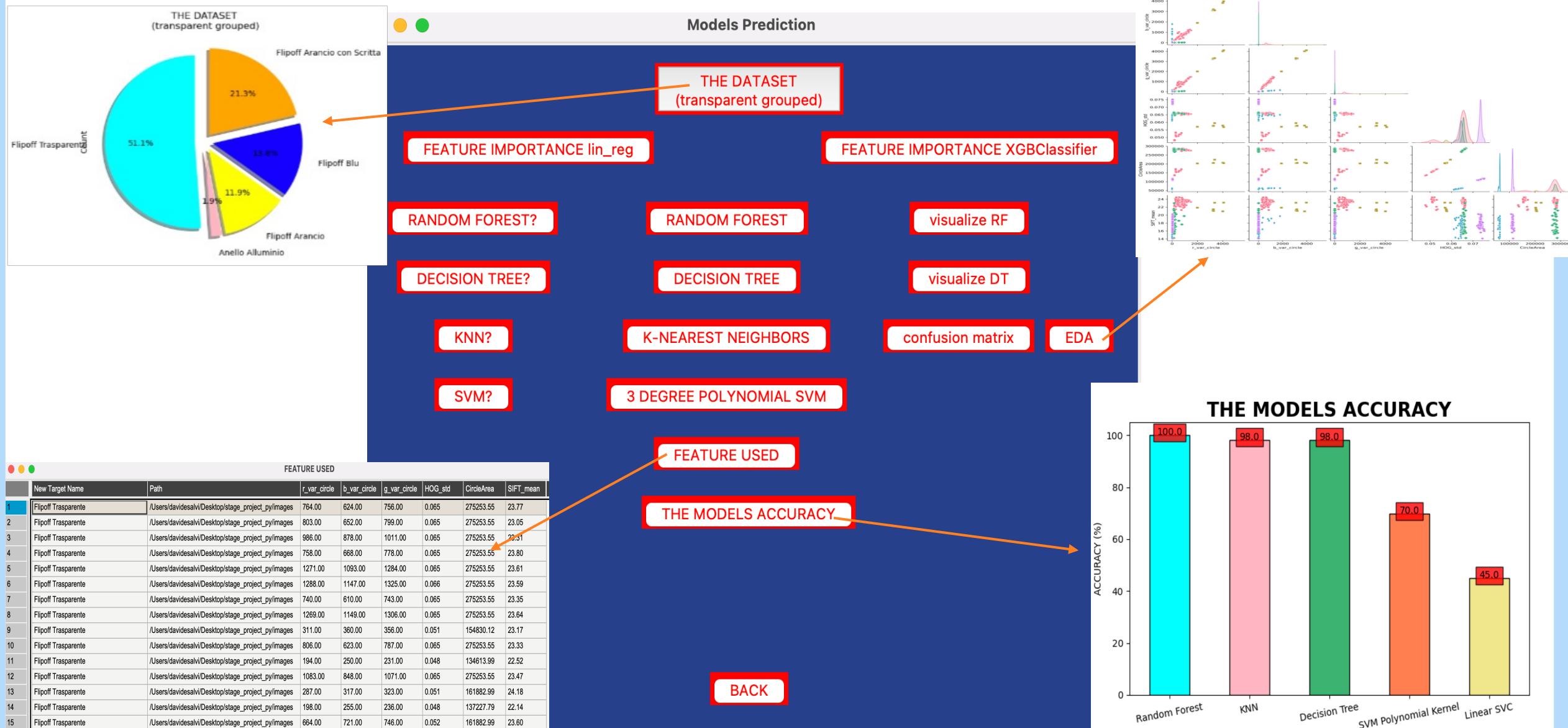
FlipOff Arancio con scritta



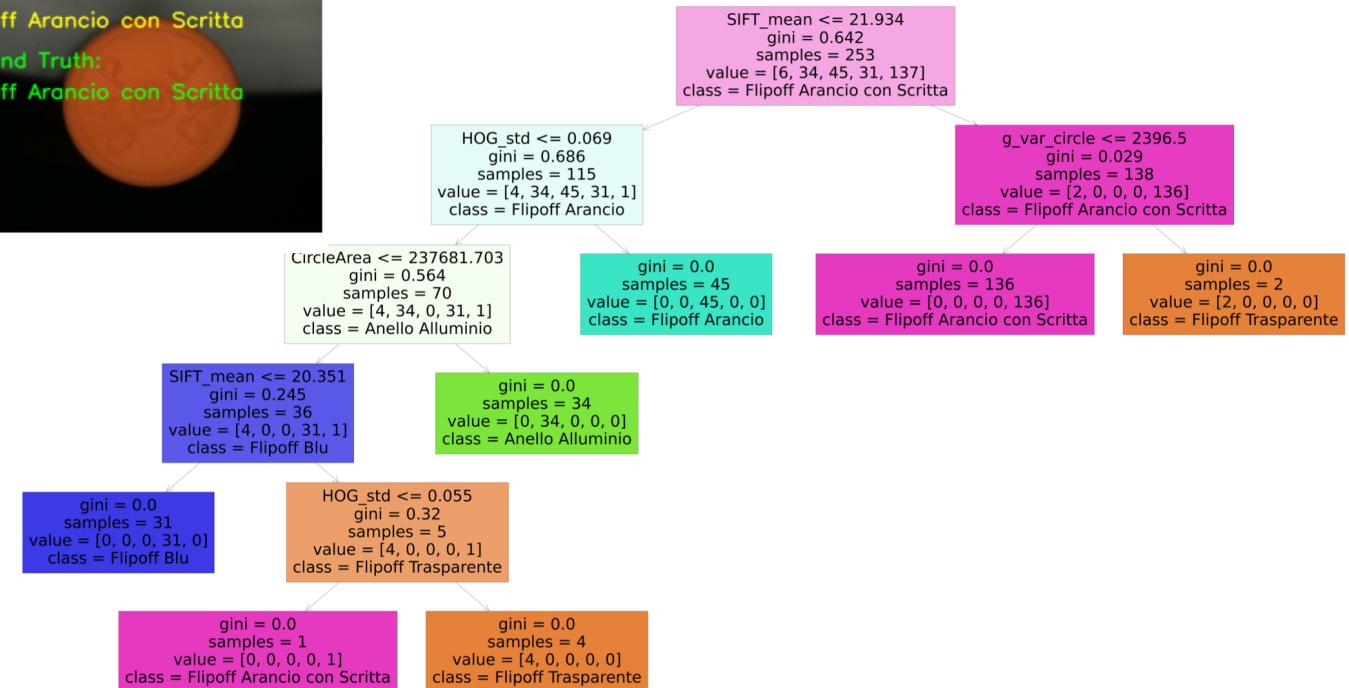
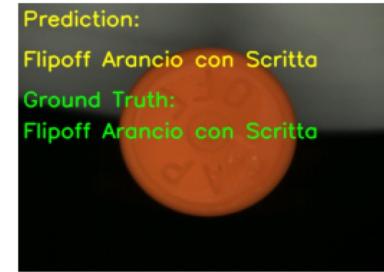
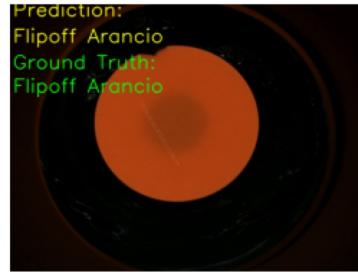
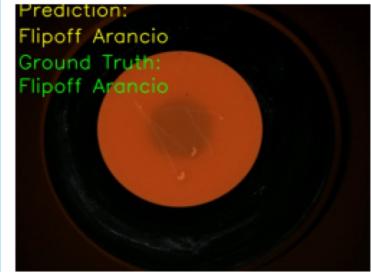
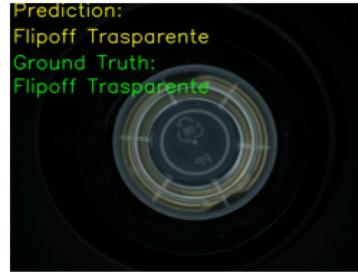
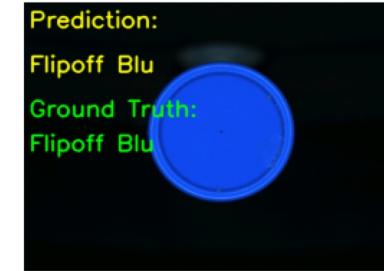
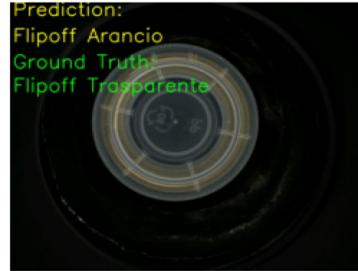
Feature importance obtained from linear regression coefficients (abs)  
(no list feature)



# THE SUB-PROBLEMS VISUALISATION



## DECISION TREE



# CNN and ViTransformer GUI

