

Extraction de caractéristiques

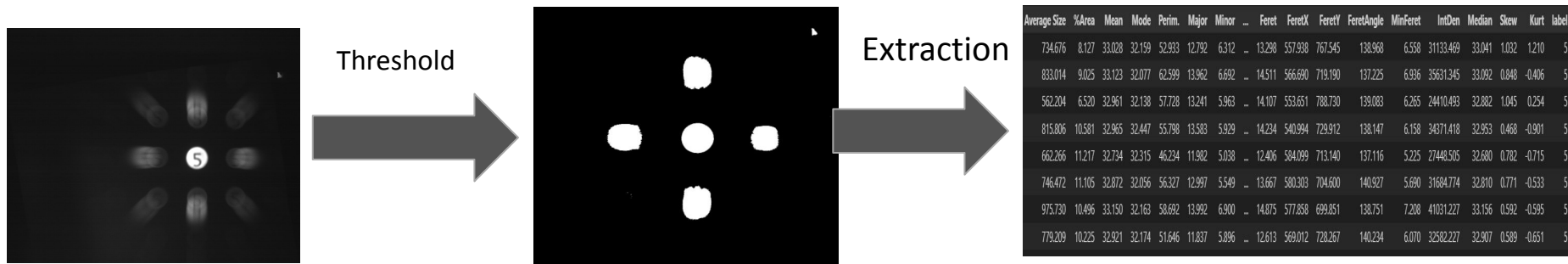
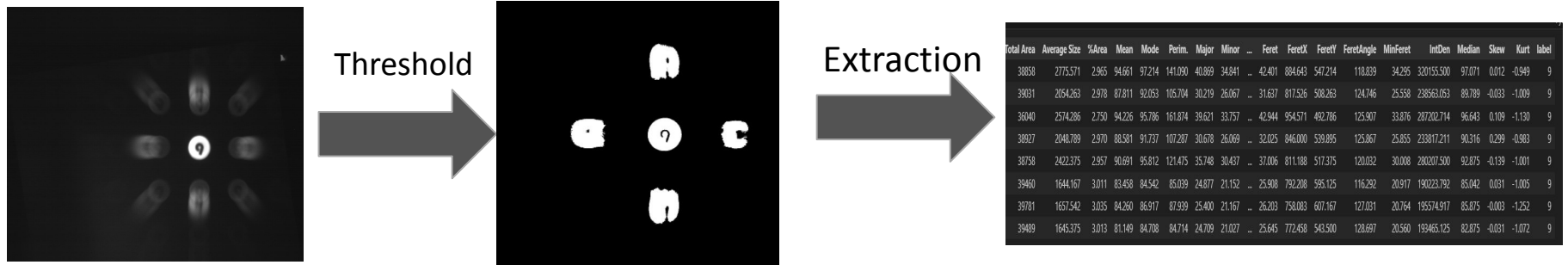
Chaanani Youness

# Présentation du problème de classification

- Dans ce problème, on dispose d'une base de données comprenant deux catégories distincts, chacune décrite par 23 caractéristiques(features).
- Notre objectif est de développer un modèle décisionnel capable de classer ces deux catégories avec précision.
- On a observé que certaines caractéristiques présentes dans la base de données pourraient perturber l'apprentissage de notre modèle.
- Par conséquent, on a envisagé, avant de construire le modèle, de sélectionner les caractéristiques les plus importantes afin d'optimiser les performances de classification.

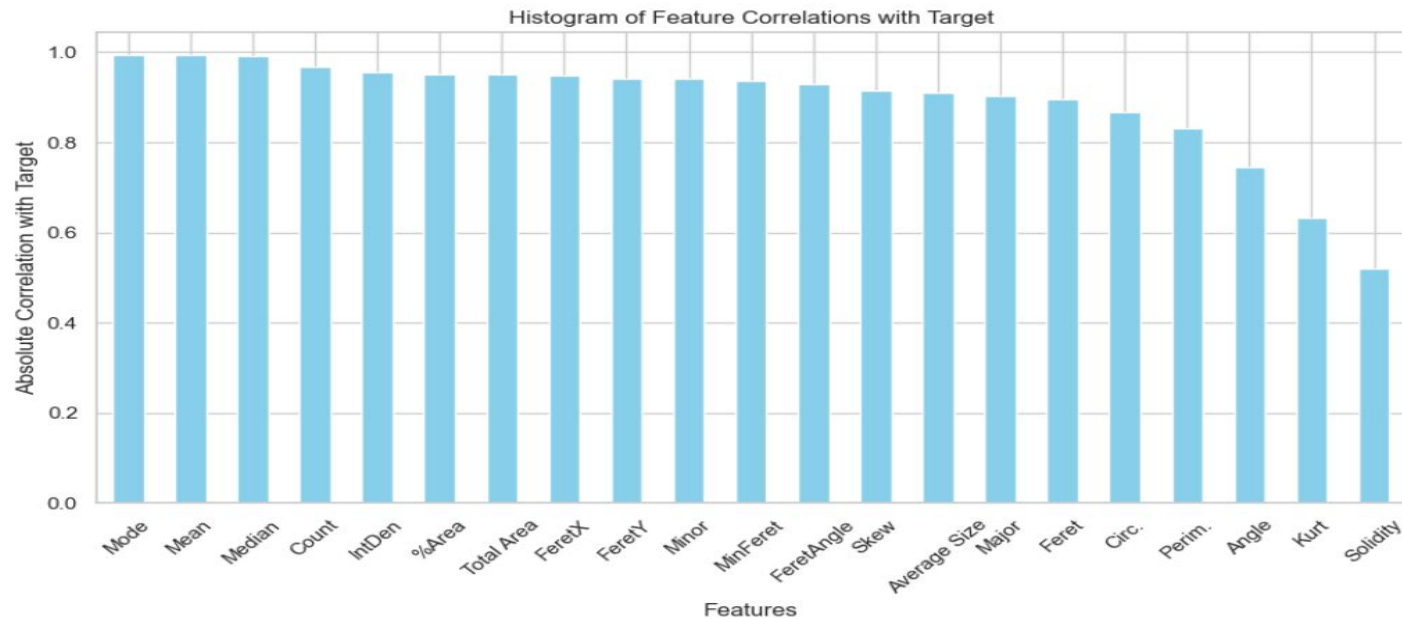


# Le pipeline d'extraction de features



# La sélection des meilleurs features

- Pour sélectionner les features, je me suis basé sur l'étude de la corrélation entre chaque feature et la variable cible, en conservant uniquement les caractéristiques ayant un seuil de corrélation élevé.
- On remarque que les features Mode, Médiane et Moyenne sont les plus corrélées avec la variable cible, ce qui fait on va conserver ces features.



# Les différentes approches pour la sélection de features

- On peut passer par le test "pearson", aussi connu sous le nom de coefficient de corrélation de Pearson, est une mesure statistique de la corrélation linéaire entre deux variables.

