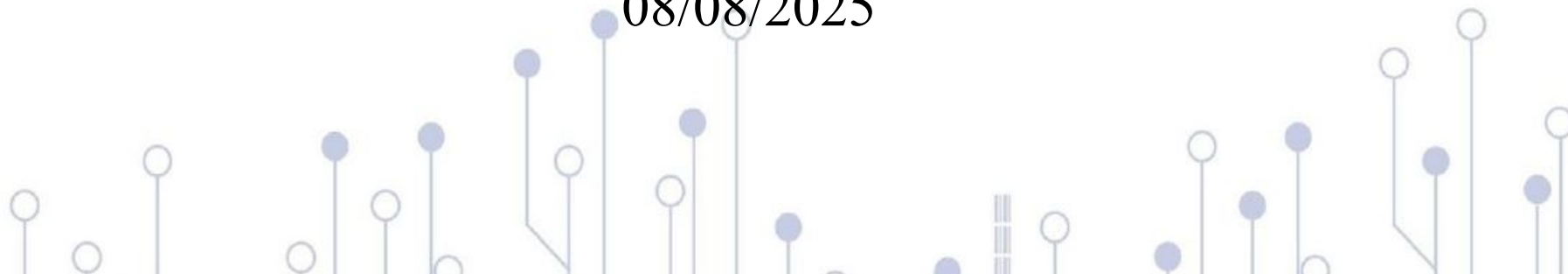


Breast Cancer Diagnosis

Mahesh Kumar Mulimani, Zhaowei Ding, Zebin Li

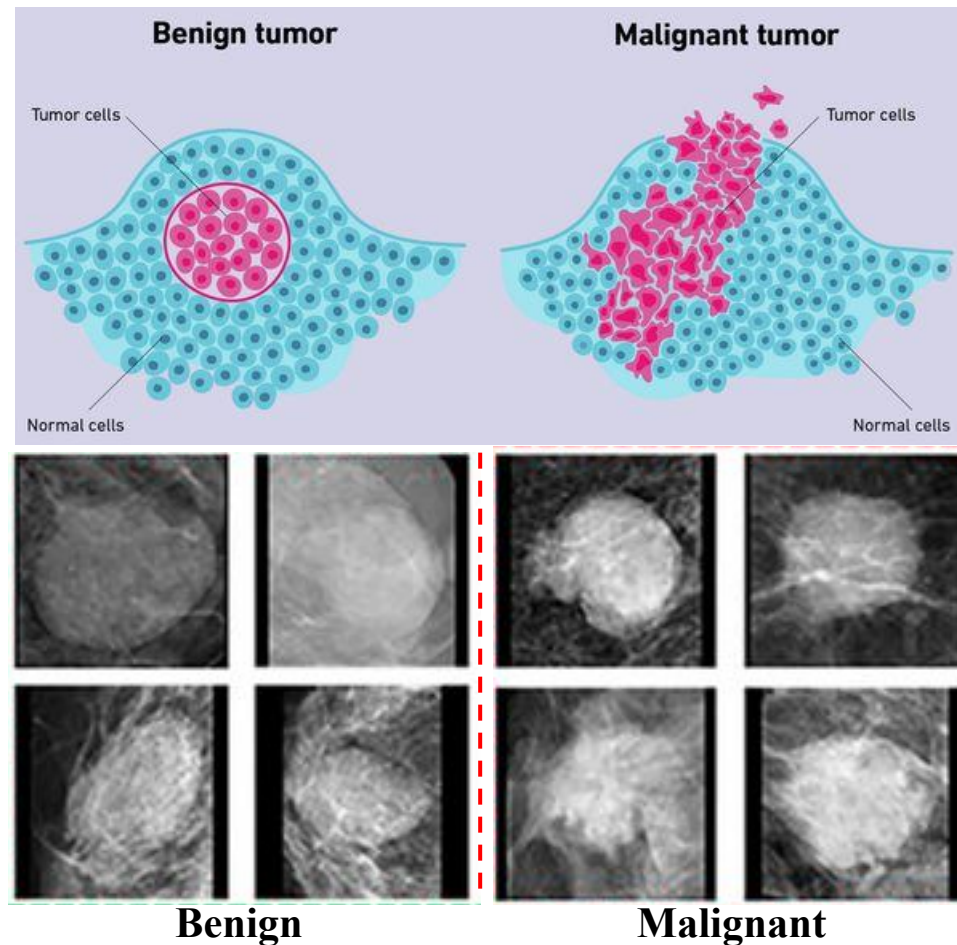
08/08/2025



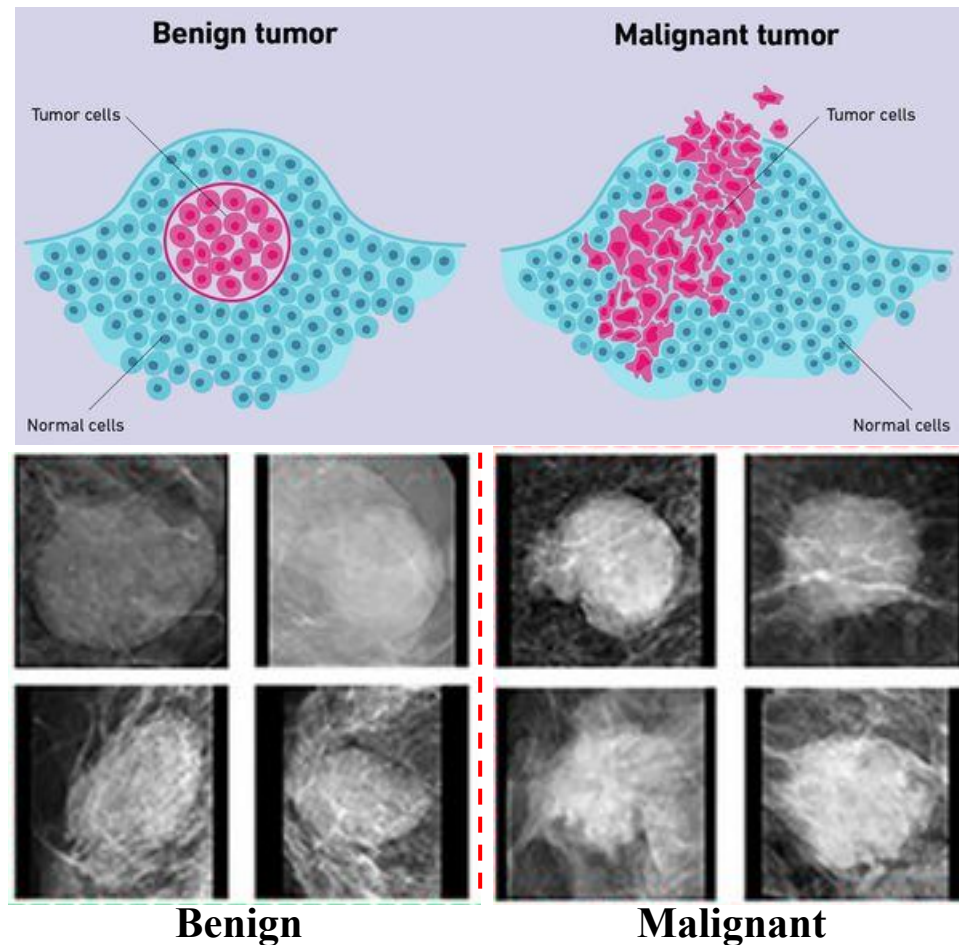
- Introduction
- Understanding the data
- Model
- Interpreting results
- Summary

Introduction

- Cancer is an unsolved disease and is one of the leading causes of death.
- Cancer cells can stay dormant for longer time and can reactivate later.



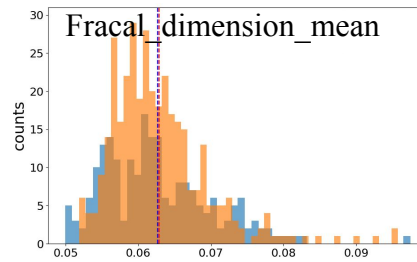
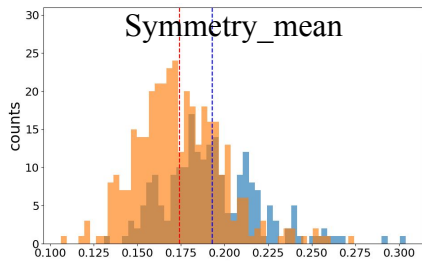
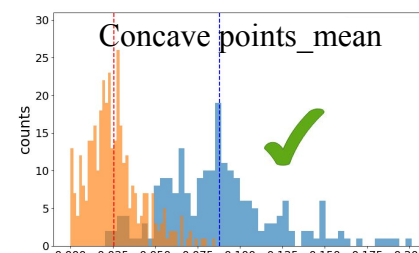
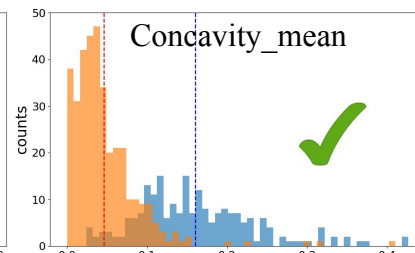
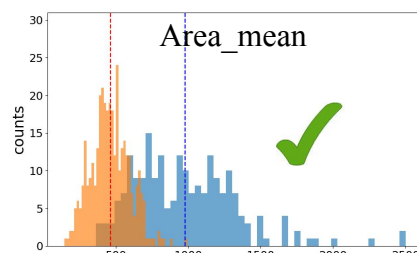
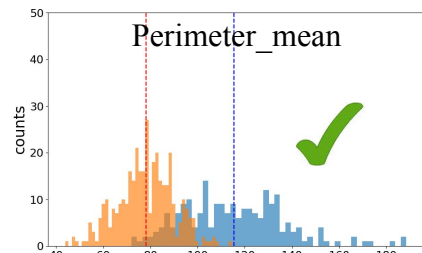
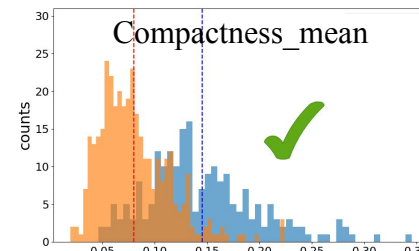
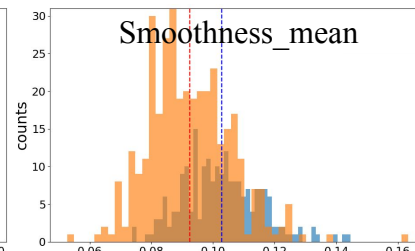
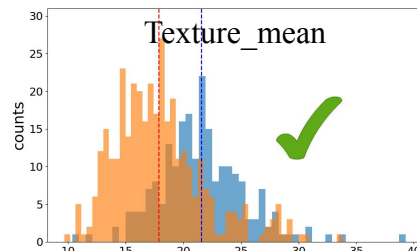
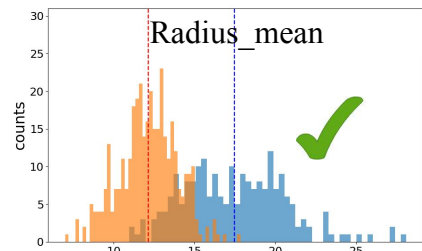
- Early screening of cancer can help in intervention that can mitigate the progression of the cancer.
- Goal: AI-assisted diagnostic tool detecting cancer cell type!



“All about the DATA!”

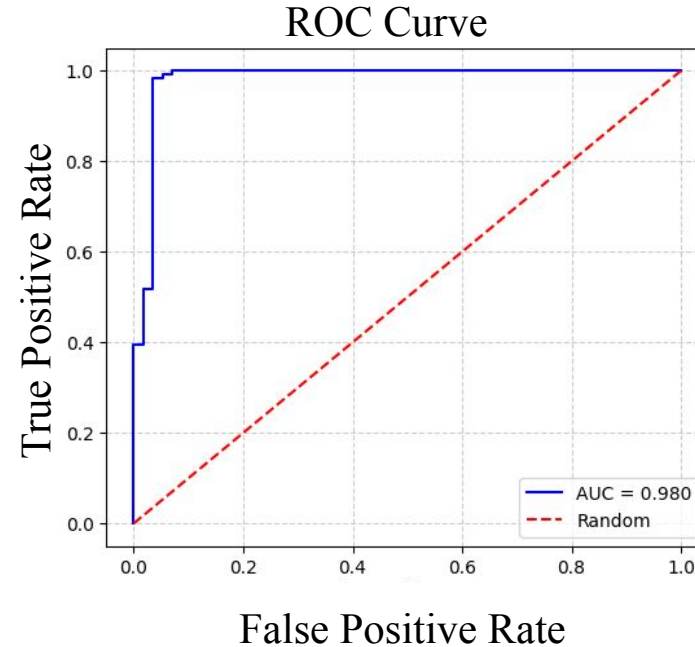
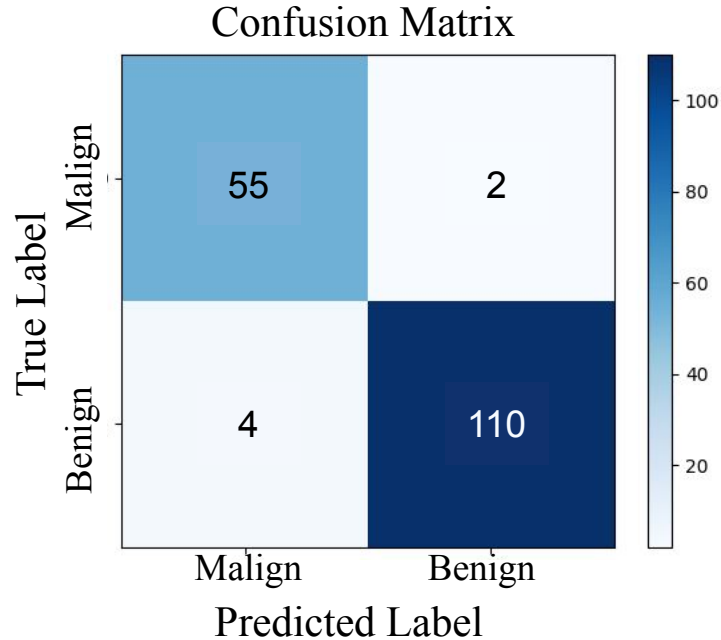
- Tabular data of 10 measured physical quantities (with standard deviations and worst values)
- Number of dataset : $N = 569$
- Physical quantities: Geometrical quantities of the cells in cancer mass
 - Radius, Texture, Perimeter, Area, Smoothness
 - Compactness, Concavity, Number of concave points
 - Symmetry and Fractal dimension

Benign Malign



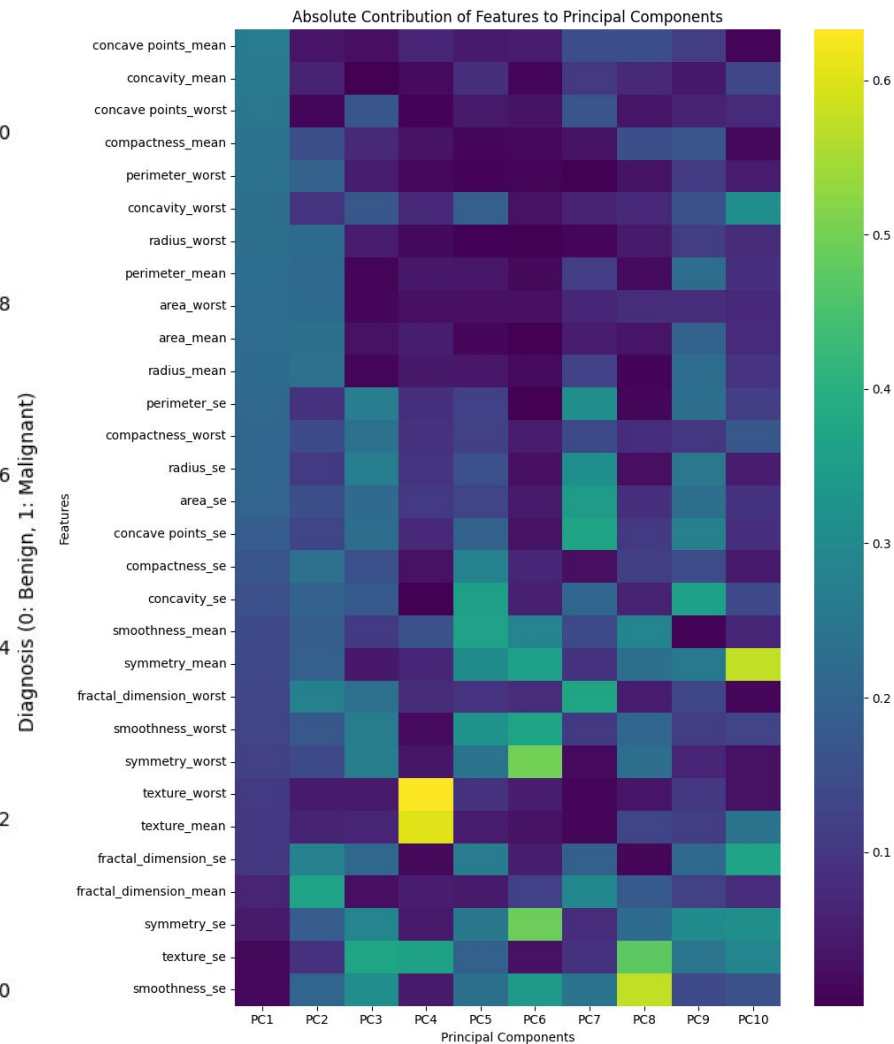
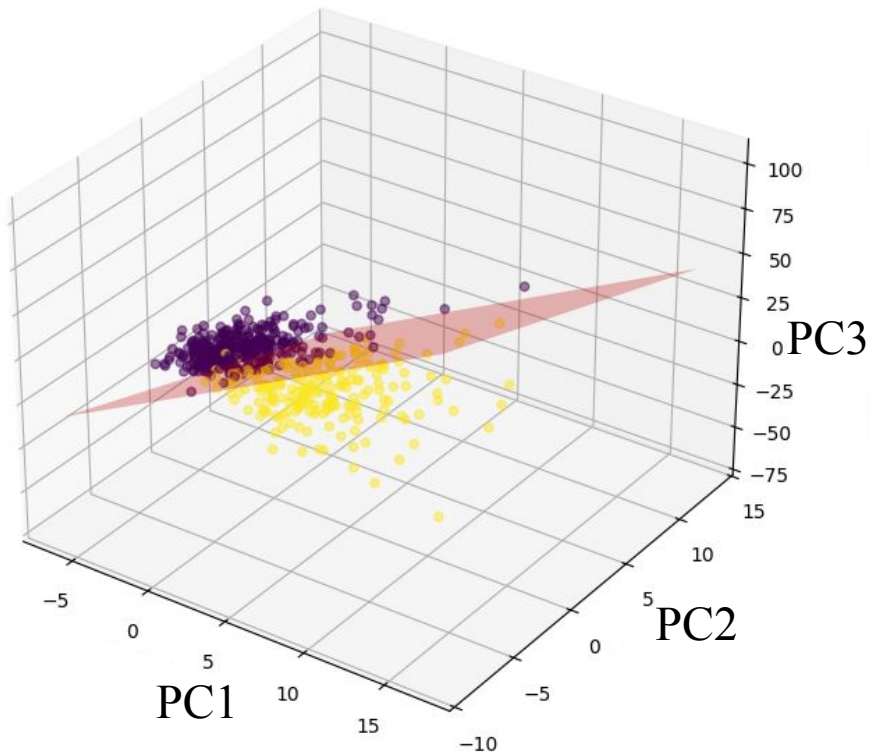
- Benign and Malign data shows difference in the data distributions in some geometric features!!

- Model: **Logistic regression** (preprocessed by normalization, training:testing=70:30)
 - We got 98.25% accuracy!



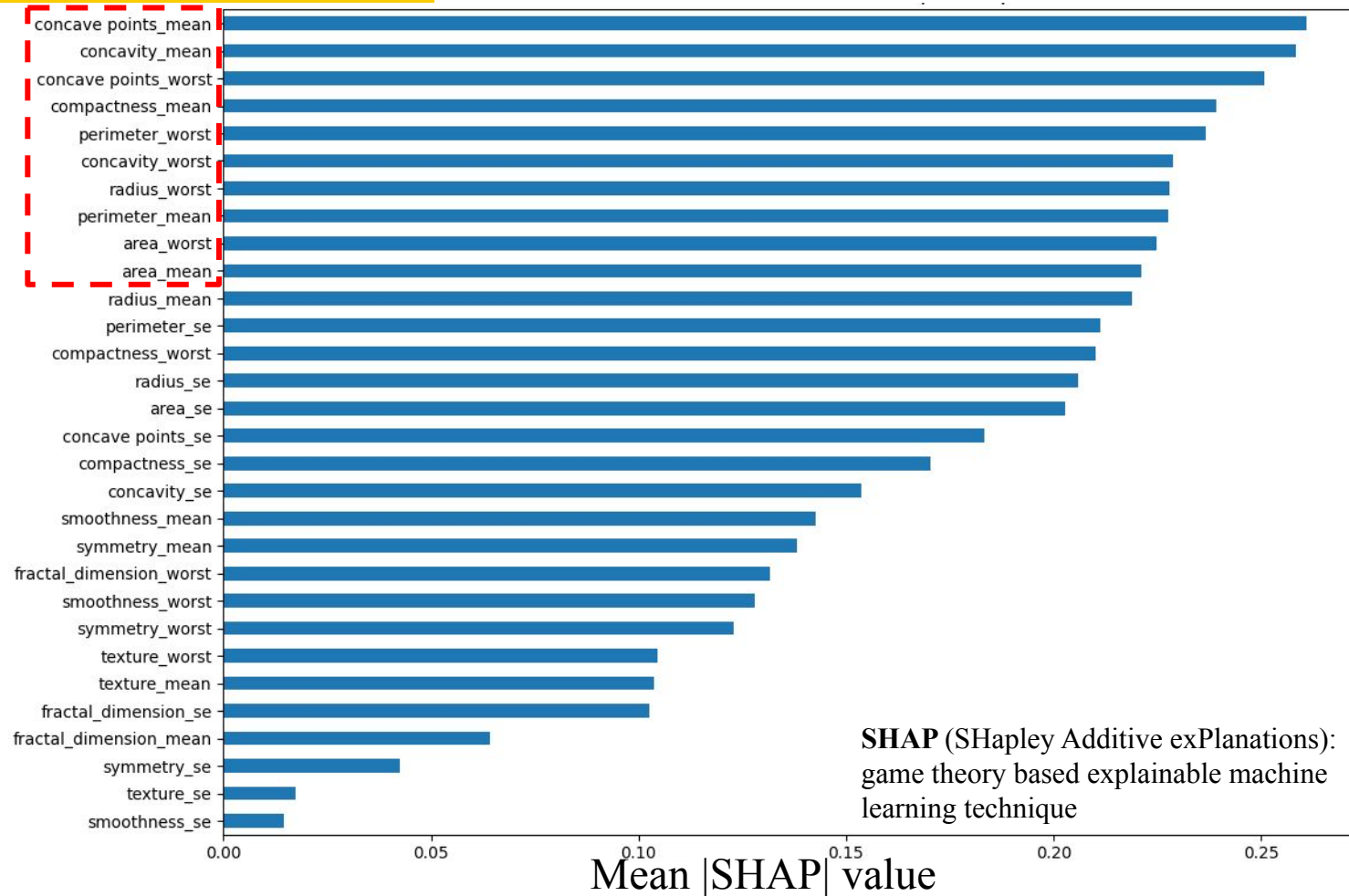
Interpreting results

Dataset Visualization via PCA



Interpreting results

SHAP Feature Importance (Logistic Regression)



Top 10

Concave_points_m

Concavity_m

Concave_points_w

Compactness_m

Perimeter_w

Concavity_w

Radius_w

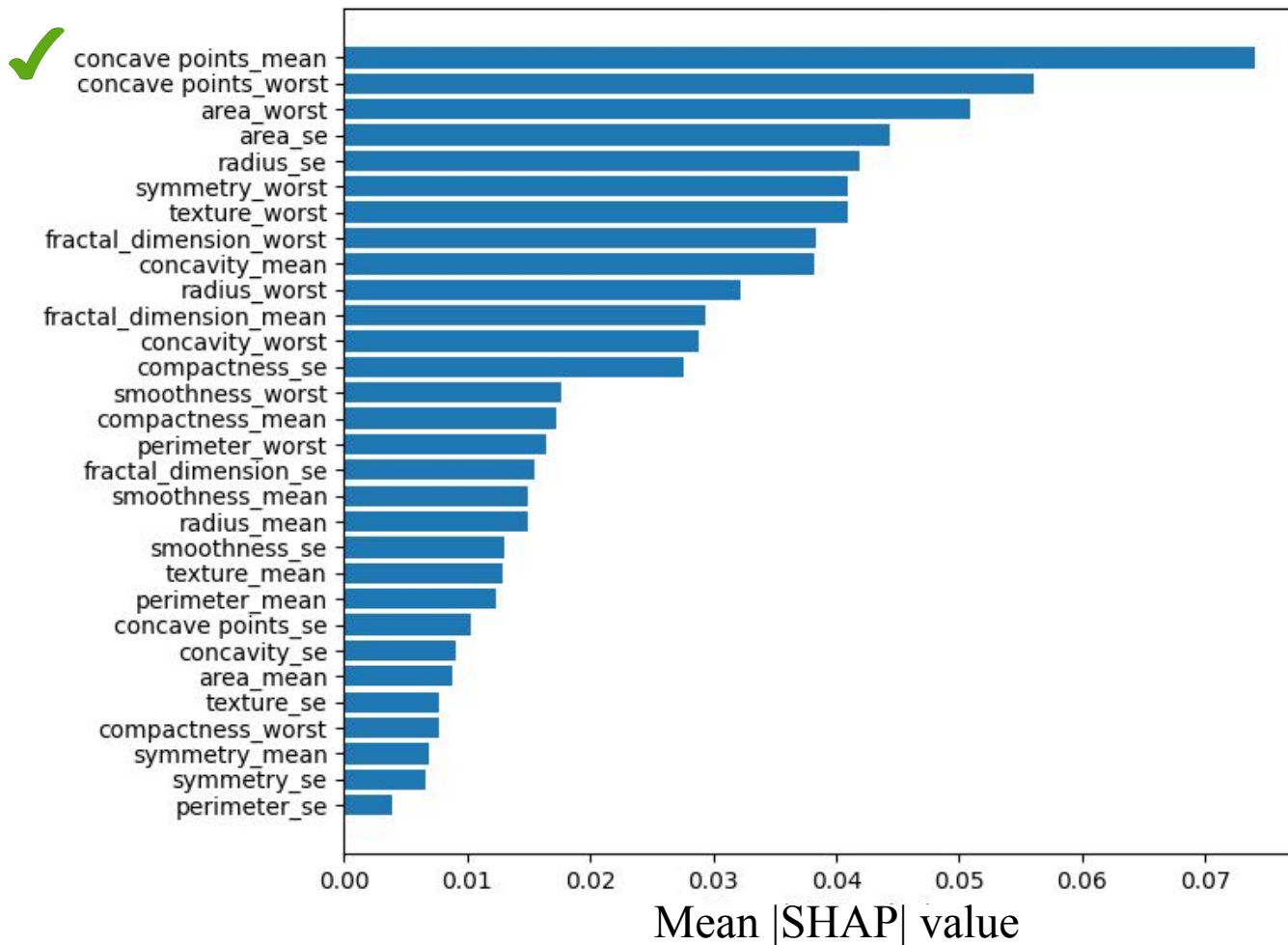
Perimeter_m

Area_w

Area_m

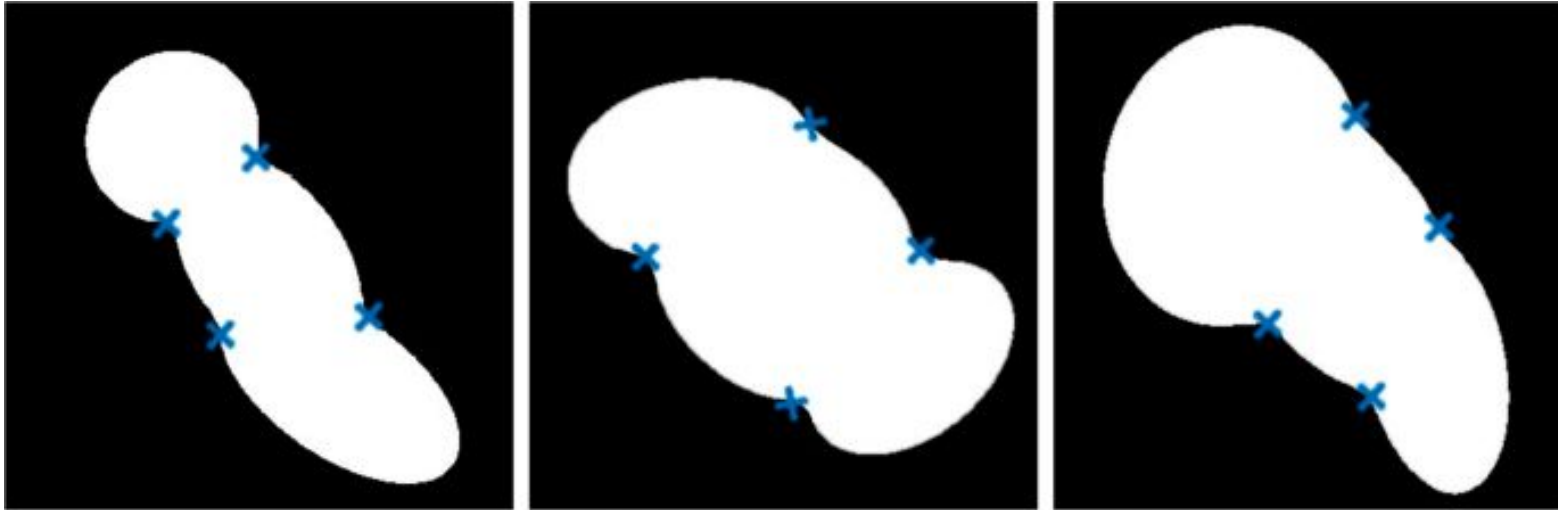
Interpreting results

SHAP Feature Importance (One-Layer Neural Network)



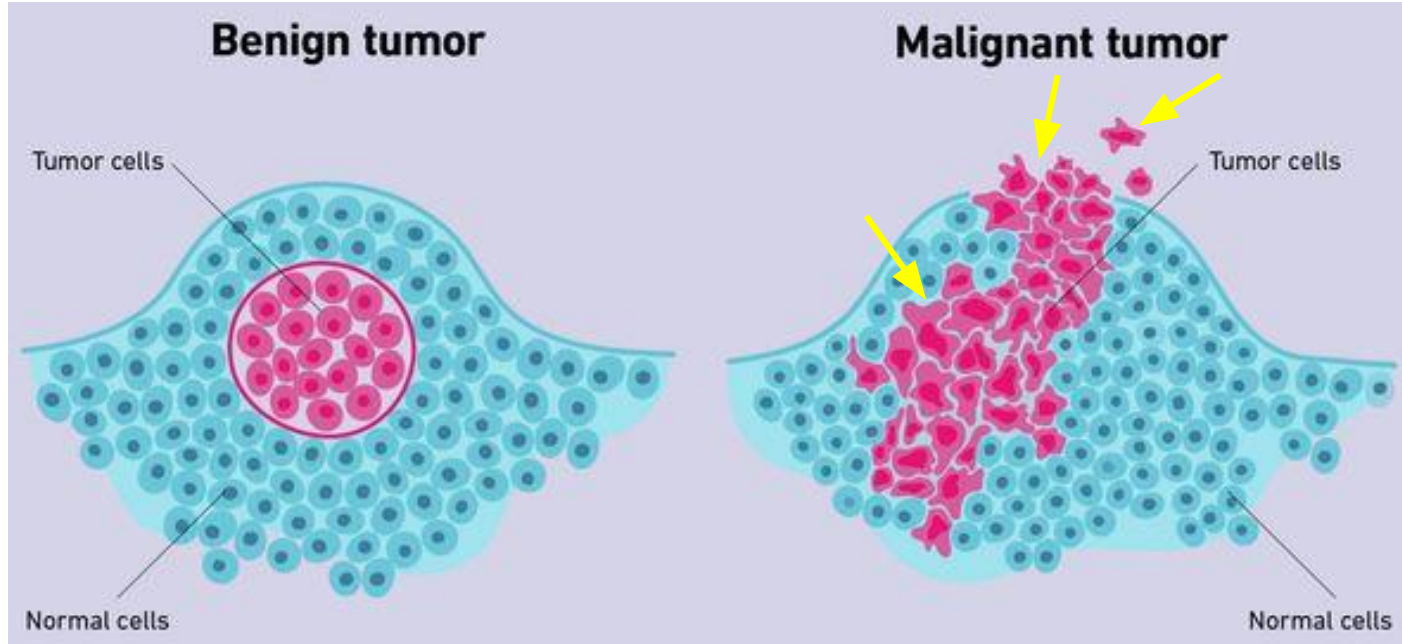
Interpreting results

- Concave points: The location where the shape bends inward like a dent.



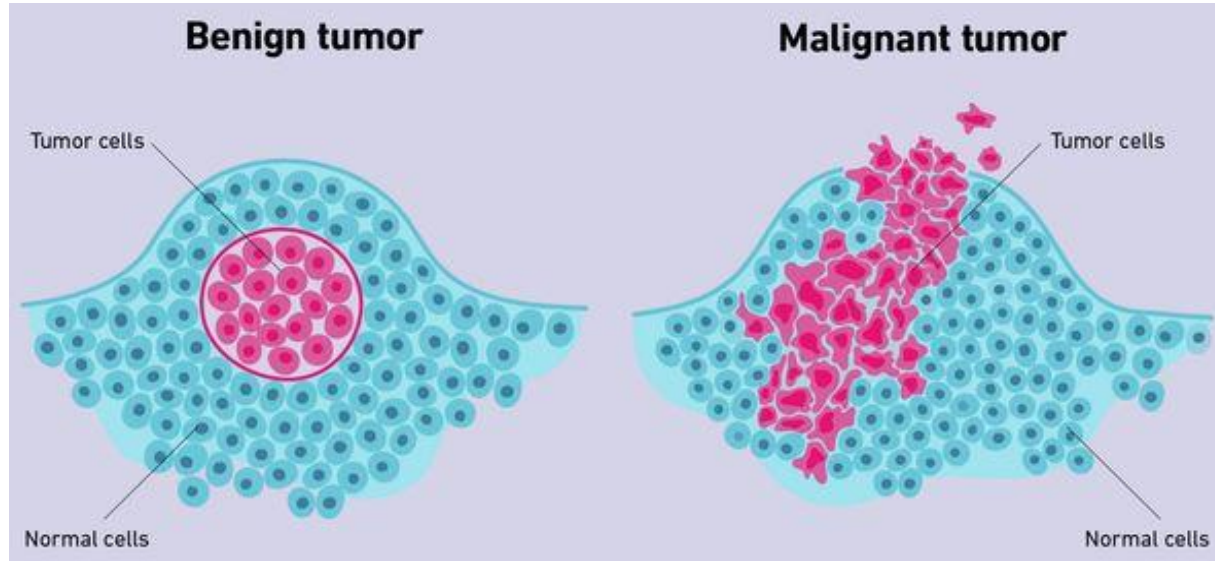
Interpreting results

- Concave points: The location where the shape bend inwards.



Interpreting results

- Geometric features that are important in distinguishing the cancer cell type.



- Radius
- Perimeter
- Area
-etc

- **Simple models are effective!**
- **Certain geometric features can be determining factors in distinguishing cancer types.**
- **More data can help in quantitative analysis.**

Thank you!