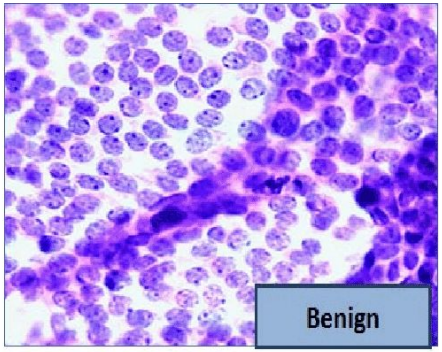
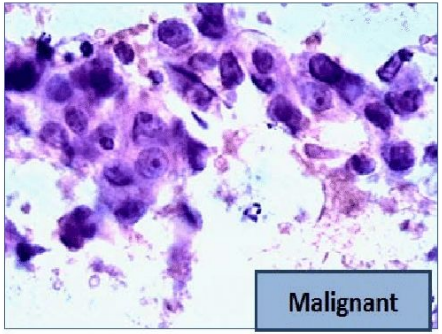
Breast Cancer Classification

James Benedito

Introduction

- **Breast Cancer**
  - United States
    - 30% of all new cancer cases in women
    - ~4 million women have a history of breast cancer
      - In 2023: ~300,000 more women are predicted to be diagnosed





https://www.researchgate.net/figure/FNA-results-for-benign-and-malignant-tumor-under-the-microscope_fig1_286571014

https://www.breastcancer.org/facts-statistics

Business Context

Problem:

- Want to increase accuracy of benign and malignant tumor classification

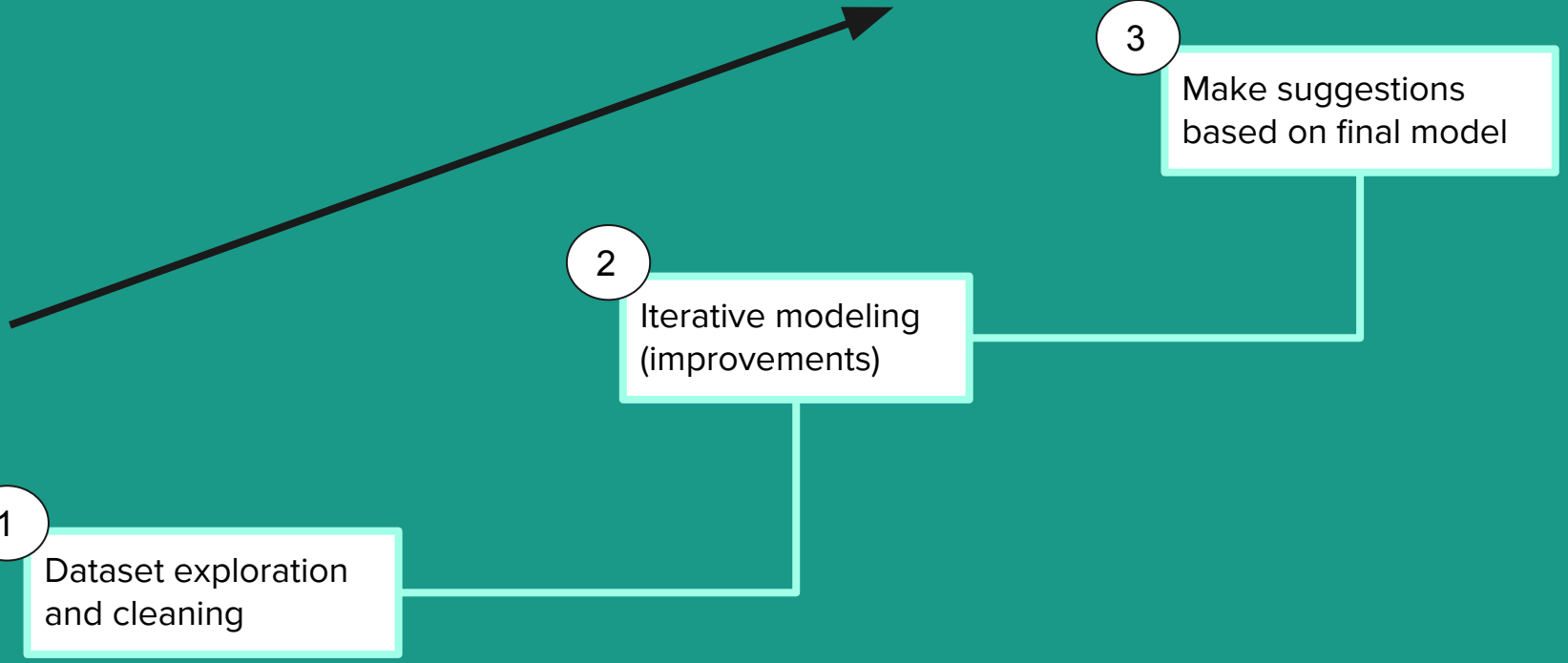
Goal:

- Generate suggestions for diagnostic “look-fors”
  - Cellular level
  - Whole-tumor level

Variables to Consider

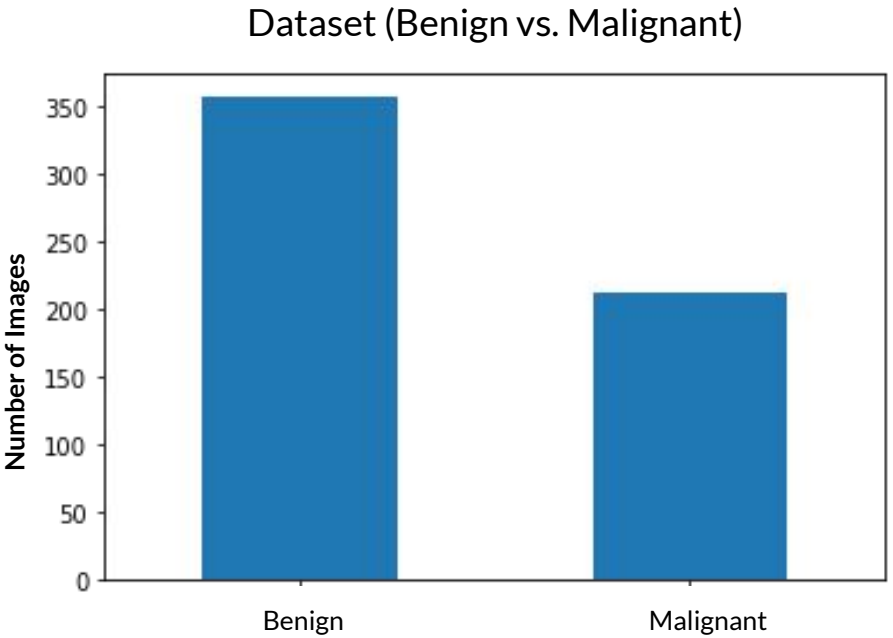
- Dependent variable: **benign (0) or malignant (1)**
  - Cancerous or not?
- Decide on best indicators of malignancy
  - Determined by **classification models**
  - Most important features (characteristics to watch out for)

Methods (Process Steps)



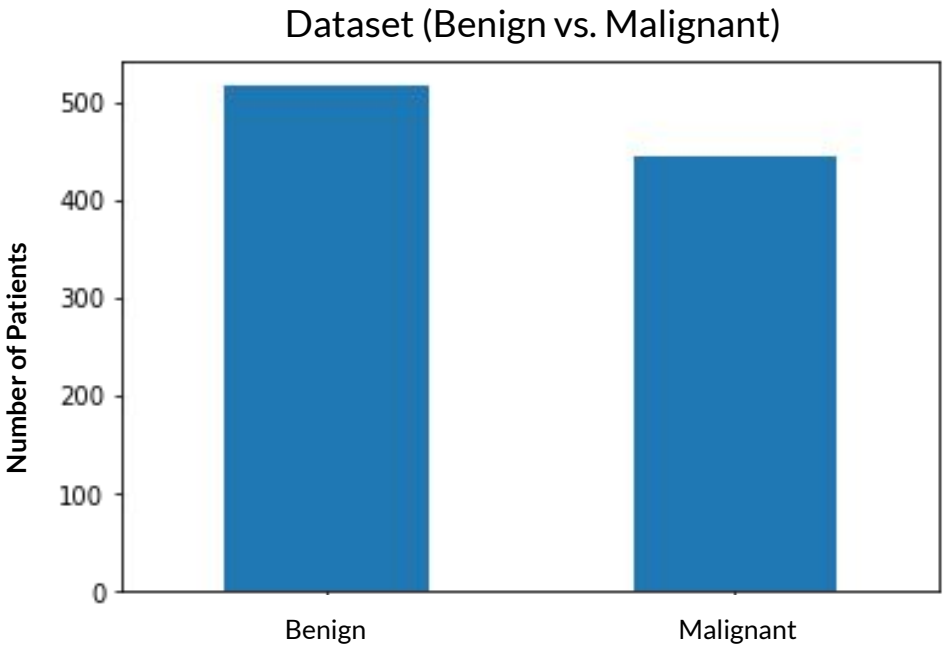
Dataset 1

- **Breast Cancer Dataset**
  - **569 FNA images**
    - 10 cellular features
      - Nuclear size, shape, texture
    - For each cellular feature:
      - Mean value
      - Standard error (SE)
      - Largest (“worst”)



Dataset 2

- **Mammographic Masses Dataset**
  - **830 patients**
    - 5 attributes
      - BIRADs score
      - Age
      - Tumor shape
      - Tumor margin
      - Density

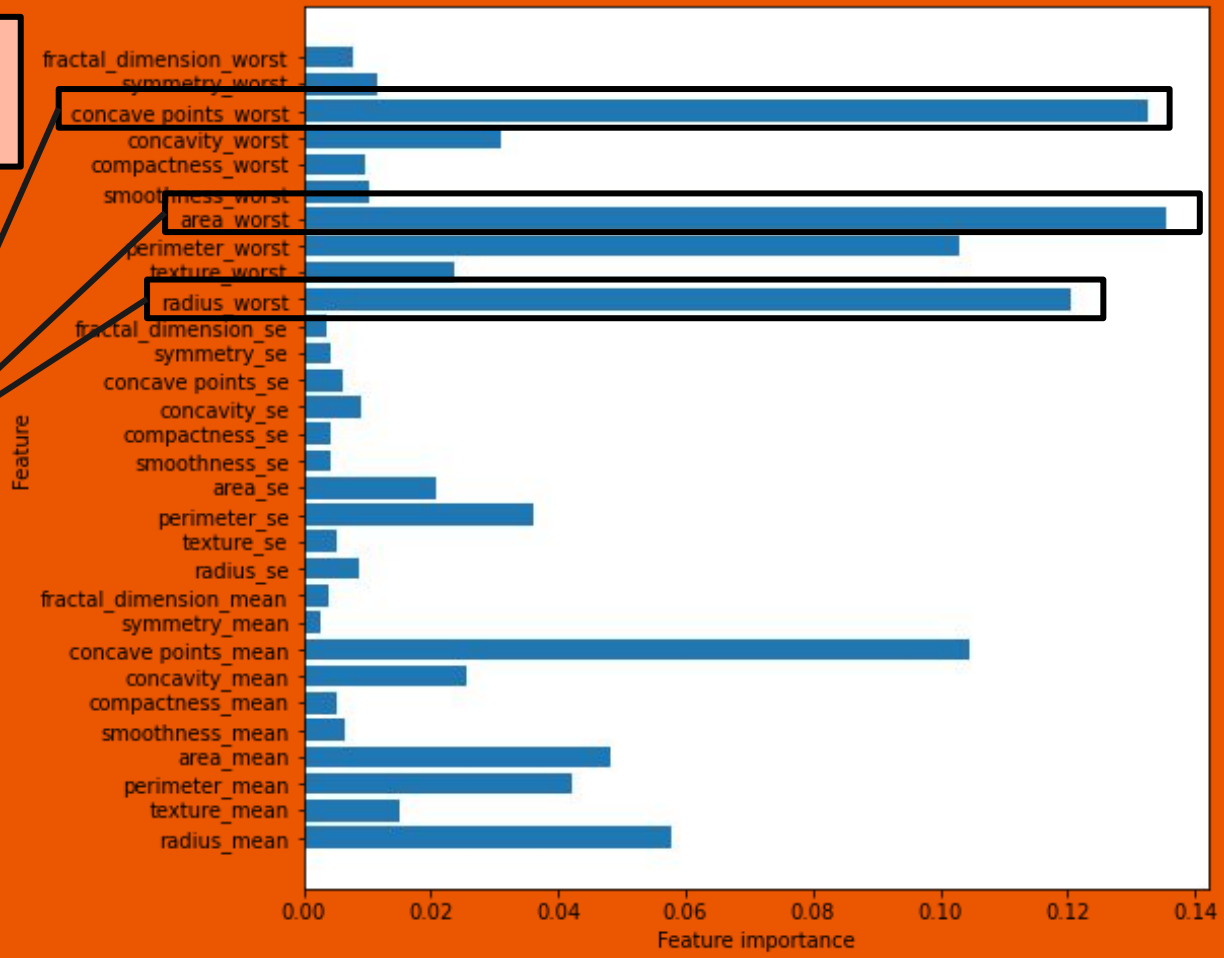


Final Classification Model for Breast Cancer Data

Had an **accuracy** of **98.2%**

Most important features:

- **Largest area**
- **Largest number of concave points**
- **Largest radius**

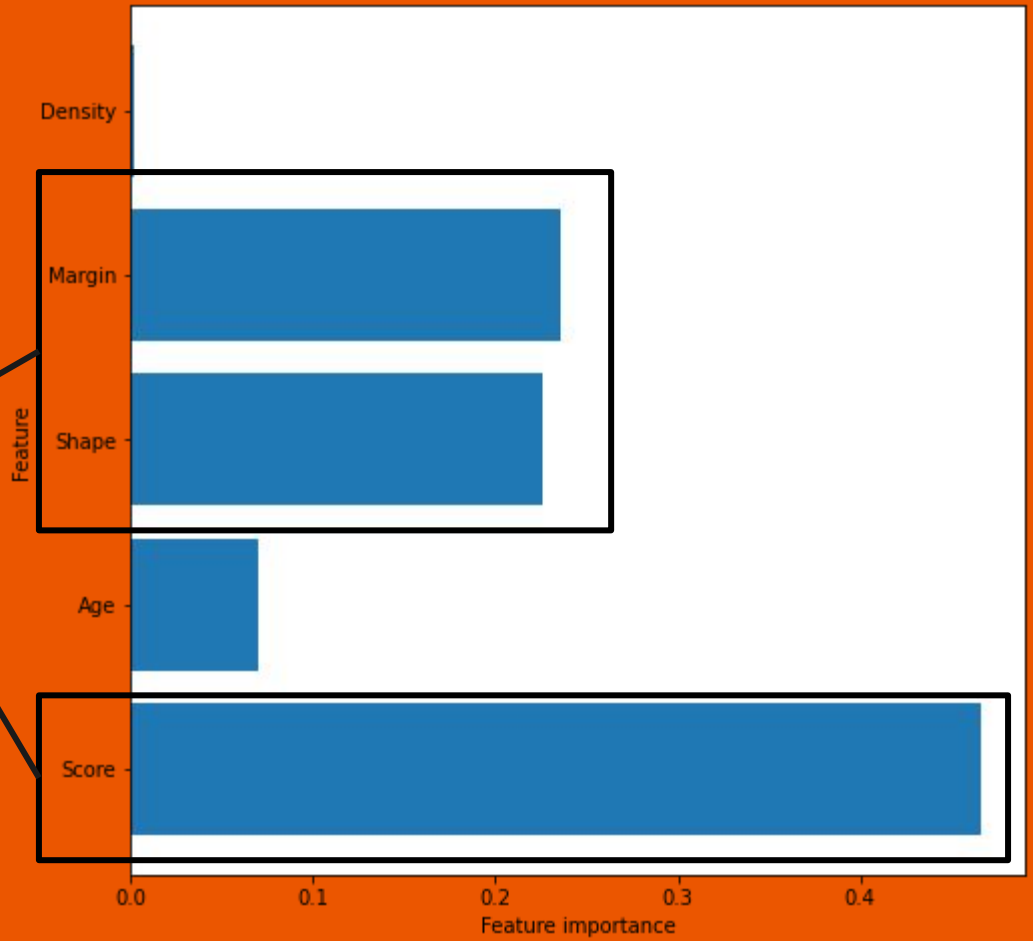


Final Classification Model for Mammographic Masses Data

Had an **accuracy** of **86.8%**

Most important features:

- **BIRADS Score**
- **Tumor Margin**
- **Tumor Shape**



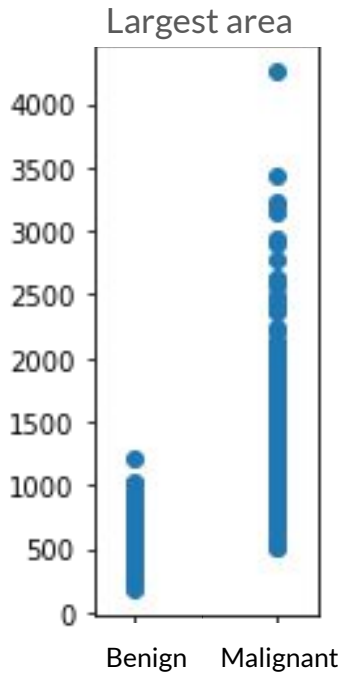
Results & Business Recommendations

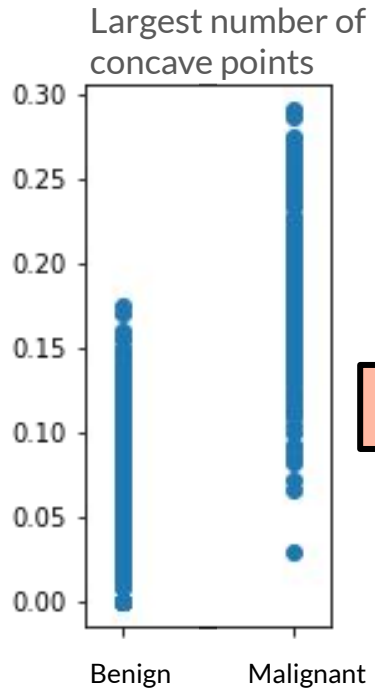
Recommendation #1

For **FNA images**, the **larger** the…
- **Largest area**
- **Largest number of concave points**
- **Largest radius**

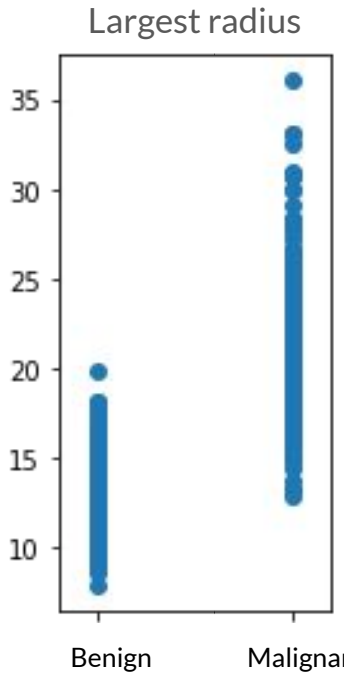
the **more likely** a tumor is **malignant**

r=0.73



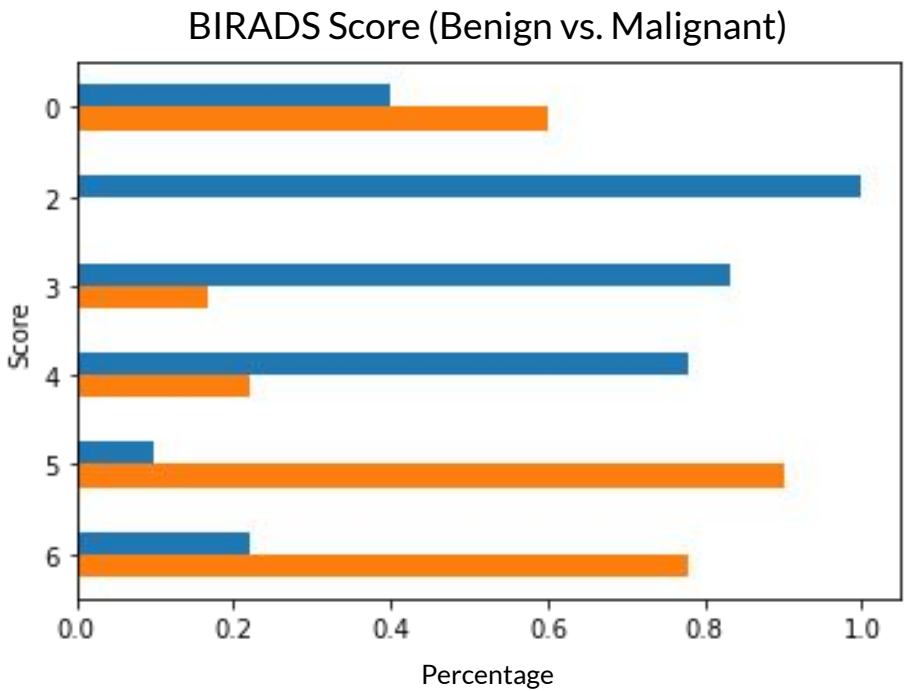


r=0.79



r=0.78

Recommendation #2: BIRADS Score

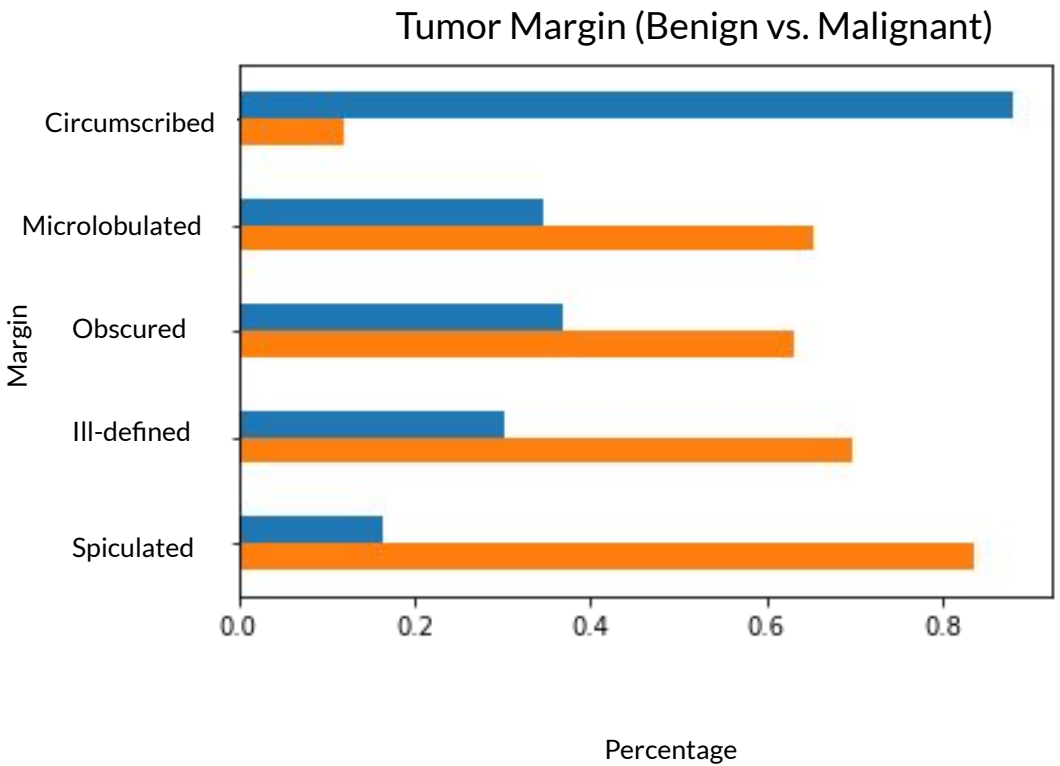


BIRADS score of **5** or **6** indicates a high chance of malignancy

Benign

Malignant

Recommendation #3: Tumor Margin

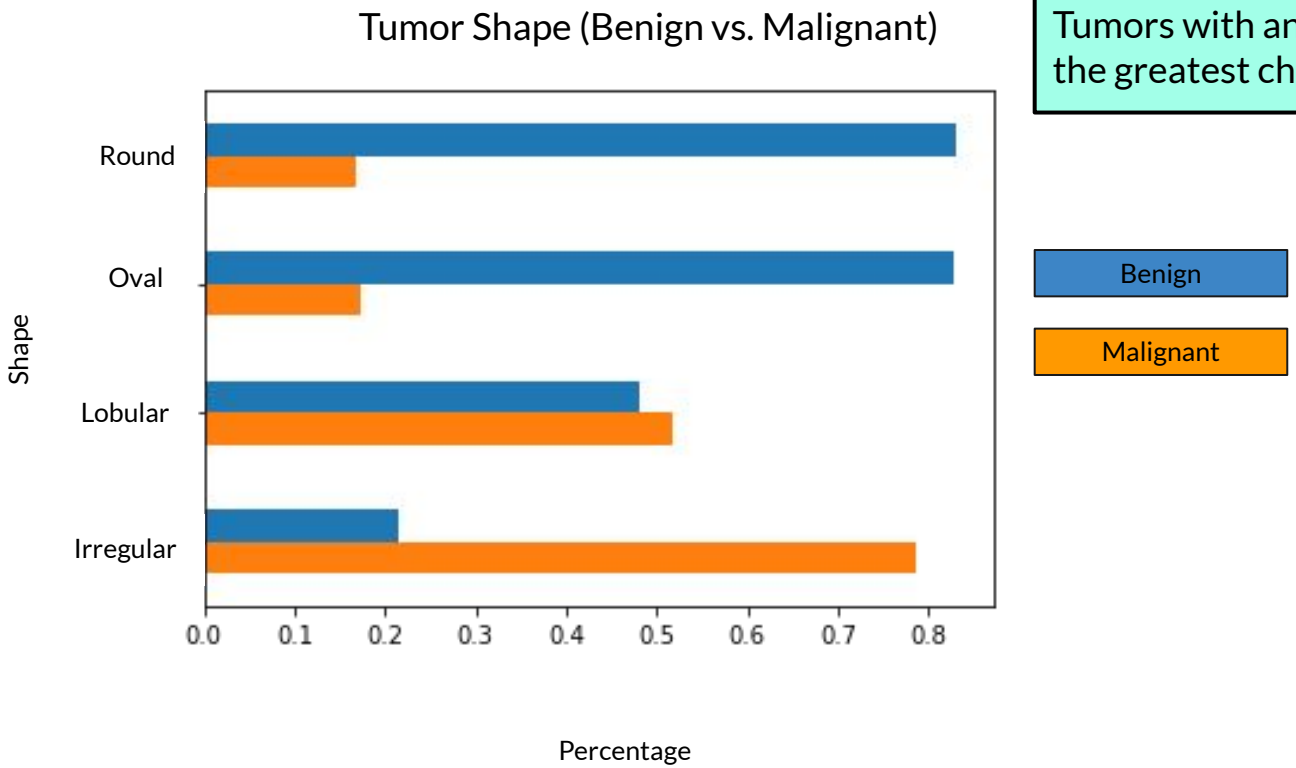


Tumors with a **spiculated margin** have the greatest chance of being malignant

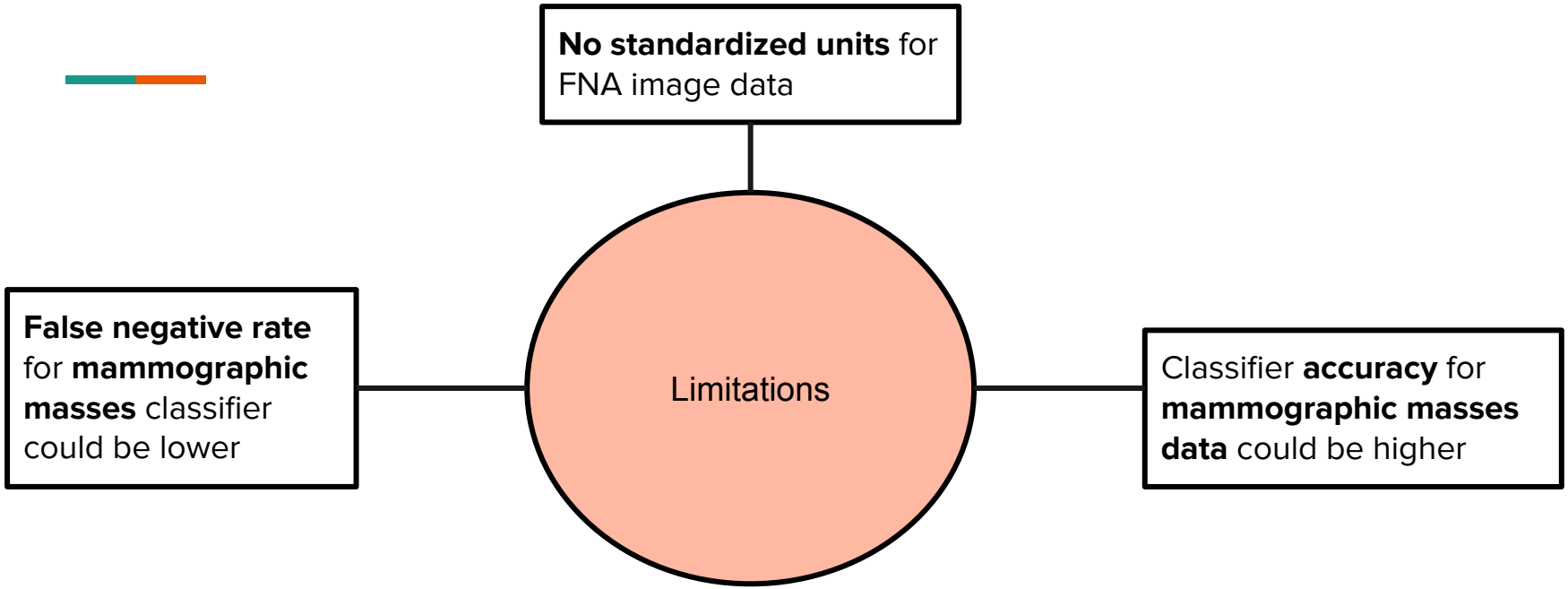
Benign

Malignant

Recommendation #4: Tumor Shape



Tumors with an **irregular shape** have the greatest chance of being malignant


No standardized units for FNA image data
False negative rate for **mammographic masses** classifier could be lower
Limitations
Classifier **accuracy** for **mammographic masses data** could be higher

Summary

- To determine malignancy, look for…
  - Cellular level
    - **Largest area**
    - **Largest number of concave points**
    - **Largest radius**
  - Whole-tumor level
    - **BIRADS score** of **5** or **6**
    - **Spiculated tumor margin**
    - **Irregular tumor shape**

Contact

James Benedito
jbenedito99@gmail.com
(702) 523-2734
LinkedIn:
www.linkedin.com/in/james-benedito-7b326b217