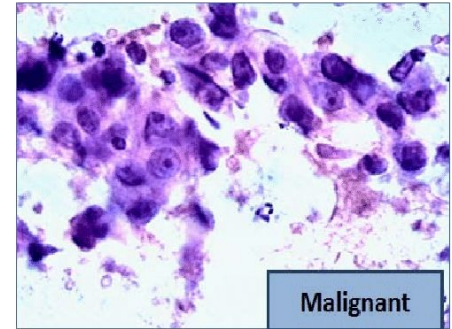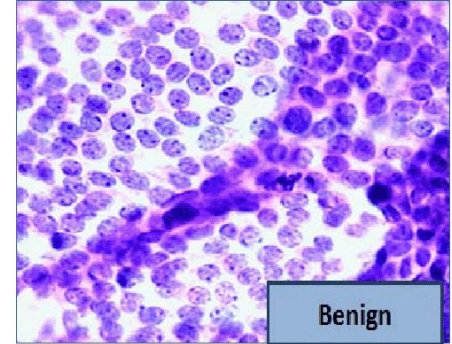# Breast Cancer Classification

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

Benign

Malignant

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
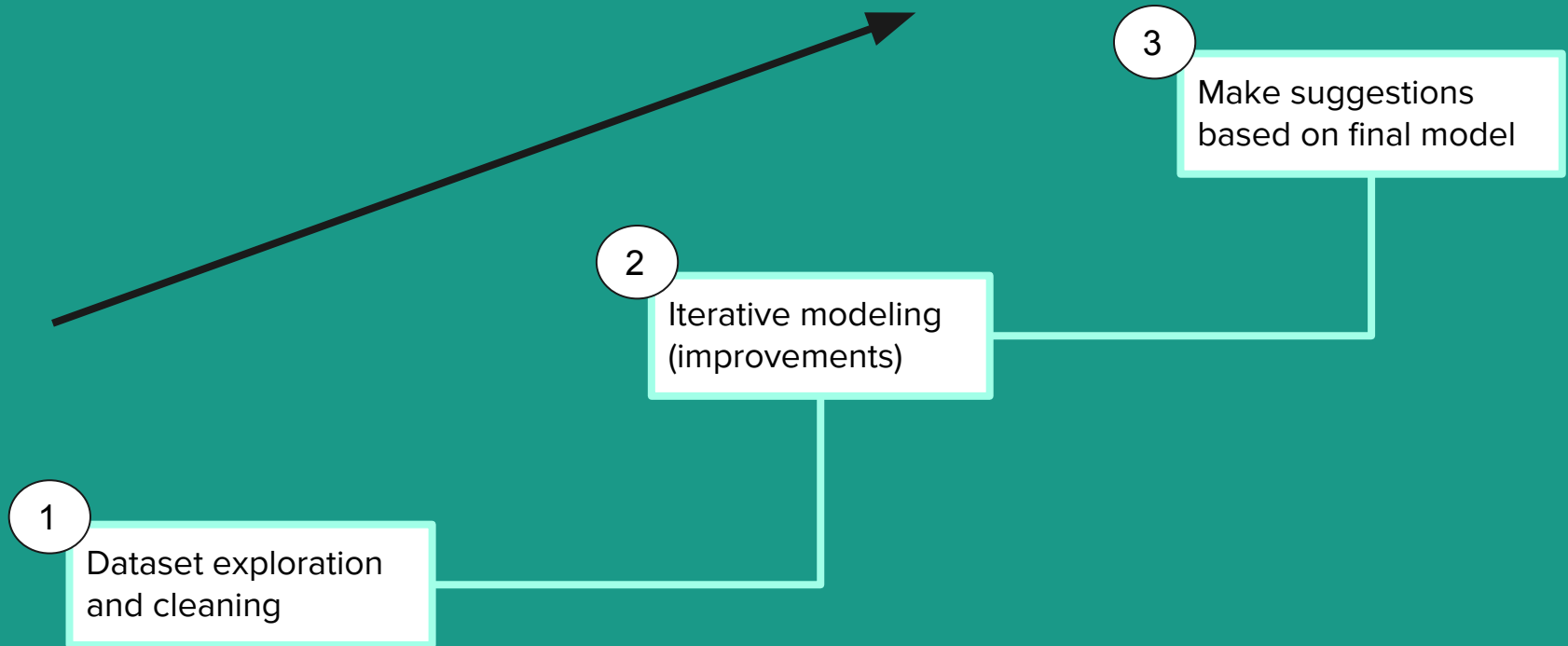
Goal:

- 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)

1 Dataset exploration and cleaning

2 Iterative modeling (improvements)

3 Make suggestions based on final model

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 (Benign vs. Malignant)
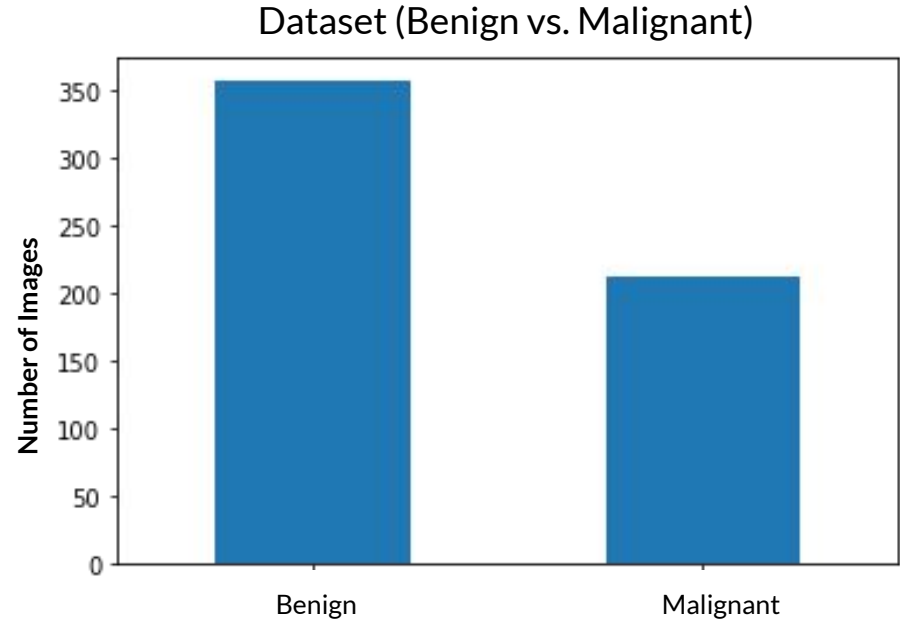
Dataset 2

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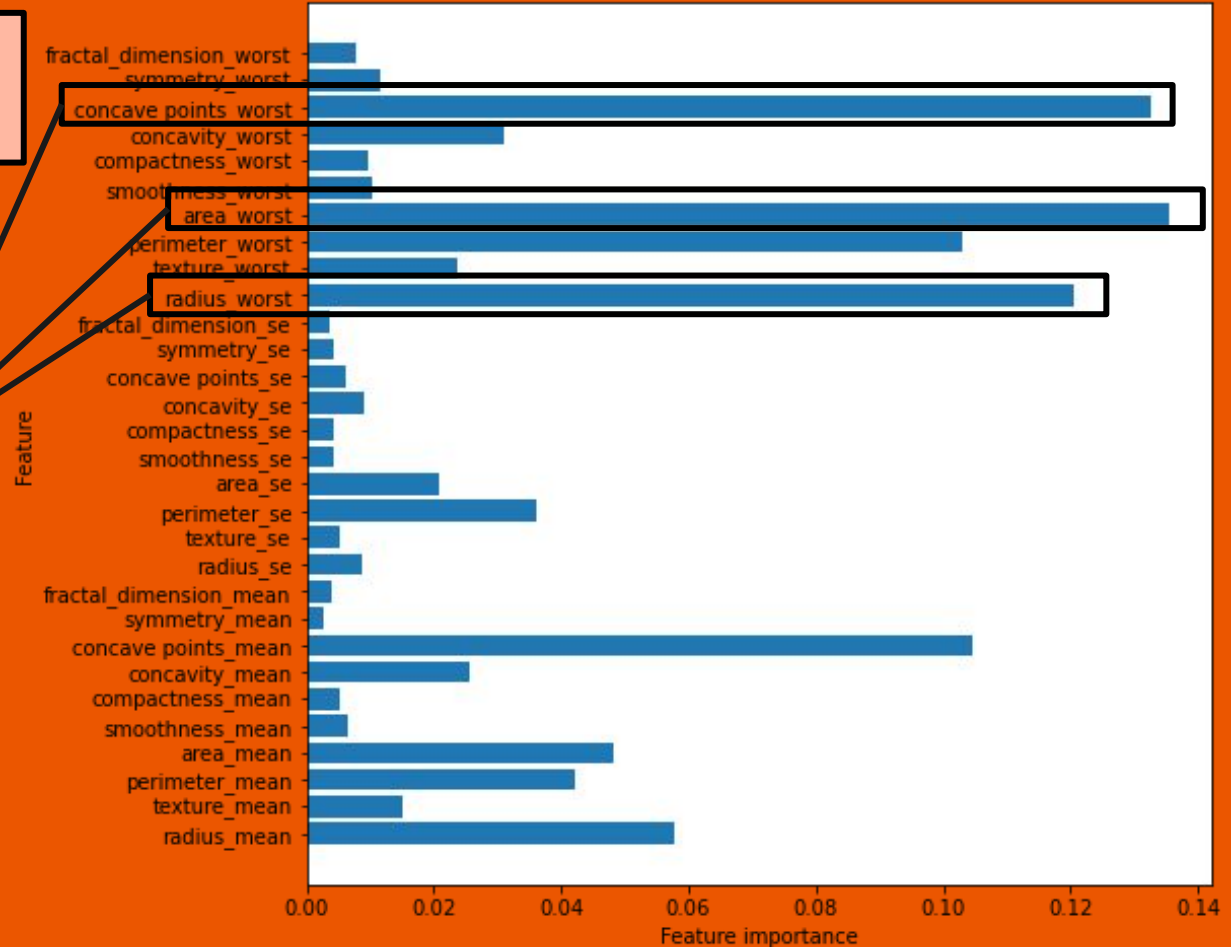   - Tumor shape
      - Tumor margin
      - Density

Dataset (Benign vs. Malignant)
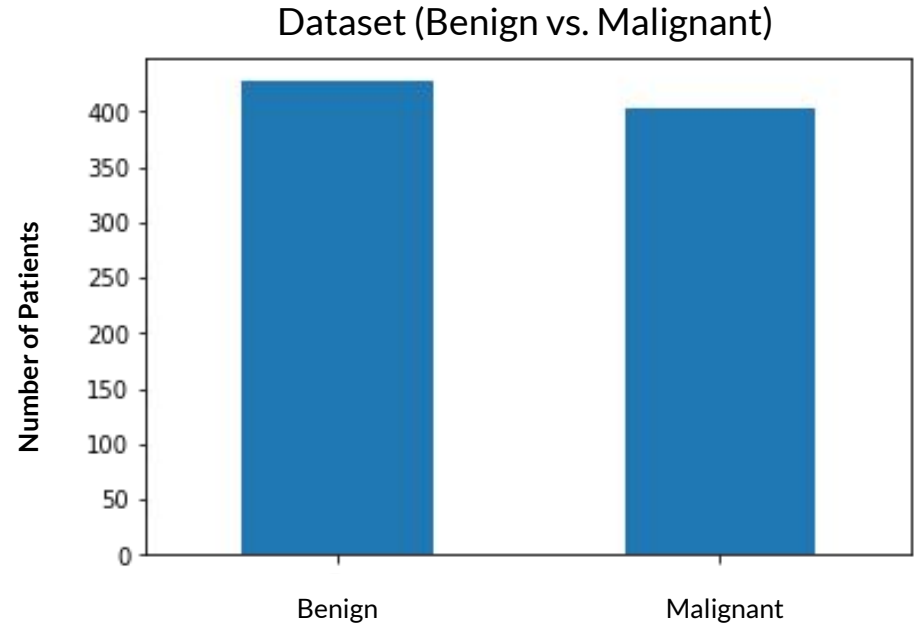
Final Classification Model for Breast Cancer Data

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

Most important features:
- **Largest area**
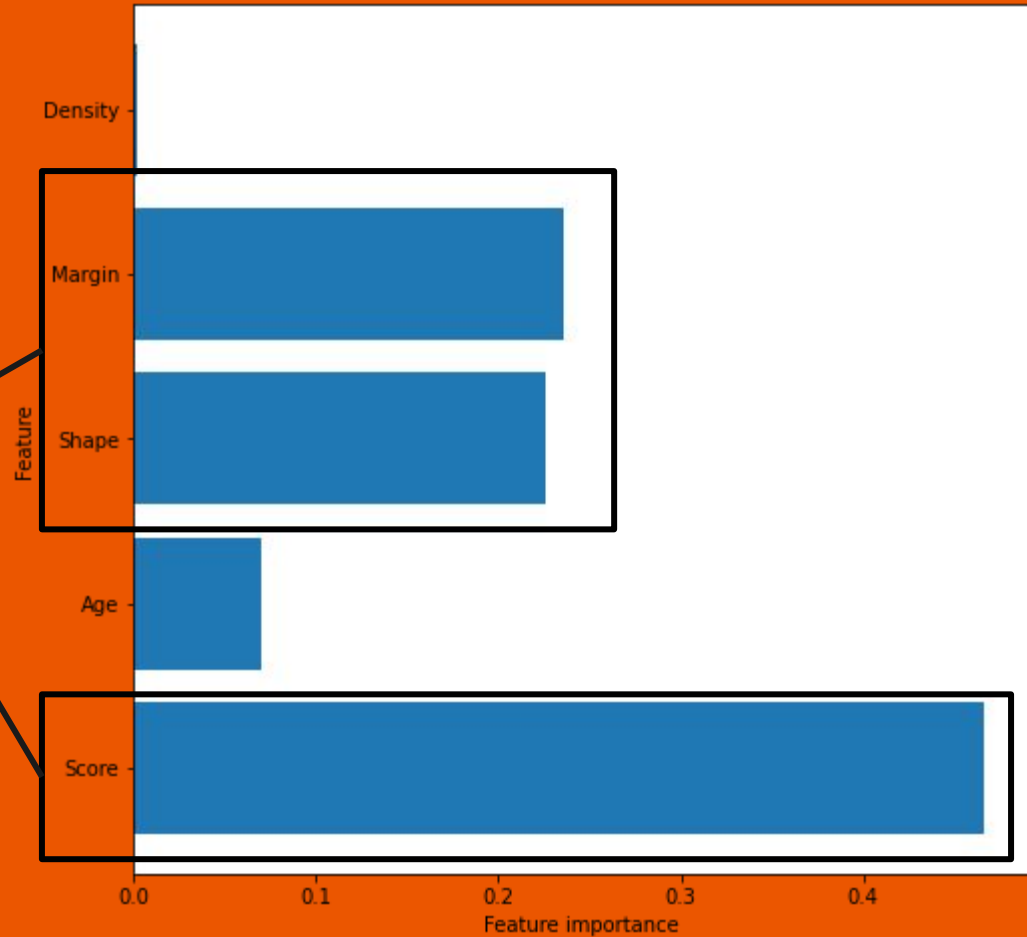- **Largest number of concave points**
- **Largest radius**

Final Classification Model for Mammographic Masses Data

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

Most important features:
- **BIRADS Score**
- **Tumor Margin**
- **Tumor Shape**
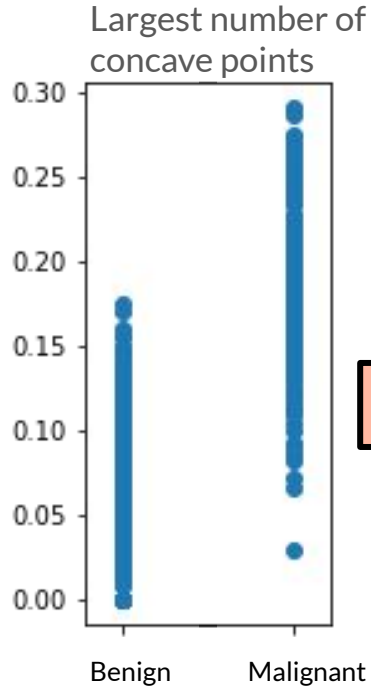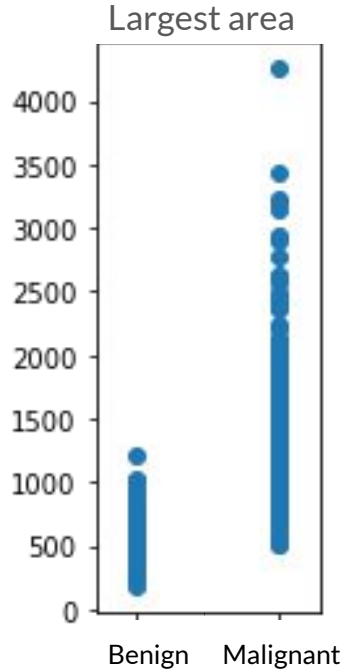
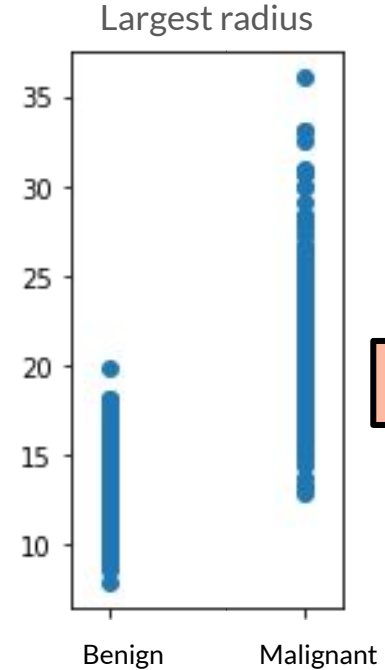# Results & Business Recommendations

Recommendation #1

Largest area

r=0.73

Largest number of concave points

r=0.79

Largest radius

r=0.78

Recommendation #2: BIRADS Score

BIRADS Score (Benign vs. Malignant)
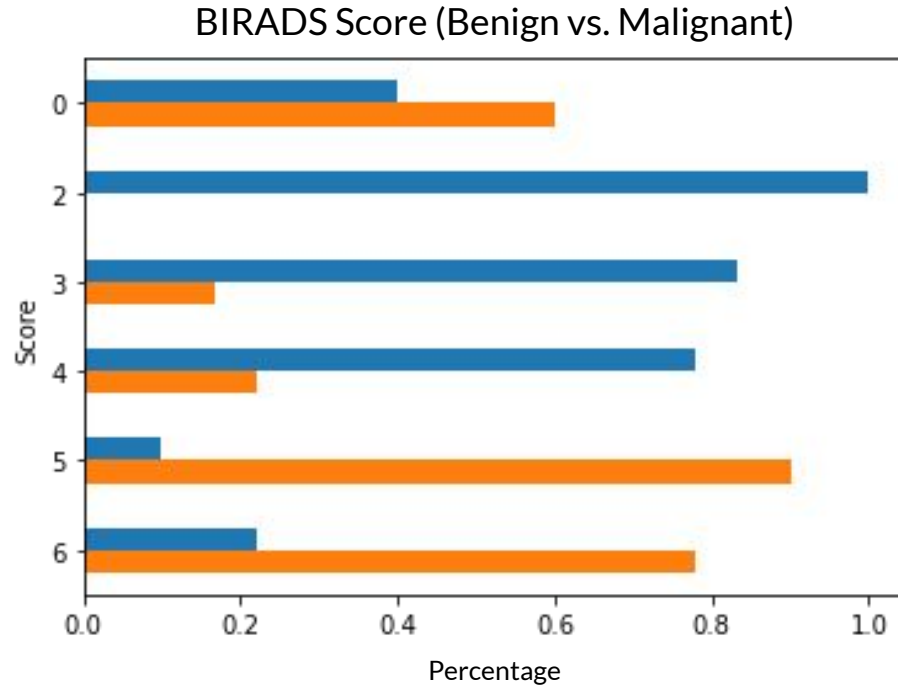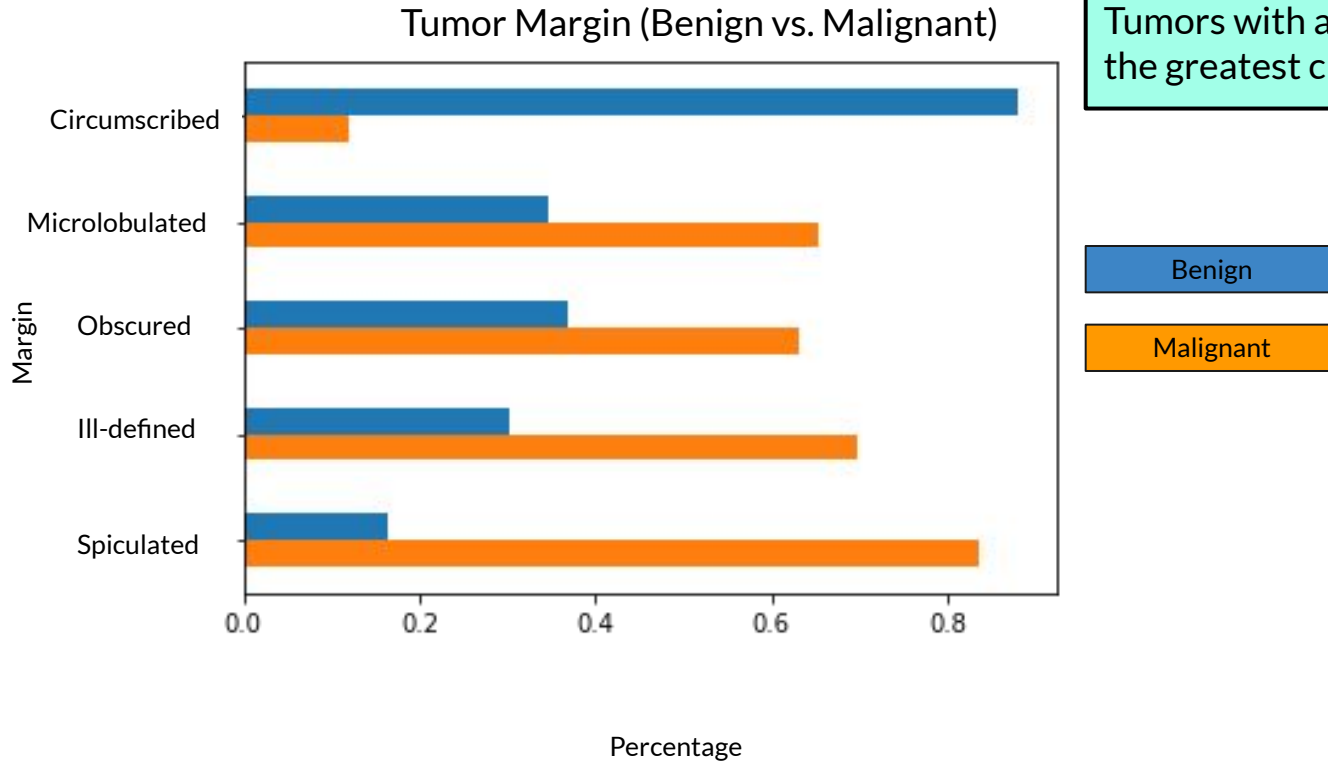

BIRADS score of **5** or **6** indicates a high chance of malignancy

Benign

Malignant

Recommendation #3: Tumor Margin

Tumor Margin (Benign vs. Malignant)
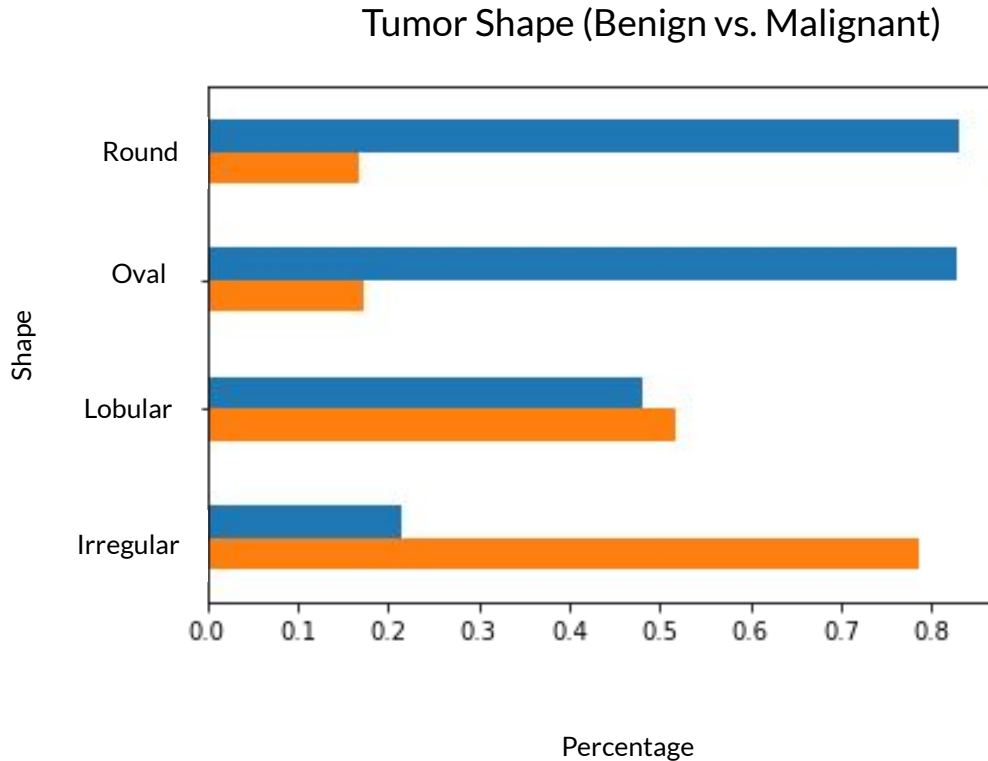
Tumors with a **spiculated margin** have the greatest chance of being malignant

Recommendation #4: Tumor Shape

Tumor Shape (Benign vs. Malignant)

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

Benign

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
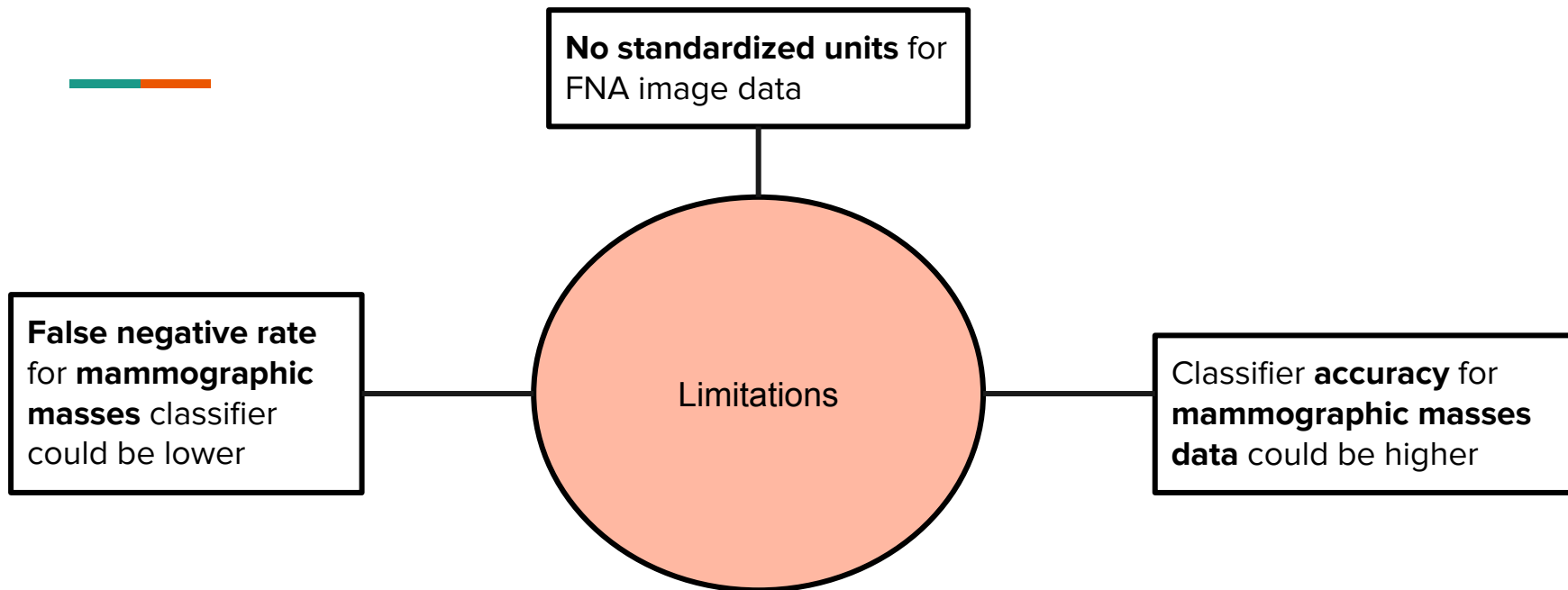
Summary

- 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