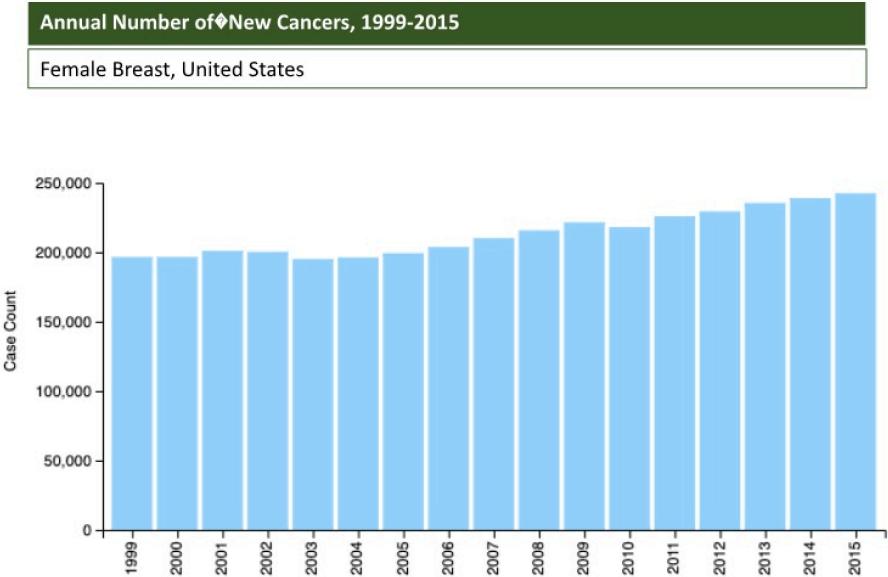


# TUMOR CLASSIFICATION IN BREAST CANCER PATIENTS

Jacob Niederer

# BACKGROUND

**CDC** Centers for Disease Control and Prevention  
CDC 24/7: Saving Lives, Protecting People™



Most common form of cancer affecting women in the United States (except for skin cancer), regardless of race or ethnicity

Every year about 240,000 cases of breast cancer are diagnosed:

- About 40,000 female and 500 male deaths each year

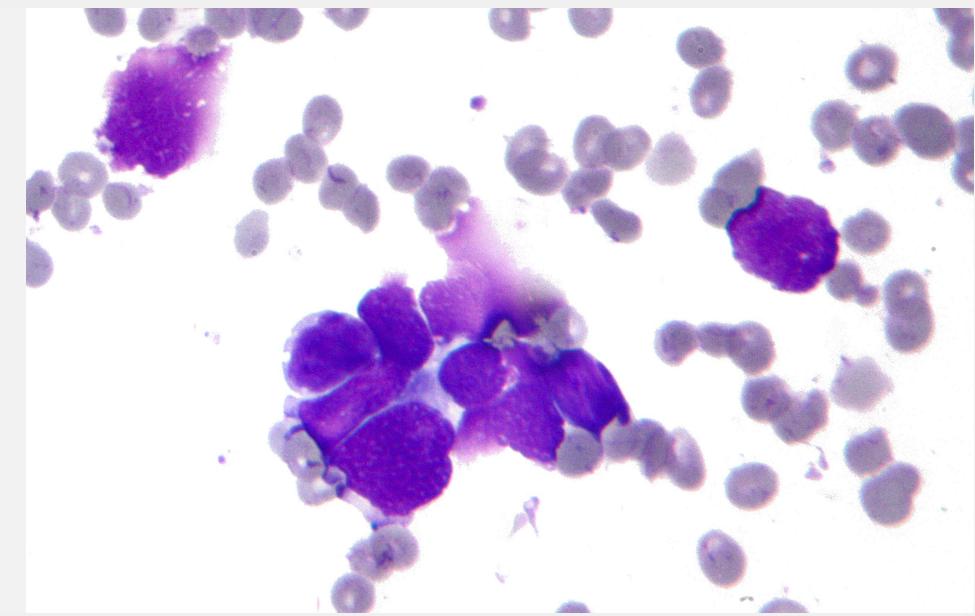
**Objective:** Identify diagnostic features in malignant breast tissue

# DIAGNOSTIC BREAST CANCER DATA

University of Wisconsin Diagnostic Data

Cell nuclei features computed from digitized  
images of fine needle aspirate (FNA)

Class Labels: Benign/Malignant



# CLASSIFICATION METRICS

Classification Models:

k-Nearest Neighbors, Logistic Regression,  
Naive Bayes, Decision Tree, Random Forest

Performance Metrics:

Recall >>> Accuracy > ROC AUC

		Predicted class	
		+	-
Actual class	+	TP True Positives	FN False Negatives Type II error
	-	FP False Positives Type I error	TN True Negatives

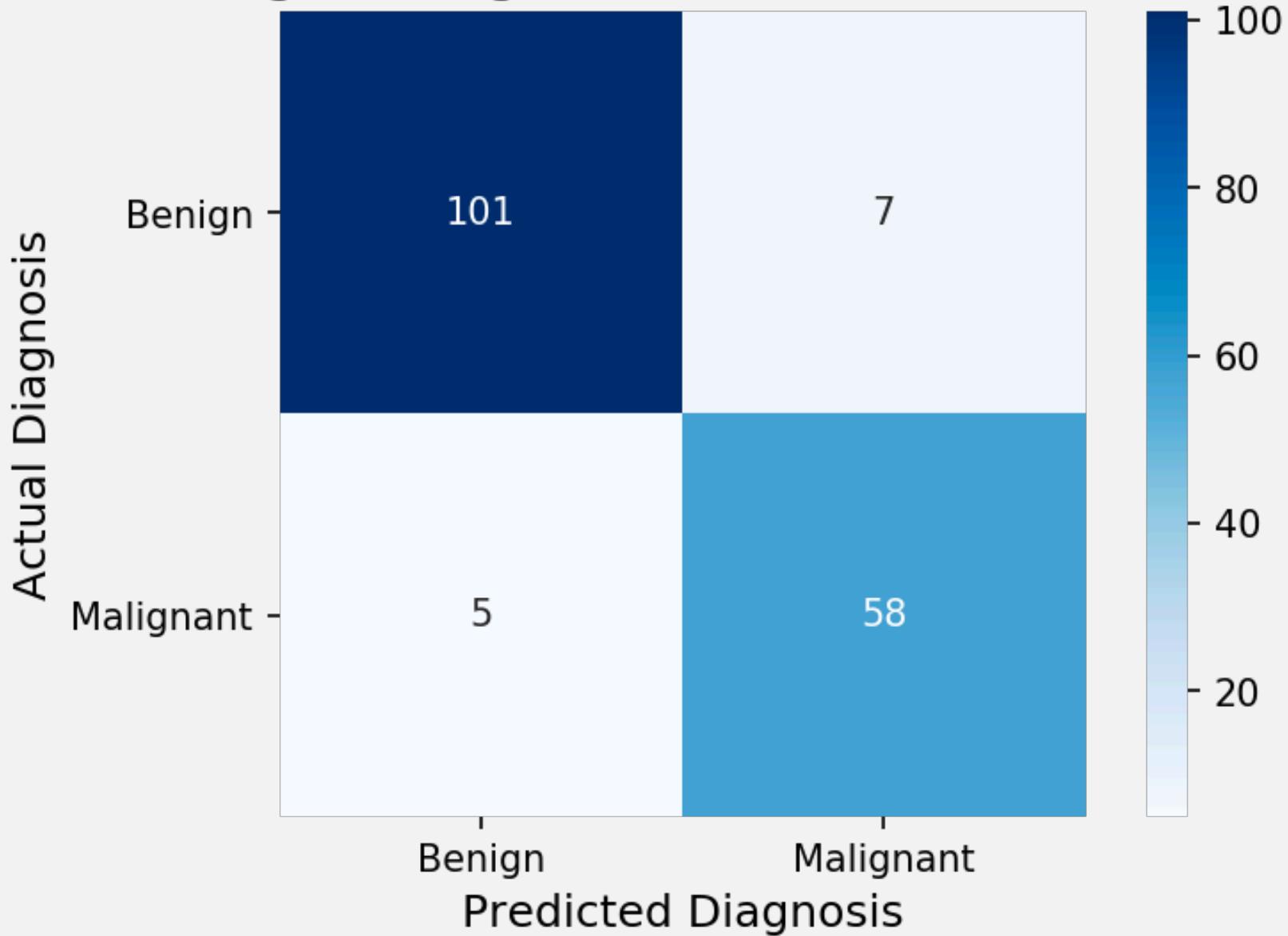
$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total}}$$

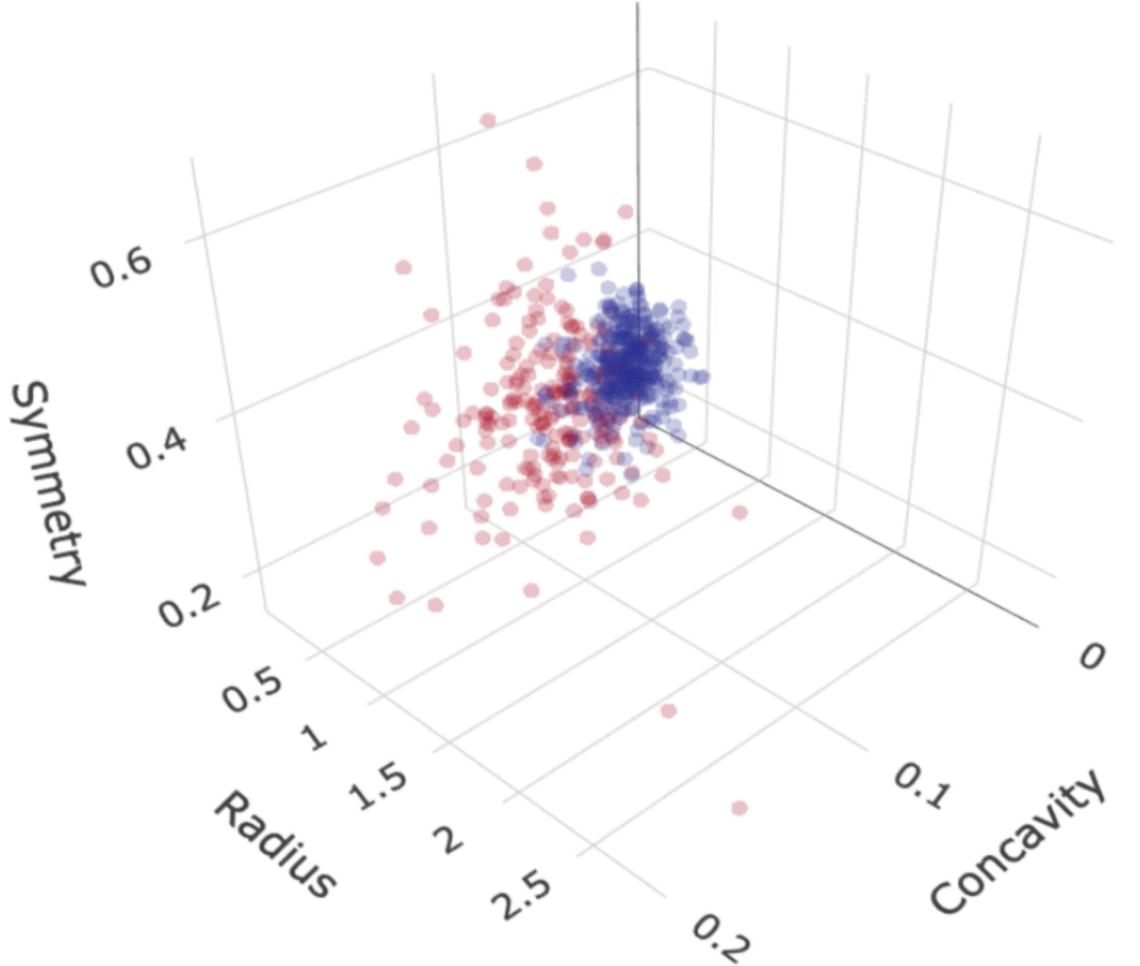
$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

	Recall	Accuracy	ROC AUC
kNN	87.3	92.39	91.99
Logistic Regression	92.06	92.98	98.58
Naive Bayes	85.71	90.05	96.66
Decision Tree	88.88	90.64	89.81
Random Forest	88.88	92.98	97.97

	Recall	Accuracy	ROC AUC
kNN	87.3	92.39	91.99
Logistic Regression	92.06	92.98	98.58
Naive Bayes	85.71	90.05	96.66
Decision Tree	88.88	90.64	89.81
Random Forest	88.88	92.98	97.97

## Logistic Regression Confusion Matrix





## CONCLUSION

- Tumor Classification with Logistic Regression:
  - Recall: 92.06%
  - Accuracy: 92.98%
  - ROC AUC: 98.58%
- Feature Importance:
  - Concavity > Radius > Symmetry
- Detect and diagnose malignant breast cancer tumors

