Breast Cancer Diagnosis

Dhvani Patel

Statistical Facts

30% Cancer

In Women will be Breast Cancer in 2020

42,170 Deaths

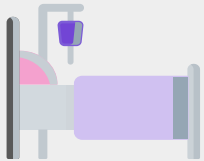
Expected In Year 2020

3.5M Cases

As of 2020 in the US alone

Ref: https://www.breastcancer.org/

Methods of Diagnosis



Surgery

100% Sensitive
Expensive
Time Consuming

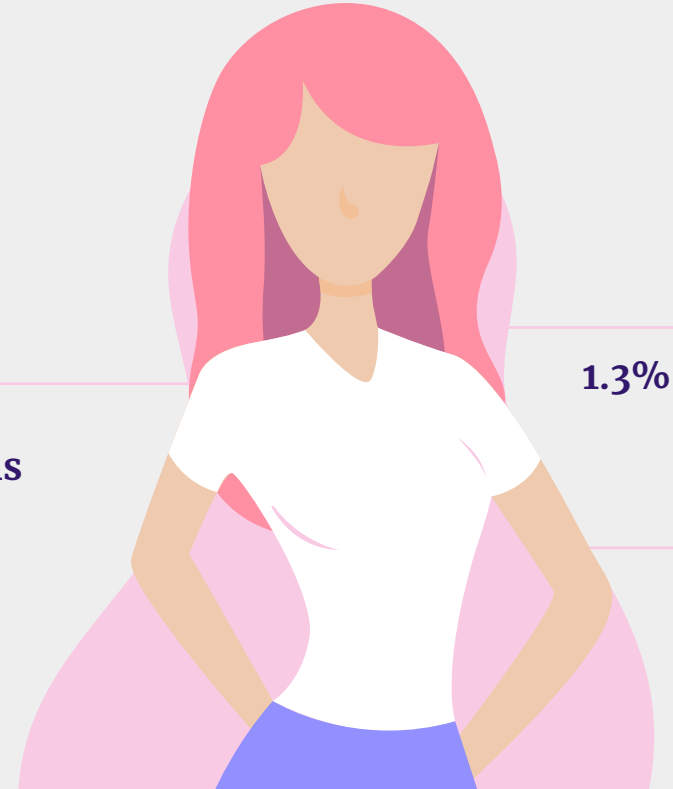
FNAC

65-98% Sensitivity
Varies Widely

Mammography

68-79% Sensitivity

More Stats !!!



93% Women

Survived Beyond 5 Years

Early Detection

1.3% Death Rate Decrease

Per Year from 2013 to 2017

Accurate Diagnosis

5Year Survival Rate

91%

NORMAL CELLS

CANCEROUS CELLS

Many cells that continue to grow and divide

Variations in size and shapes of cells

Nucleus that is larger and darker than normal

Abnormal number of chromosomes arranged in a disorganized fashion

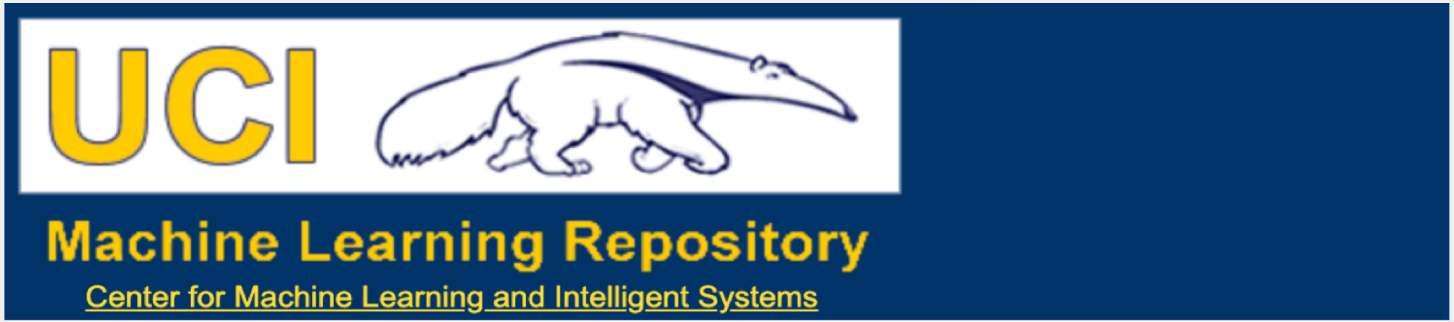
Cluster of cells without a boundry

Dataset Collection !!!

Fine Needle Aspiration from Breast Tumors

Stained on Glass Slide

Computer Vision Diagnostic System



Dataset Features

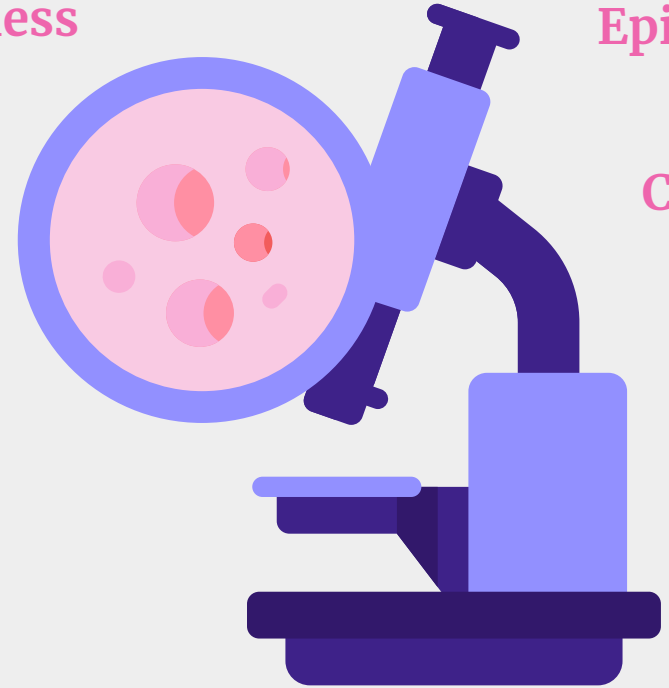
Clump Thickness

Cell Adhesion

Normal Nuclei

Bland Chromatin

Mitosis



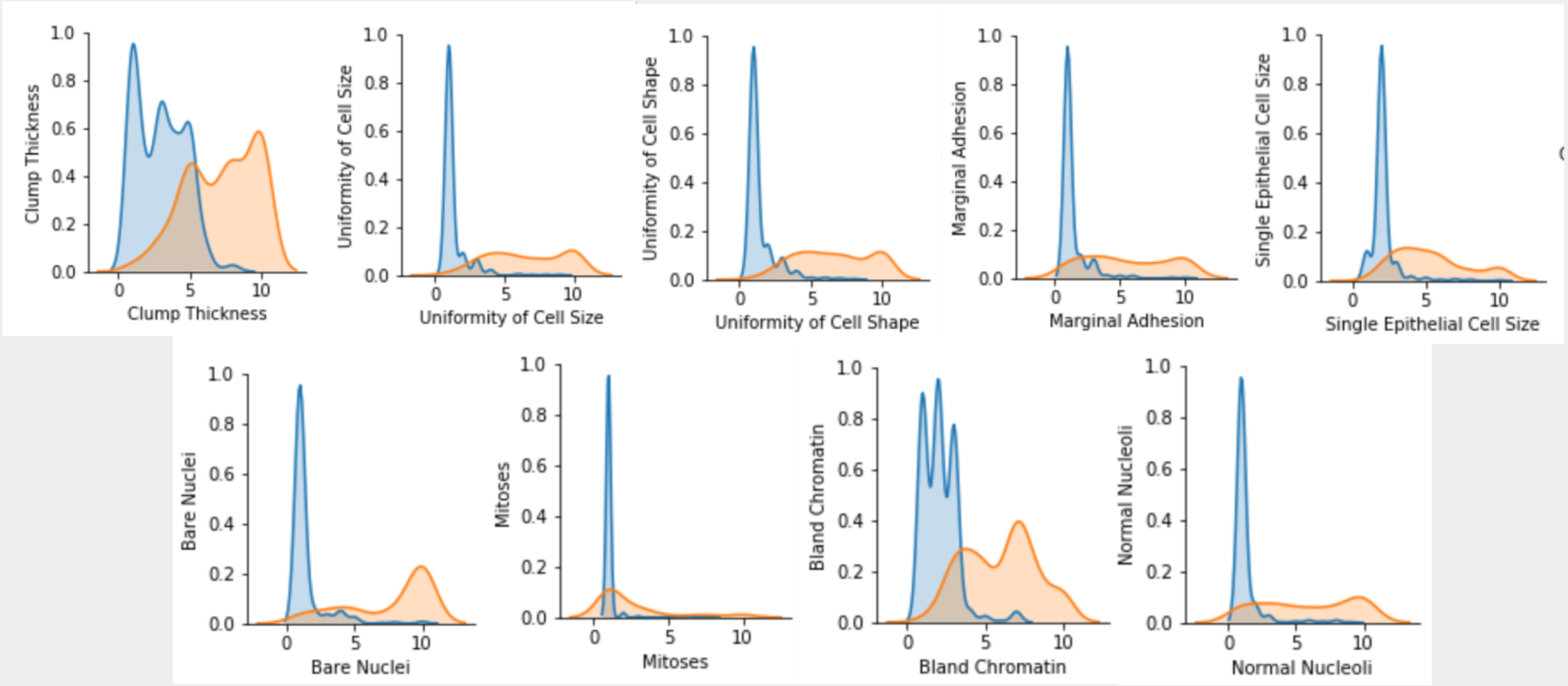
Epithelial Cell Size

Cell Size Uniformity

Cell Shape Uniformity

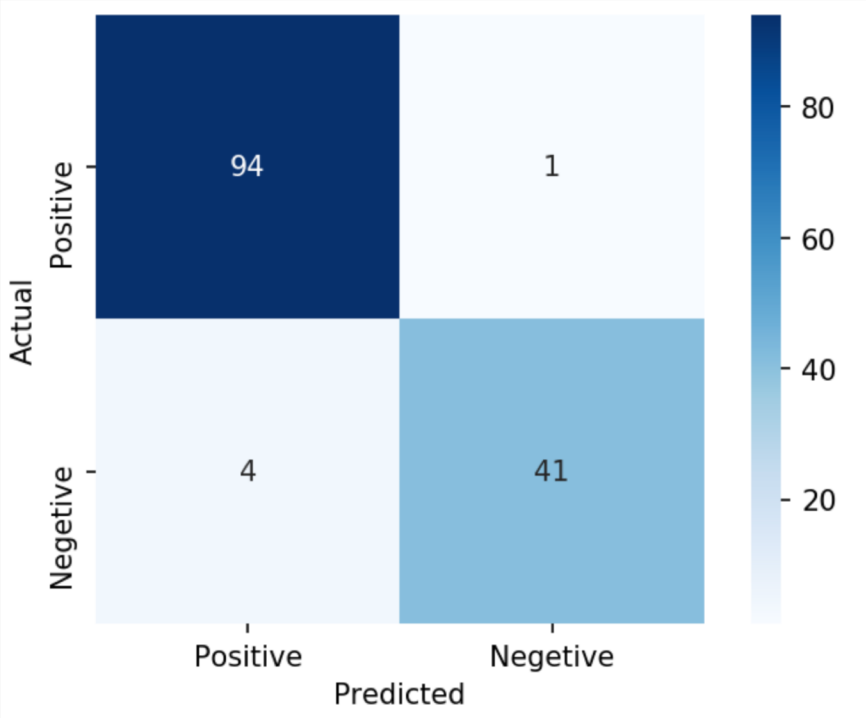
Bare Nuclei

High Separation in Features



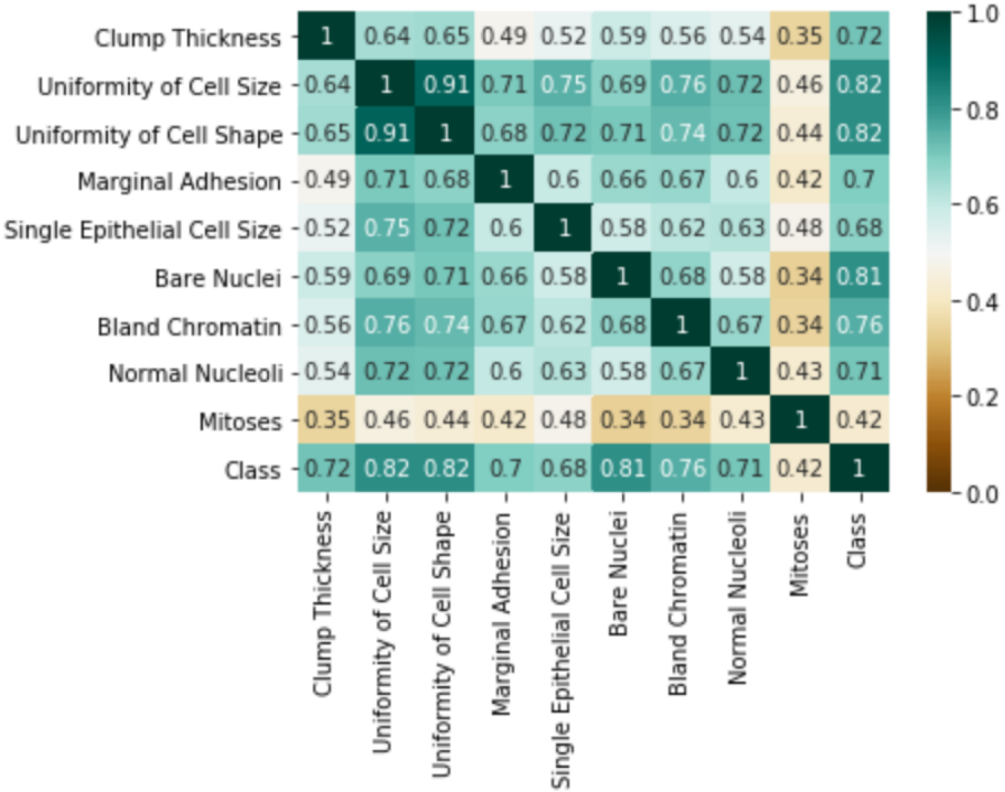
Logistic Regression

Random Forest

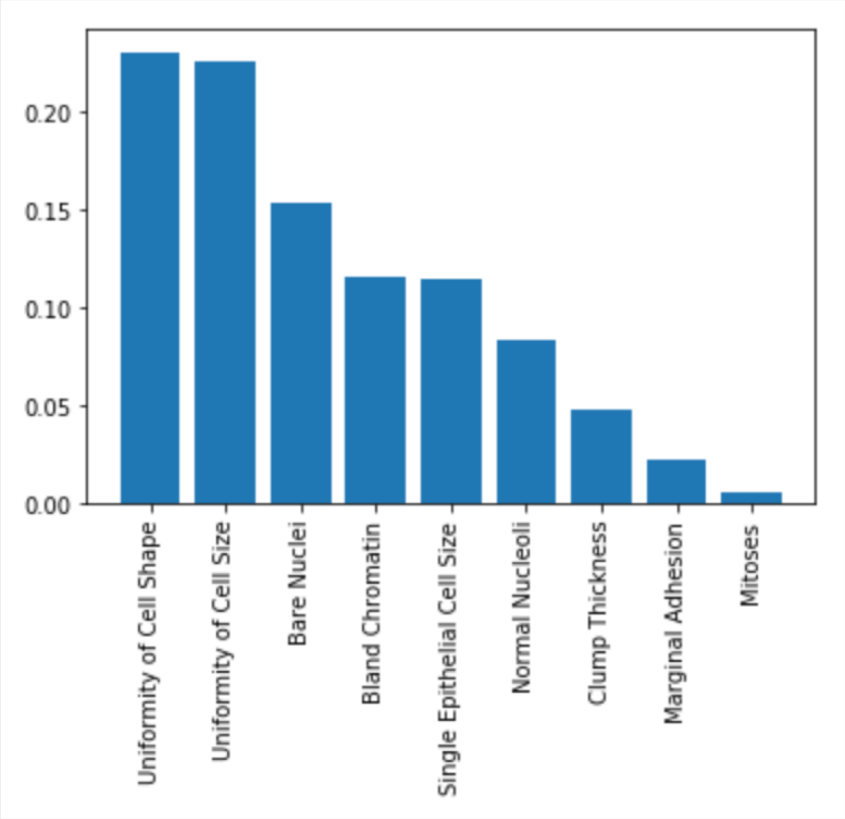


- Accuracy : 0.962
- Precision : 0.976
- Recall : 0.911

Feature Correlation

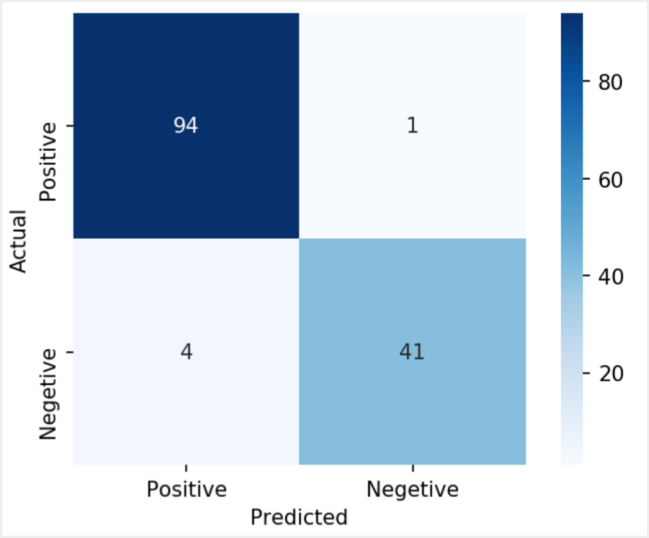


Feature Importance



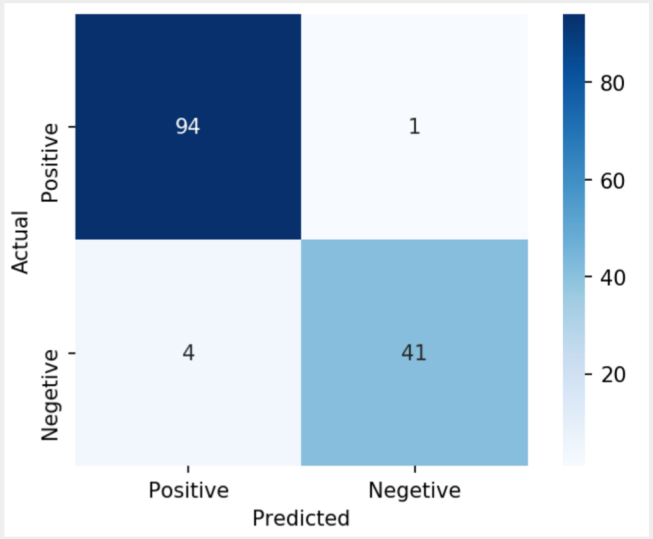
Logistic Regression

Random Forest



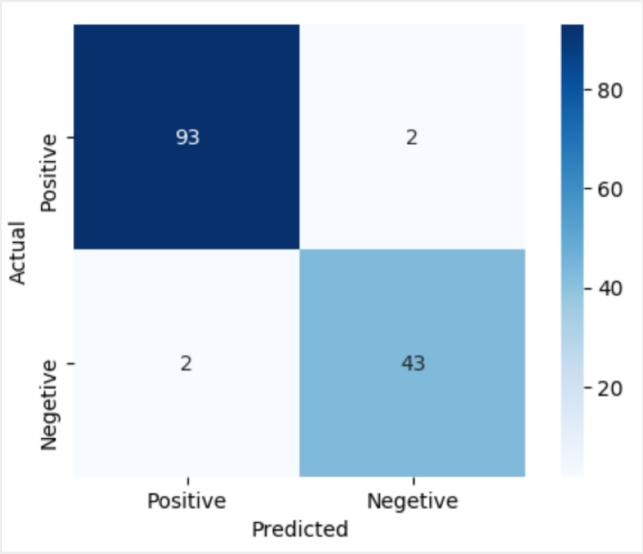
- Accuracy : 0.962
- Precision : 0.976
- F1 Score : 0.942
- Recall : 0.911

Logistic Regression



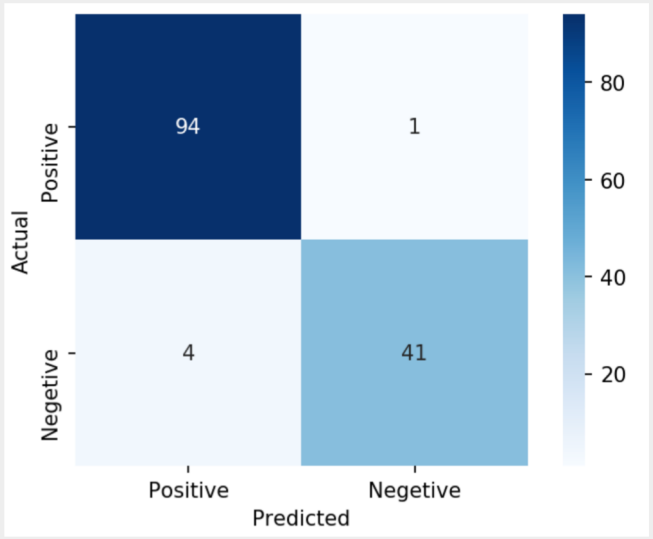
- Accuracy : 0.962
- Precision : 0.976
- F1 Score : 0.942
- Recall : 0.911

Random Forest



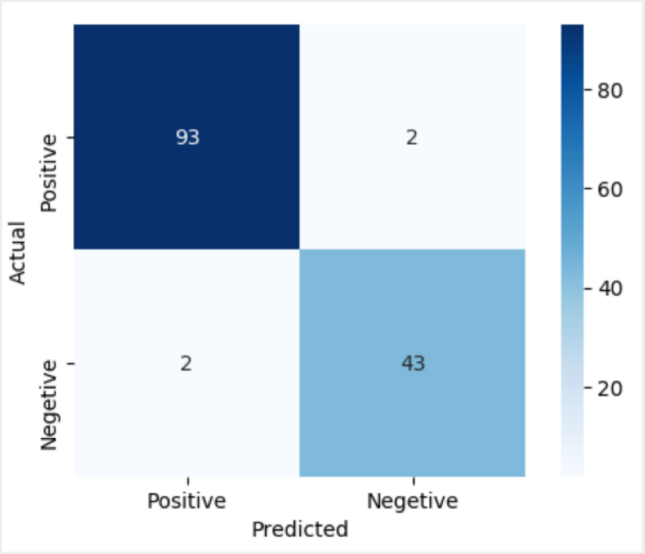
- Accuracy : 0.971
- Precision : 0.955
- F1 Score : 0.955
- Recall : 0.955

Logistic Regression



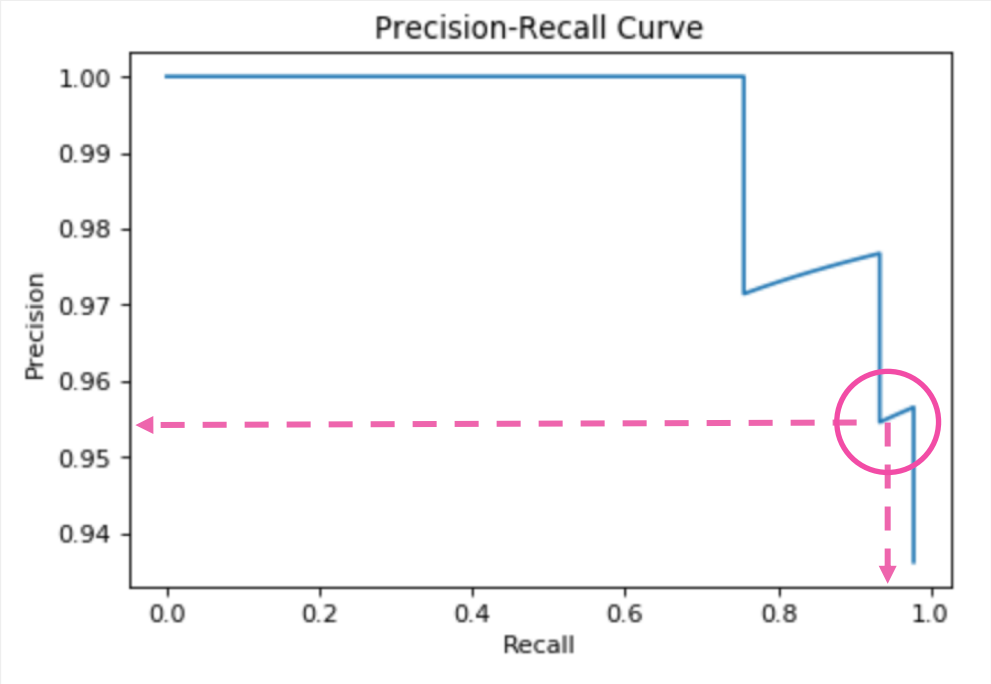
- Accuracy : 0.962
- Precision : 0.976
- F1 Score : 0.942
- Recall : 0.911

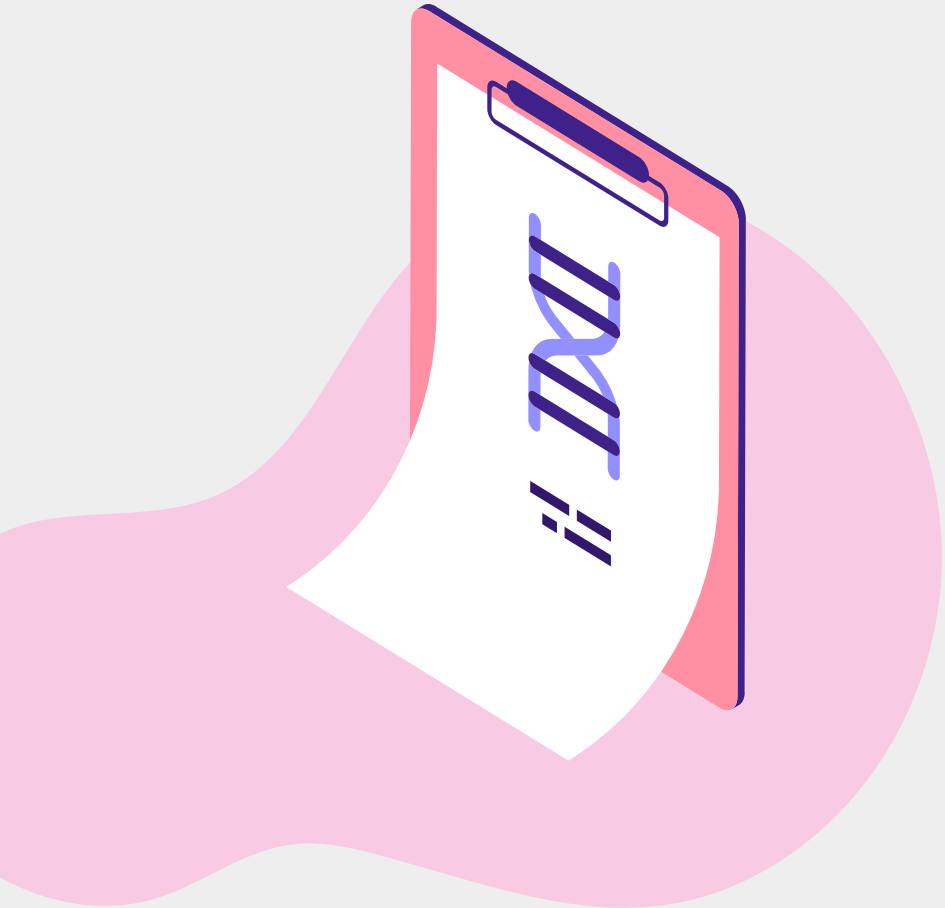
Random Forest



- Accuracy : 0.971
- Precision : 0.955
- F1 Score : 0.955
- Recall : 0.955

Precision –Recall Sweet Spot





Future Work

- SVM – Feature Transformation
- Prediction of Breast Cancer from Demographic Info
- Survival Rate Prediction based on Lifestyle
- Prognosis Prediction based on Cell Morphology

Thank You

Appendix

Nuclear Feature Extraction For Breast Tumor Diagnosis *

W. Nick Street † William H. Wolberg ‡ O. L. Mangasarian §

December 28, 1992

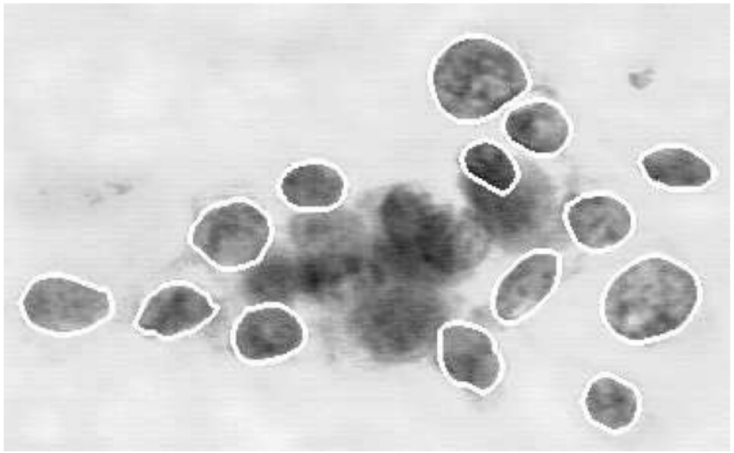


Figure 1: A magnified image of a malignant breast fine needle aspirate. Visible cell nuclei are outlined by a curve-fitting program. The **Xcyt** system also computes various features for each nucleus and accurately diagnoses the sample. Interactive diagnosis takes about 5 minutes.

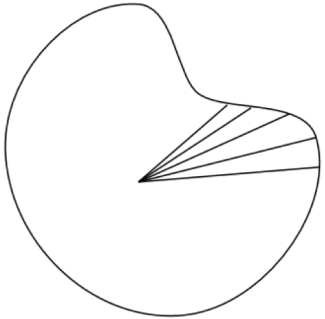
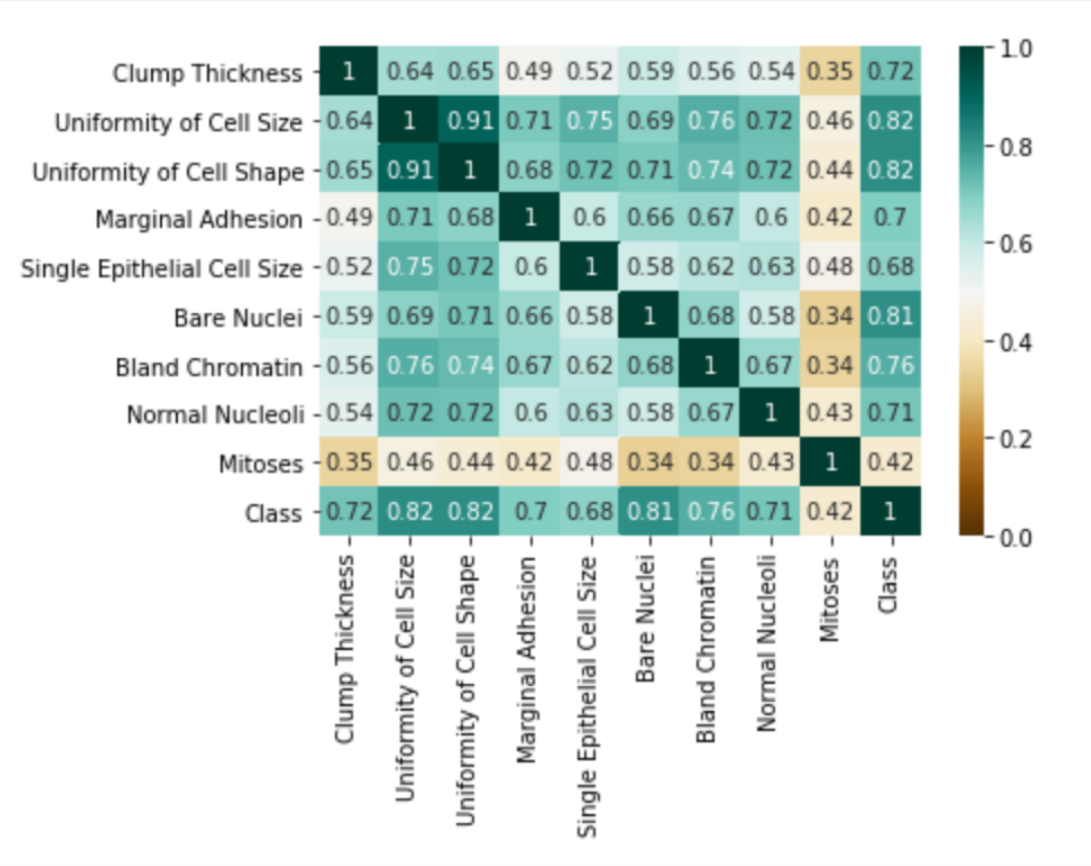


Figure 3: Radial Lines Used for Smoothness Computation

Feature Correlation



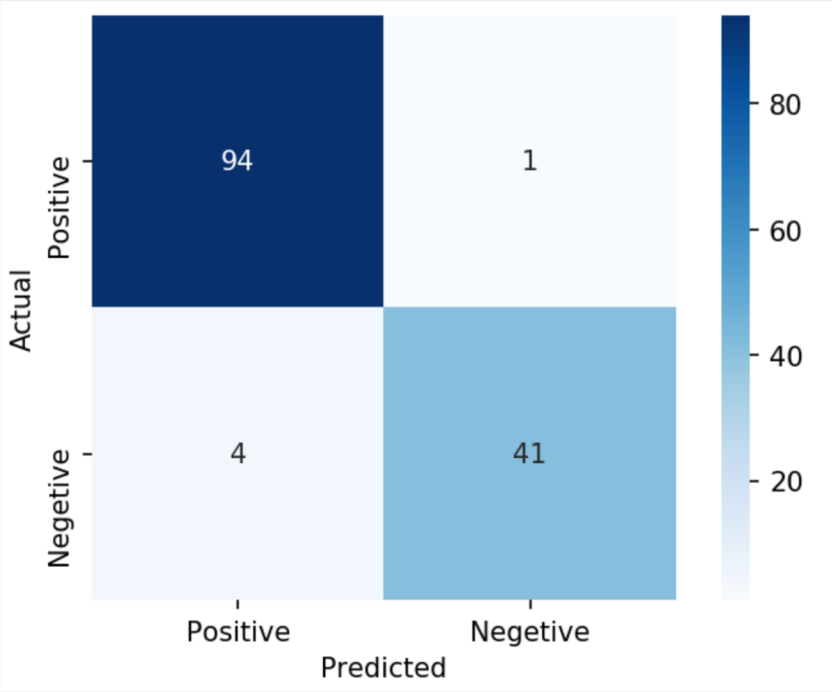
Logistic Regression – Individual Features

	Features	Accuracy Score	Precision Scores
0	Clump Thickness	0.878571	1.000000
1	Uniformity of Cell Size	0.928571	0.948718
2	Uniformity of Cell Shape	0.907143	0.900000
3	Marginal Adhesion	0.900000	0.969697
4	Single Epithelial Cell Size	0.842857	0.828571
5	Bare Nuclei	0.900000	0.942857
6	Bland Chromatin	0.892857	0.894737
7	Normal Nucleoli	0.885714	0.871795
8	Mitoses	0.807143	0.875000

Confusion Matrix

```
[[95  0]
 [17 28]] Clump Thickness
[[93  2]
 [ 8 37]] Uniformity of Cell Size
[[91  4]
 [ 9 36]] Uniformity of Cell Shape
[[94  1]
 [13 32]] Marginal Adhesion
[[89  6]
 [16 29]] Single Epithelial Cell Size
[[93  2]
 [12 33]] Bare Nuclei
[[91  4]
 [11 34]] Bland Chromatin
[[90  5]
 [11 34]] Normal Nucleoli
[[92  3]
 [24 21]] Mitoses
```

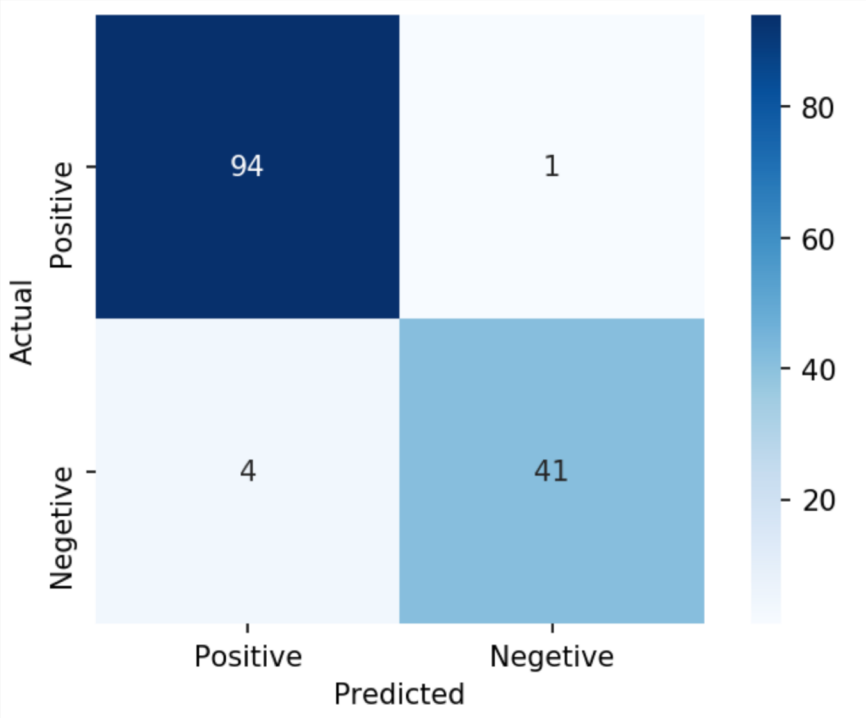
Logistic Regression – All Features



- Accuracy : 0.957
- Precision : 0.975
- Recall : 0.888
- F1 Score : 0.93

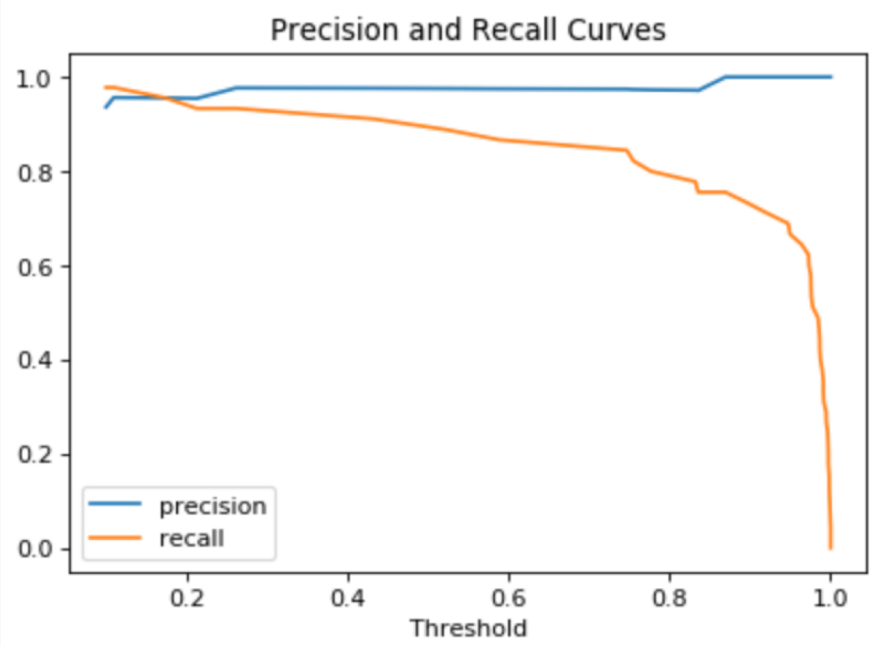
Logistic Regression – Hyperparameter Tuning

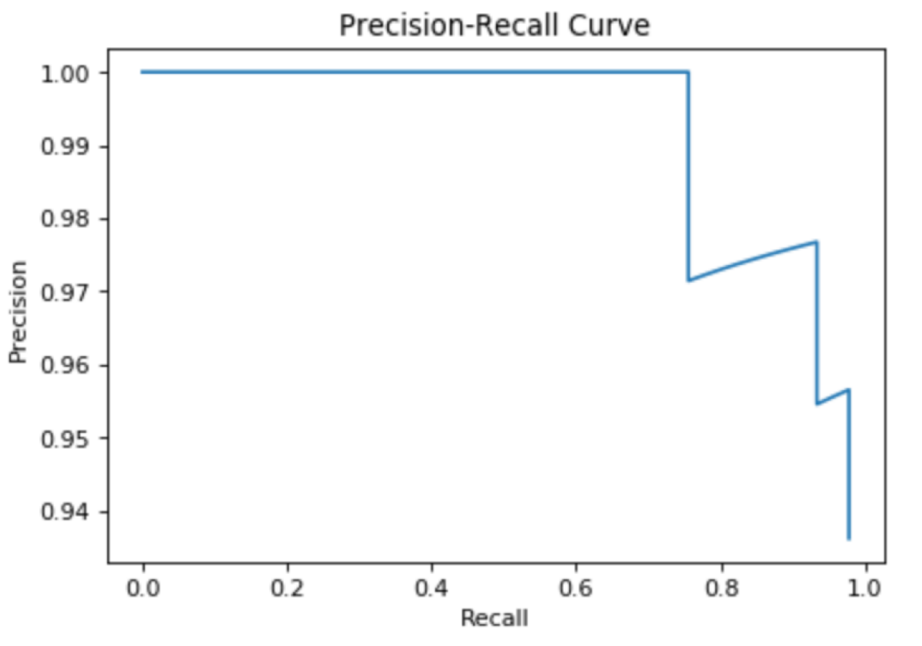
‘C’ = 0.1



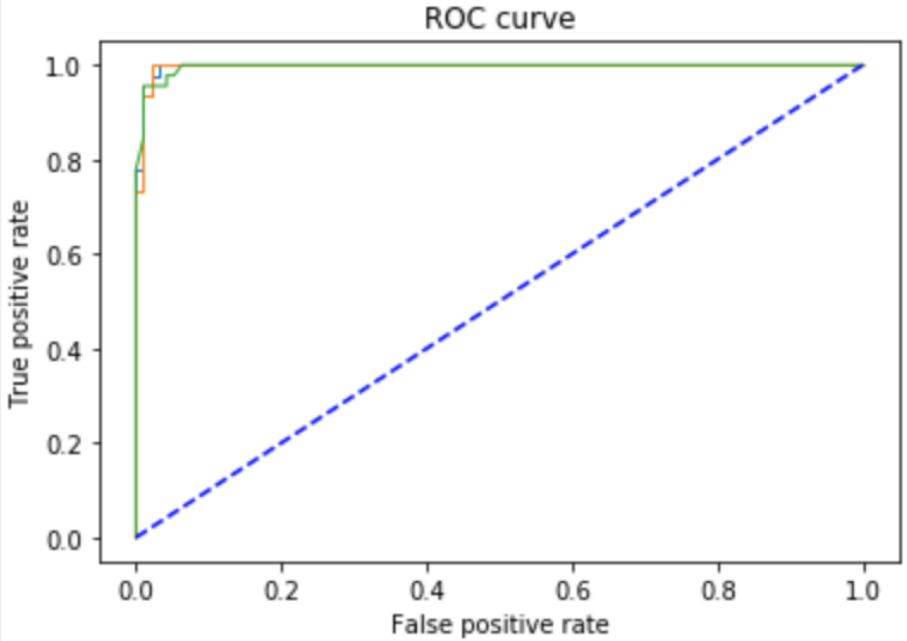
- Accuracy : 0.962
- Precision : 0.976
- Recall : 0.911
- F1 Score : 0.942

Precision and Recall – Log Reg





ROC AUC Curve




```
ROC AUC score (lm_all)=  0.9967251461988305
ROC AUC score (lm_fe1)=  0.9964912280701754
ROC AUC score (rf_fe1)=  0.9962573099415205
```