



Imarticus



alva.healthtech@gmail.com



123-456-7890

ALVA

LATIN FOR "LIFE"

EMPOWERING BREAST CANCER PREDICTION

Through cell Nuclei detection and classification



Presented By :
Ishika Bhatia

MOTIVATION & PROBLEM STATEMENT

EARLY DETECTION OF BREAST CANCER WITH MACHINE LEARNING

- Breast Cancer is the most commonly diagnosed cancer among women globally.
- Early and accurate diagnosis significantly improves survival rates
- However, interpreting cytological data (eg. cell nuclei measurement) can be time-consuming and error-prone
- Goal: Build a machine learning-powered diagnostic tool that predicts whether a breast mass is benign or malignant, based on cell nucleus features.



DATA & PREPROCESSING



- Source: Wisconsin Breast Cancer Dataset
- Total Samples: 569
- Features : 30 numeric predictors + 1 target
- Target Classes: Malignant (M) = 1,
Benign (B) = 0
- Features grouped into:
 - Mean values (eg - radius_mean)
 - Standard Error (eg - perimeter_se)
 - Worst values (eg - smoothness_worst)
- Preprocessing Steps:
 - Dropped irrelevant columns
 - id, unnamed index
 - Label encoding for diagnosis column
 - StandardScaler applied for model compatibility



MODEL ARCHITECTURE & TRAINING

LOGISTIC REGRESSION MODEL



- Model: Logistic Regression
 - Chosen for interpretability and performance on the classification
- Scaled Features using StandardScaler
- Data split: 80% train, 20% test
- Model serialized with Pickle for reuse in app/web
 - Tried different models
- Performance:
 - Accuracy: ~99% on test set
 - Classification Report:
 - Precision, Recall
 - F1-score > 0.98 for both classes

Model Comparison:
Logistic Regression: 0.9807 (+/- 0.0065)
SVM: 0.9736 (+/- 0.0147)
Random Forest: 0.9596 (+/- 0.0189)
KNN: 0.9649 (+/- 0.0096)
XGBoost: 0.9701 (+/- 0.0119)

Tuning Logistic Regression...

Tuned Logistic Regression Report:

	precision	recall	f1-score	support
0	0.99	1.00	0.99	71
1	1.00	0.98	0.99	43
accuracy			0.99	114
macro avg	0.99	0.99	0.99	114
weighted avg	0.99	0.99	0.99	114

Accuracy: 0.9912280701754386



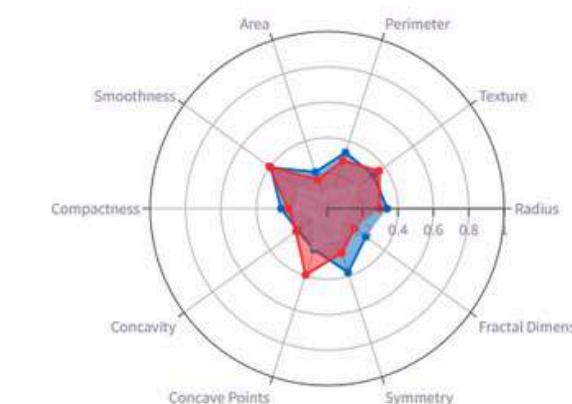
STREAMLIT APP OVERVIEW



Breast Cancer Predictor

The Breast Cancer Diagnosis app is a machine learning-powered tool designed to support medical professionals in diagnosing breast cancer. By analyzing a set of tumor measurements, the app predicts whether a breast mass is likely to be benign or malignant.

Please Connect this app to your cytology lab to help diagnose breast cancer from your tissue sample. This app predicts using a machine learning model whether a breast mass is benign or malignant based on the measurements it receives from your cytology lab. You can also update the measurements by hand(manually) using the sliders in the sidebar



Cell Nuclei Prediction

The model predicts that the breast mass(Cell Nuclei) is:

Benign

The probability of the Cell Nuclei being Benign is: 0.5841898244385364

The probability of the Cell Nuclei being Malignant is: 0.4158101755614636

This app can assist medical professionals in making informed decisions or a diagnosis, but should not be used as a substitute for professional medical advice/diagnosis. Always consult a qualified healthcare provider for any medical concerns or questions.

MODELS EVALUATED

- Logistic Regression – Ideal for binary classification, interpretable, and fast.
- Support Vector Machine (SVM) – Effective in high-dimensional spaces.
- Random Forest – Ensemble of decision trees, good at handling complex patterns.
- K-Nearest Neighbors (KNN) – Predicts based on closest neighbors, simple but effective.
- XGBoost – Advanced boosting method, learns from past errors for high accuracy.





CROSS-VALIDATION RESULTS

- Logistic Regression: 0.9807
- SVM: 0.9736
- Random Forest: 0.9596
- KNN: 0.9649
- XGBoost: 0.9701



Model Tuning:

- Used GridSearchCV to optimize Logistic Regression.

Post-Tuning Performance (on test data):

- Accuracy: 99.12%
- Precision, Recall, F1-score: All ≈ 0.99 for both classes (malignant & benign)



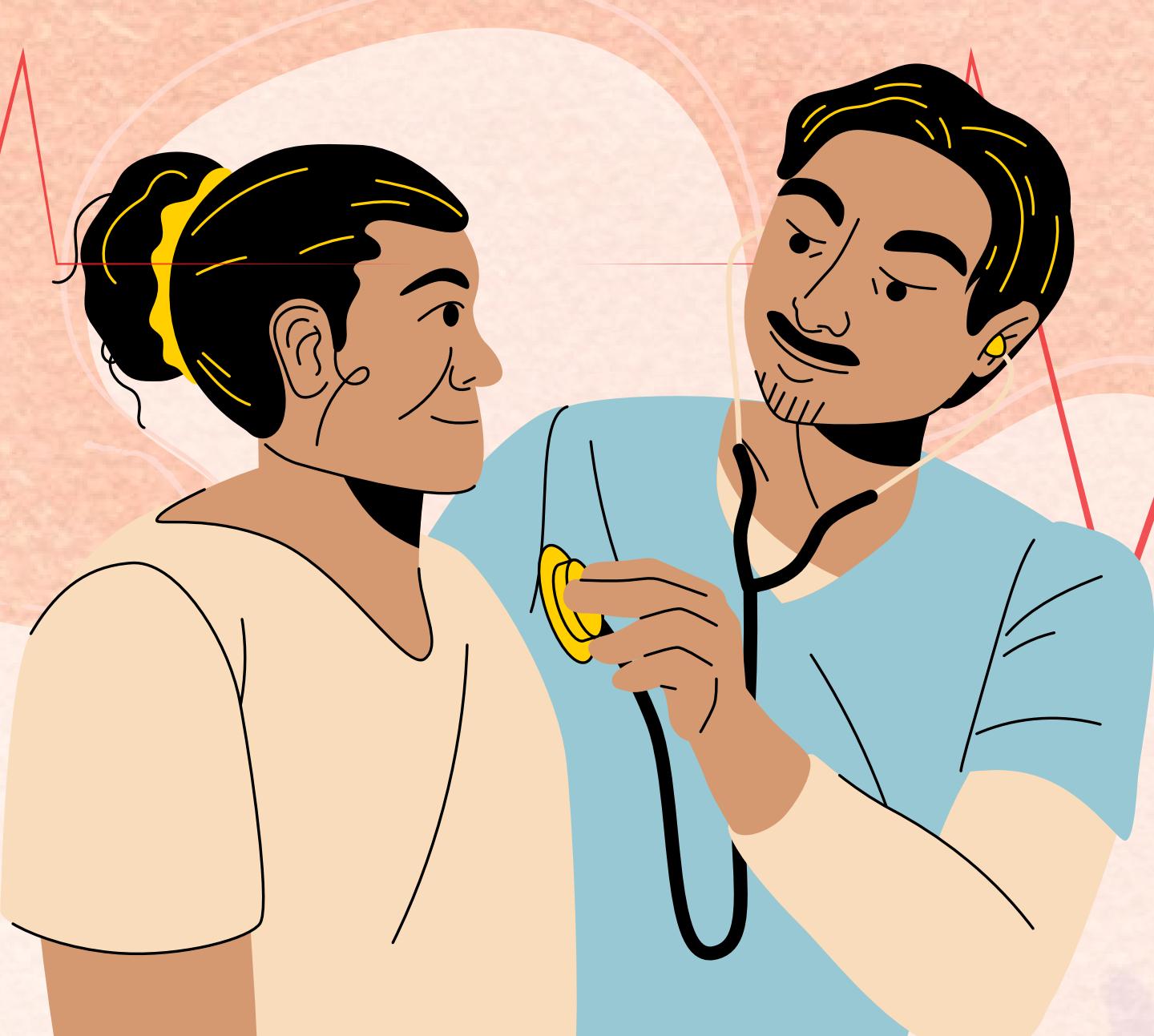
FINAL CONCLUSION

- Several models were tested, but Logistic Regression gave the best combination of performance, simplicity, and interpretability.
- After tuning, it achieved ~99% accuracy, which is highly acceptable in the healthcare domain.
- Almost perfect balance between catching all malignant cases (recall) and avoiding false positives (precision).





THANK YOU Q AND A SESSION



Presented By :

Ishika Bhatia