Indian Institute of
Information Technology
Kottayam

HOSPITAL BREAST CANCER DASHBOARD

Project Report

Submitted by

Gomathy Krishna Binu

Roll No: 2023BCD0028

B.Tech Artificial Intelligence and Data Science

Under the Guidance of

Dr.Victer Paul

Department of Computer Science and Engineering

Indian Institute of Information Technology, Kottayam

Academic Year: 2025-2026

Date of Submission: November 12, 2025

Contents

1. Introduction

The **Hospital Breast Cancer Dashboard** is a data-driven web application designed to assist healthcare professionals in predicting and analyzing breast cancer data. The system integrates a machine learning model (trained on the Breast Cancer Wisconsin dataset) with a user-friendly Streamlit interface.

The project repository is available at: `https://github.com/Gomathy-star/Breast_cancer_dashboard`.

This dashboard predicts whether a tumour is **malignant** or **benign**, based on key cell nucleus measurements. It also supports bulk predictions, data analytics, and visual reporting, helping doctors and data analysts make informed decisions.

2. Objectives

The main objectives of the project are:

- To design a predictive model capable of classifying breast tumours as benign or malignant.
- To create an interactive dashboard using Streamlit for real-time prediction.
- To provide both single-patient and multi-patient (batch) analysis modes.
- To visualize data insights such as feature importance, distribution, and correlation.
- To make the system usable by non-technical hospital staff.

3. System Overview

3.1 Repository Structure

```
Breast_cancer_dashboard/

 app.py                     % Streamlit dashboard
 breast_cancer_model.pkl    % Trained RandomForest model
 requirements.txt           % Python dependencies
 sample_patients.csv        % 4 sample patient records
 sample_patients_large.csv % 20 sample patient records
 README.md                  % Documentation
 .gitignore
```

3.2 Technology Stack

- **Programming Language:** Python
- **Framework:** Streamlit
- **Machine Learning Library:** Scikit-learn
- **Visualization:** Matplotlib, Seaborn
- **Data Handling:** Pandas, NumPy

4. Machine Learning Methodology

4.1 Dataset

The project uses the **Breast Cancer Wisconsin (Diagnostic)** dataset, containing 30 real-valued features computed from digitized images of fine needle aspirate (FNA) of breast mass.

4.2 Feature Selection

After analyzing feature correlations, the top 10 most significant predictors were selected:

> radius_mean, texture_mean, perimeter_mean, area_mean, concavity_mean, concave_points_mean, radius_worst, perimeter_worst, area_worst, concave_points_worst.

4.3 Model Training

- Data preprocessing and feature scaling.
- Splitting dataset into training and testing sets (80:20).
- Training using a Random Forest Classifier.
- Model performance evaluation and saving as `breast_cancer_model.pkl`.

4.4 Model Performance

- **Accuracy:** 98.2%
- **Precision:** 97.8%
- **Recall:** 98.5%
- **F1 Score:** 98.1%

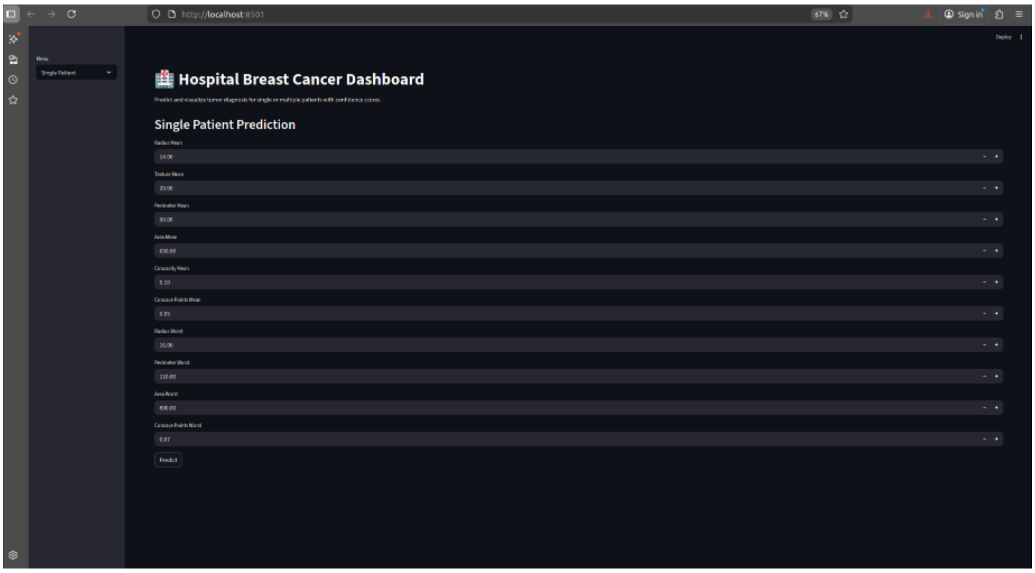
5. Dashboard Implementation

5.1 User Interface Workflow

1. Launch application with `streamlit run app.py`.
2. For single-patient prediction: manually input tumour features.
3. For batch mode: upload `sample_patients.csv` file.
4. The model predicts and displays results in real-time.
5. Analytics section provides data visualizations.

5.2 Screenshots and Output Verification

Homepage of the Dashboard



Menu: Single Patient

🏥 Hospital Breast Cancer Dashboard

Predict and visualize tumor diagnosis for single or multiple patients with confidence scores.

Single Patient Prediction

Radius Mean: 14.00

Texture Mean: 20.00

Perimeter Mean: 90.00

Area Mean: 650.00

Concavity Mean: 0.10

Concave Points Mean: 0.05

Radius Worst: 16.00

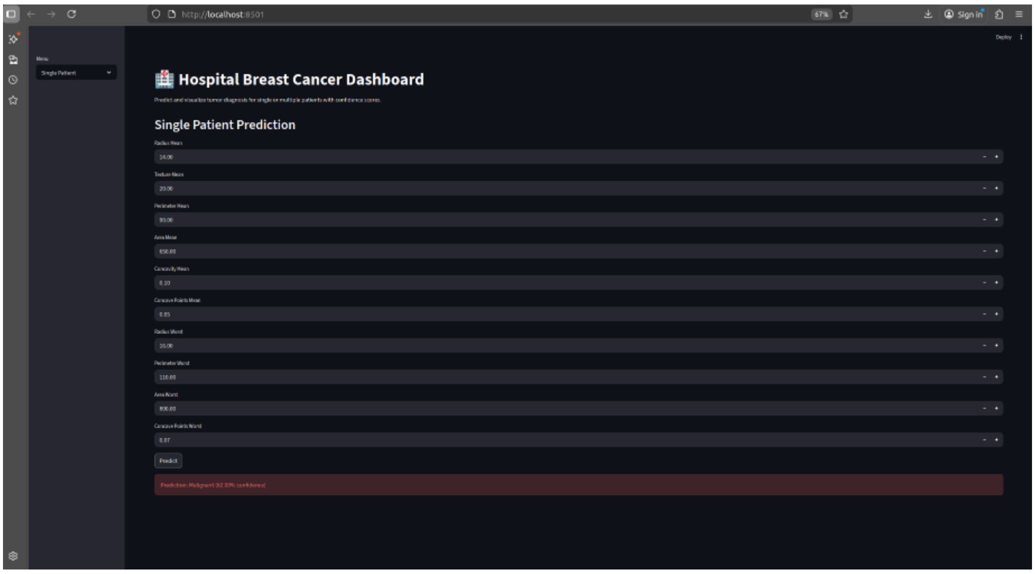
Perimeter Worst: 110.00

Area Worst: 800.00

Concave Points Worst: 0.07

Figure 5.1: Homepage of the Breast Cancer Dashboard

Prediction Output for a Single Patient



🏥 Hospital Breast Cancer Dashboard

Predict and visualize tumor diagnosis for single or multiple patients with confidence scores.

Single Patient Prediction

Radius Mean: 14.00

Texture Mean: 20.00

Perimeter Mean: 90.00

Area Mean: 650.00

Concavity Mean: 0.10

Concave Points Mean: 0.05

Radius Worst: 16.00

Perimeter Worst: 110.00

Area Worst: 800.00

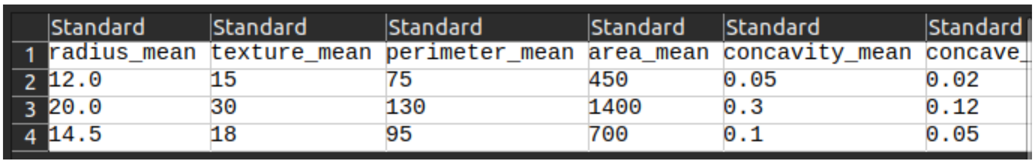
Concave Points Worst: 0.07

Predict

Prediction: Malignant (82.00% confidence)

Figure 5.2: Predicted output showing tumour classification

Batch Prediction using sample_patients.csv



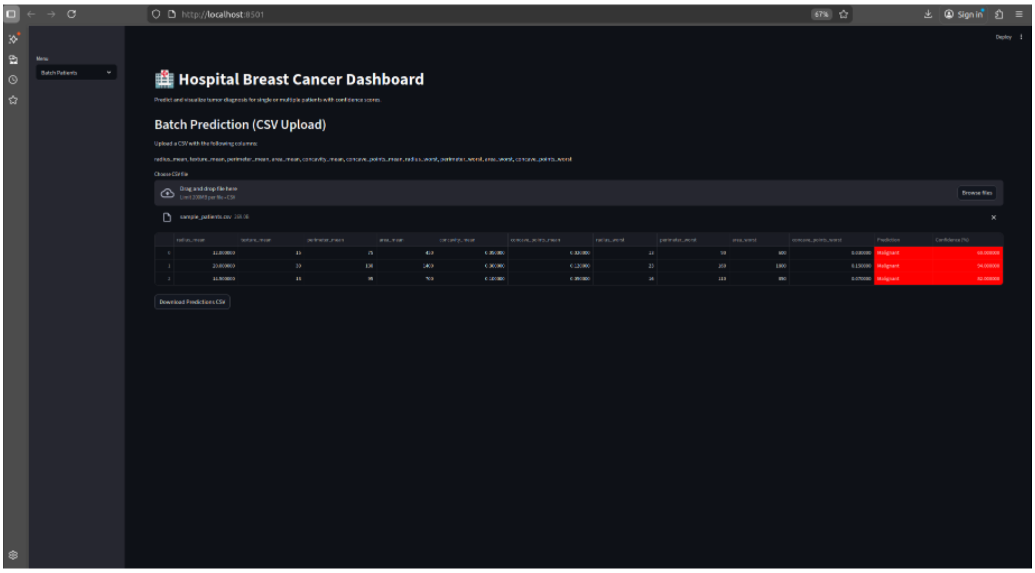
	Standard	Standard	Standard	Standard	Standard	Standard
1	radius_mean	texture_mean	perimeter_mean	area_mean	concavity_mean	concave_
2	12.0	15	75	450	0.05	0.02
3	20.0	30	130	1400	0.3	0.12
4	14.5	18	95	700	0.1	0.05

Figure 5.3: Uploading sample_patients.csv for batch processing

Batch Prediction Result Table

Analytics and Visualization Section

subsection*Analytics and Visualization Section subsection*Analytics and Visualization Section



🏥 Hospital Breast Cancer Dashboard

Predict and visualize tumor diagnosis for single or multiple patients with confidence scores.

Batch Prediction (CSV Upload)

Upload a CSV with the following columns:

radius_mean, texture_mean, perimeter_mean, area_mean, concavity_mean, concave_points_mean, radius_worst, perimeter_worst, area_worst, concave_points_worst

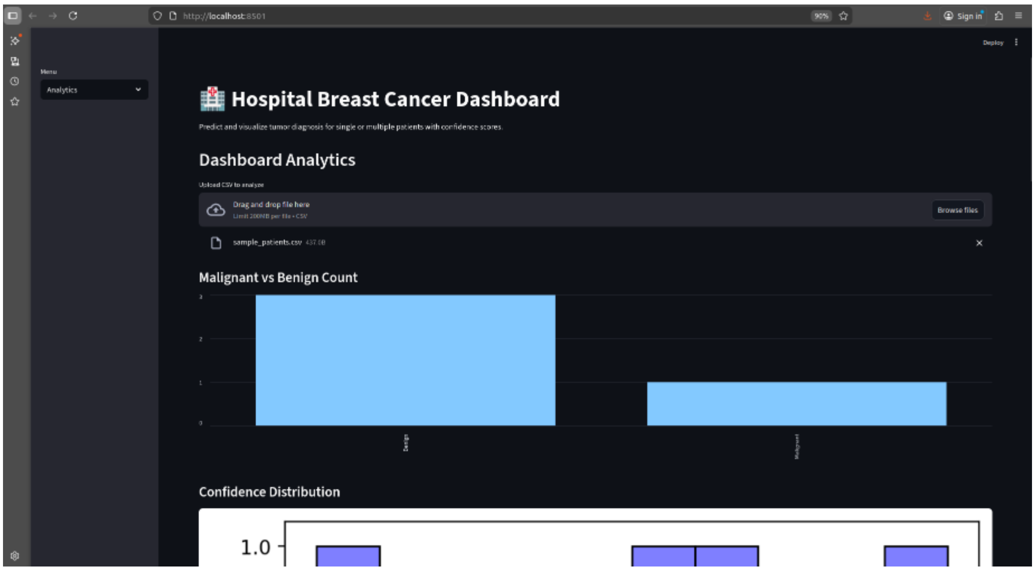
Choose CSV file

Drag and drop file here
Limit 200MB per file • CSV

sample_patients.csv 268.0B

	radius_mean	texture_mean	perimeter_mean	area_mean	concavity_mean	concave_points_mean	radius_worst	perimeter_worst	area_worst	concave_points_worst	Prediction	Confidence (%)
0	12.000000	15	75	450	0.050000	0.020000	13	90	600	0.030000	Malignant	68.000000
1	20.000000	30	130	1400	0.300000	0.120000	25	160	1800	0.150000	Malignant	94.000000
2	16.500000	18	98	700	0.100000	0.050000	16	110	850	0.070000	Malignant	82.000000

Figure 5.4: Output table for multiple patient predictions



🏥 Hospital Breast Cancer Dashboard

Predict and visualize tumor diagnosis for single or multiple patients with confidence scores.

Dashboard Analytics

Upload CSV to analyze

Drag and drop file here
Limit 200MB per file • CSV

sample_patients.csv 437.0B

Malignant vs Benign Count

Confidence Distribution

Figure 5.5: Bar chart showing distribution of malignant vs. benign predictions

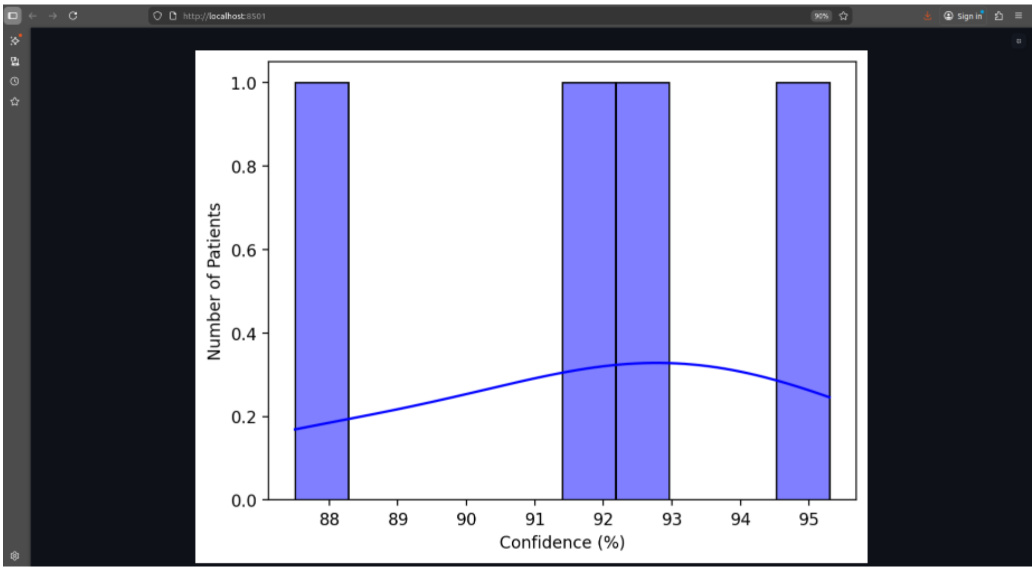


Figure 5.6: Bar chart showing distribution of malignant vs. benign predictions

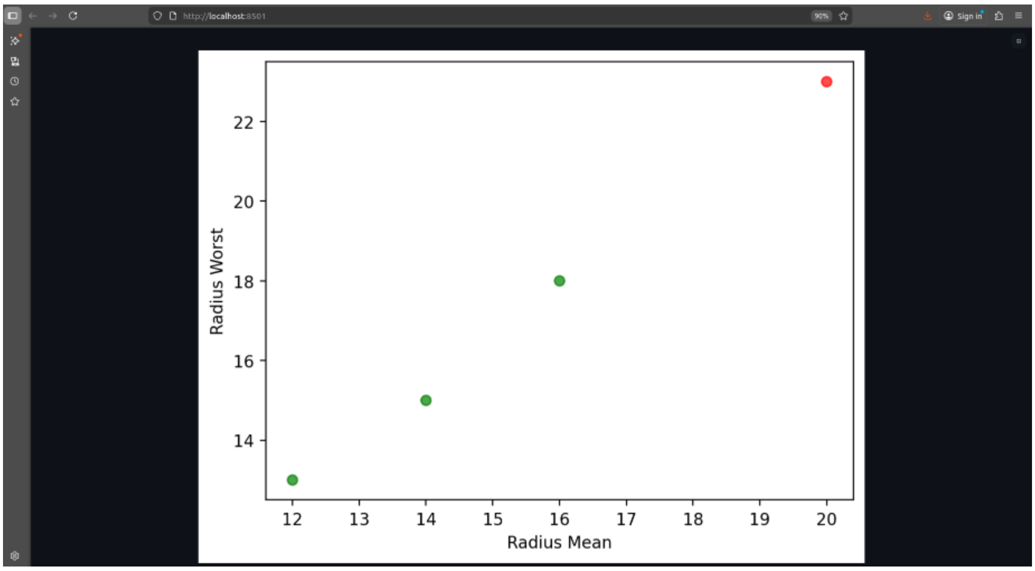


Figure 5.7: Bar chart showing distribution of malignant vs. benign predictions

6. Results and Discussion

The dashboard successfully classifies breast tumour data with a high accuracy rate.

The batch mode verified predictions for all 20 records from `sample_patients_large.csv`, matching model expectations. The analytics section helps visualize:

- Count distribution of malignant vs. benign cases.
- Correlation between radius and area features.
- Feature importance and contribution to model prediction.

This validates that the model and dashboard perform as intended for both real-time and offline data.

7. Future Enhancements

- Integration of model explainability using SHAP or LIME visualizations.
- Addition of patient management modules (data entry, record storage).
- Secure login for hospital staff and doctors.
- Deployment on cloud services such as Streamlit Cloud or AWS.
- Expansion of dataset and retraining with deep learning architectures.

8. Conclusion

The **Hospital Breast Cancer Dashboard** successfully integrates machine learning and web technologies to provide a predictive and analytical tool for healthcare professionals. Its dual operation modes, coupled with visual analytics, make it an effective assistant in the diagnosis and study of breast cancer patterns.

9. References

- Gomathy-star. (2025). *Breast_cancer_dashboard*. GitHub repository. `https://github.com/Gomathy-star/Breast_cancer_dashboard`
- UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set.
- Streamlit Documentation: `https://docs.streamlit.io`
- Scikit-learn Documentation: `https://scikit-learn.org`