■ Breast Cancer Detection using Machine Learning

[Breast Cancer Detection - AI Project] [Python 3.10] [Machine Learning: Sklearn | XGBoost] [Flask Web App]

■ About the Project
Early detection saves lives. This project uses **Machine Learning models** to detect whether a breast tumor is **benign** or **malignant** using the Breast Cancer Wisconsin (Diagnostic) dataset. We compared multiple ML algorithms and found that **XGBoost outperformed others** in terms of accuracy, precision, and computational efficiency. To make it practical, a **Flask web app** was also developed where clinicians can input patient data and instantly get predictions.

■ Features
✔■ Data preprocessing and feature scaling
✔■ Multiple ML algorithms compared (XGBoost, SVM, Random Forest, Logistic Regression, MLP)
✔■ Interactive **web application** for real-time predictions
✔■ Visualizations of dataset, feature importance, and model performance
✔■ Well-documented & easy to extend

■ Results

Model	Accuracy	Precision	Recall	F1-score	AUC
XGBoost	97.3%	94.7%	93.2%	93.9%	0.98
Random Forest	96.1%	93.0%	91.5%	92.2%	0.97
SVC	95.8%	92.5%	90.8%	91.6%	0.96
Logistic Regression	94.5%	89.6%	88.1%	88.8%	0.93
MLP Classifier	95.1%	91.4%	89.7%	90.5%	0.95

■■ Demo Screenshots
■ Malignant Prediction Example

■ Benign Prediction Example

■■ Tech Stack
- Python ■
- Scikit-learn & XGBoost
- Flask (for deployment)
- Pandas, NumPy (data handling)
- Matplotlib / Plotly (visualizations)

■ Project Structure
■ Breast-Cancer-Detection-ML
■ ■ README.md
■ ■ requirements.txt
■ ■ app.py # Flask web app
■ ■ model_training.ipynb
■ ■ breast_cancer_data.csv
■ ■ static/ # CSS, JS, images

■ ■ templates/ # HTML files

■■ Installation & Usage
1. Clone the repo:
```
git clone https://github.com/your-username/Breast-Cancer-Detection-ML.git
cd Breast-Cancer-Detection-ML
```

2. Install dependencies:
```
pip install -r requirements.txt
```

3. Run the app:
```
python app.py
```

4. Open in browser:
```
http://127.0.0.1:5000/
```

■ Future Improvements
- Add explainable AI (XAI) for model transparency
- Integrate with larger and diverse datasets
- Real-time cloud deployment
- Enhanced UI for clinicians

■ Contributing
Pull requests are welcome! If you'd like to add new models, improve deployment, or enhance the UI, feel free to fork this repo.

■ Acknowledgements
- Dataset: Breast Cancer Wisconsin (Diagnostic)
- Libraries: Scikit-learn, XGBoost, Flask, Pandas, NumPy
- Research & guidance from MSc AI program

■■■ Author
Jash Pal
■ MSc in Artificial Intelligence