Predicting Breast Cancer diagnosis

By Rolddy SURPRIS and Naël Yssa Iben Ahmed Robert

Content

Overview

Project Overview : From Data to Insight

01

Breast cancer is one of the most common and deadly diseases among women

02

Early detection greatly improves chances of survival.

03

Machine learning can help identify tumors faster and more accurately.

04

Our project explores how data can support doctors in early breast cancer diagnosis.

Dataset

Source

We used the Breast Cancer Wisconsin dataset, a trusted collection of medical records widely used for research and diagnostics.

Content

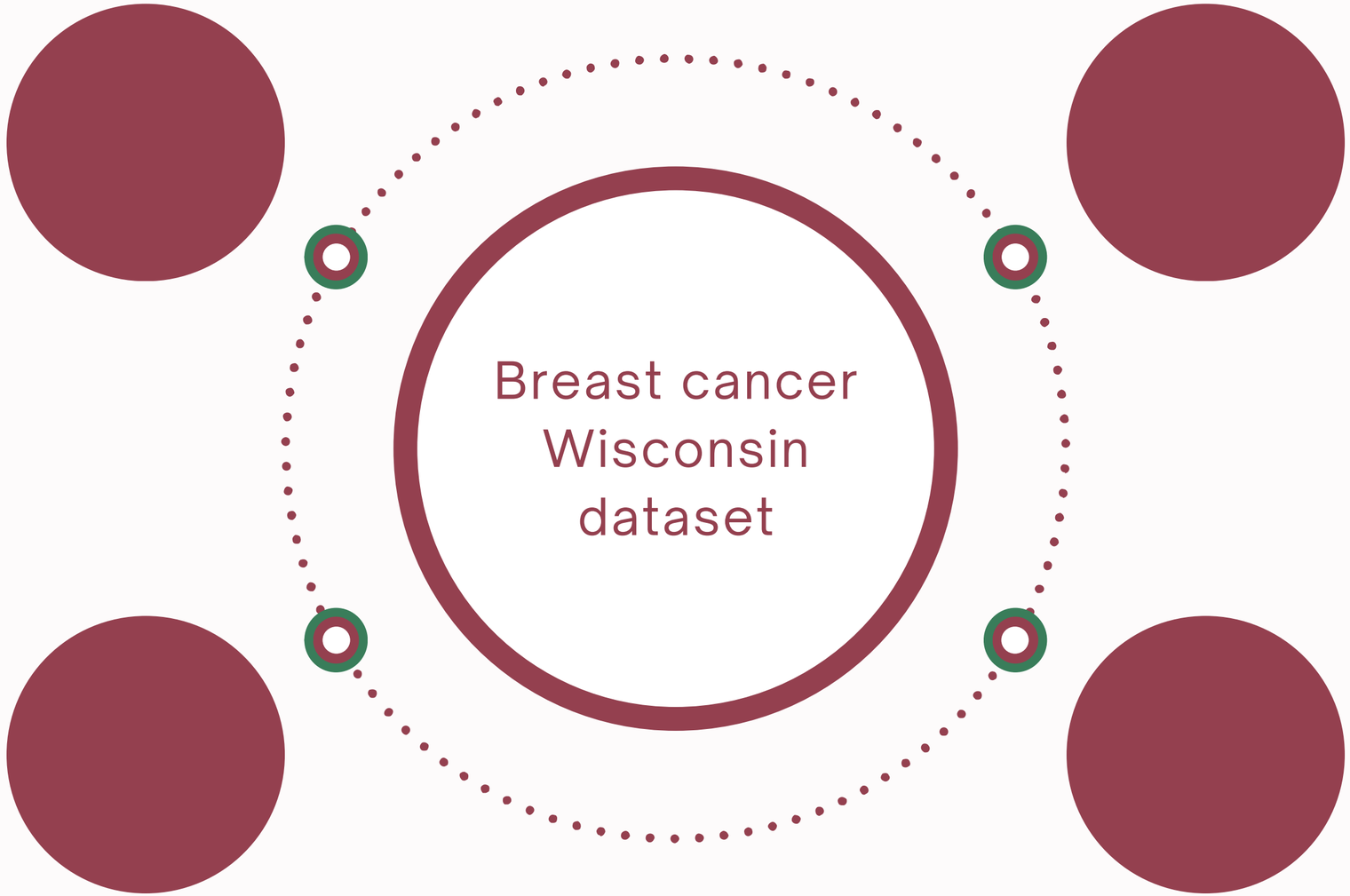
It contains measurements of 569 tumors, each described by 30 numerical features related to size, texture, and shape.

Goal

The data helps distinguish between benign (non-cancerous) and malignant (cancerous) tumors.

Preparation

We cleaned and organized the data to make it ready for analysis and model training.



Target Audience

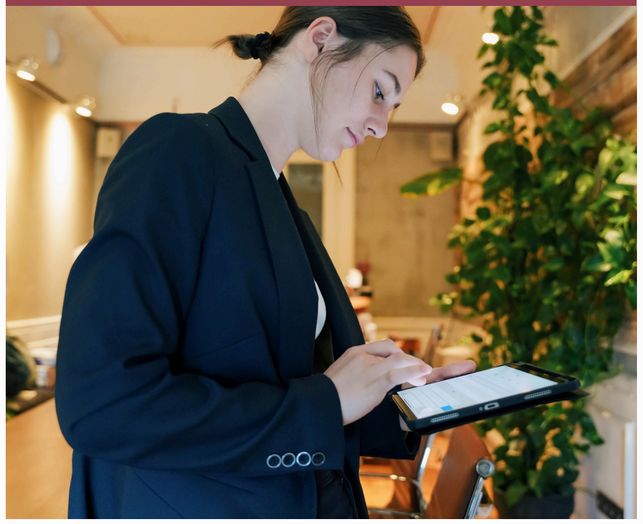
Target Audience #1

Medical Professionals: Doctors, nurses, and radiologists who want tools to support early cancer detection.

Target Audience #2

Data Science Enthusiasts: Learners and practitioners exploring how AI can assist in healthcare.

Target Audience #3



Healthcare Decision Makers: Hospitals, clinics, and organizations interested in improving patient outcomes.

Methodology

We developed a machine learning model to help detect breast cancer early. The process combines data preparation, model training, and testing to ensure reliable and understandable predictions.

4- Evaluation

Checked the model's predictions and adjusted settings to improve reliability.

3- Model Training

Tested multiple models and optimized the best one to detect cancer accurately.

2- Feature Analysis

Examined important measurements that help distinguish malignant from benign cases.

1- Data Collection and Cleaning

Gathered breast cancer data and prepared it for analysis.

RESULTS AND INSIGHT

Model Performance

Our logistic regression model reliably detected most malignant cases, showing that machine learning can support early breast cancer detection.

Key Insights

The few missed cases highlighted the unpredictable nature of cancer, not a limitation of the model itself. This reinforces the need for clinical judgment alongside AI tools.

Clinical Potential

The results demonstrate that such models can help doctors prioritize attention and make faster, more informed decisions, while remaining interpretable and trustworthy.

Limitations and challenges

Limited Data Diversity

The dataset comes from a specific population and may not represent all patient groups.

False Negatives

Some malignant cases remain undetected due to the unpredictable nature of cancer.

Feature Limitations

Certain biological signals are not captured in the available data, which can affect model accuracy.

Clinical Integration

Implementing AI in real medical practice requires careful validation, trust, and collaboration with healthcare professionals.

Impact and future work

Support Early Detection

Help doctors identify suspicious cases more quickly and reliably.

Increase Clinical Decisions

Provide interpretable insights to assist, not replace, medical judgment.

Broader Datasets

Future work will include more diverse data to improve model reliability across populations.

Innovation and Collaboration

Expand AI applications in healthcare and actively contribute to the growing data science revolution.

THANK YOU

Rolddy SURPRIS

rolddysurpris@gmail.com

Naël Yssa Iben Ahmed ROBERT

rnaelyssa@gmail.com

https://github.com/Rawldyh/Predicting-Breast-Cancer-Diagnosis.git