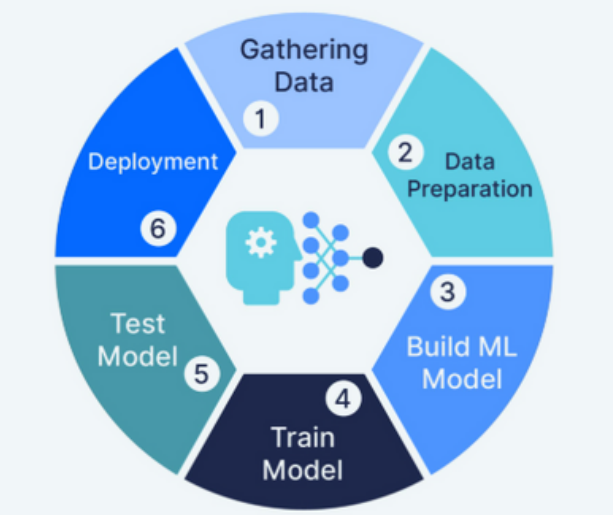
BREAST CANCER DETECTION

OBJECTIVE

To develop a web-based tool that detects breast cancer using both numerical patient data and histopathology images with high accuracy.



DATASET

Numerical Data:
Breast Cancer Wisconsin (Diagnostic) Original Dataset
➤ 699 records (typically 683 after cleaning)
➤ 10 biological attributes (rated 1–10) ➤ Class Labels: Benign (2), Malignant (4)
Image Data:
Breast Histopathology Images
➤ 277,000+ 50x50 pixel images
➤ Structured in Patient ID folders → subfolders:
• 0 = Benign images
• 1 = Malignant images

MODELS USED:

Machine Learning (Numerical Data):
• SVM: 95.2% accuracy
• Also tested: KNN, Decision Tree, Naive Bayes
Deep Learning (Image Data):
• CNN Model: 92.8% accuracy
• Trained on preprocessed histopathology images

INTEGRATION

• Combined both models into a unified Flask web application
• Built two separate routes:
1 For numerical data input and ML predictions
2 For image upload and CNN-based diagnosis
• Responsive UI with HTML/CSS
• Real-time predictions for both input types