Early Detection of Breast Cancer Using Machine Learning and Deep Learning

Objective:
The objective of this project is to develop predictive models that classify breast tumors as either malignant or benign. By applying exploratory data analysis, machine learning methods, deep learning, and neural network modeling, the goal is to evaluate which techniques deliver the most accurate and reliable predictions for a medical diagnostic context.

Dataset Description:

Source: The Breast Cancer Wisconsin Diagnostic (WDBC) dataset, available in scikit-learn via load_breast_cancer().

Structure: The dataset contains 569 samples of tumors with 30 numeric features extracted from digitized images of fine needle aspirate (FNA) of breast masses. Each instance is labeled as malignant (1) or benign (0).

Key Features: Features include measurements such as mean radius, mean texture, mean perimeter, mean area, smoothness, compactness, concavity, symmetry, and fractal dimension, with separate statistics for mean, standard error, and worst values for each property. These features are numeric and allow for predictive modeling. No missing values and dataset is balanced enough for training reliable models.

Business Problem:
Breast cancer is a leading cause of death among women globally. Misdiagnosis can result in delayed treatment or unnecessary procedures.Early detection of malignant breast tumors is critical for patient treatment and survival. Accurate classification models can assist medical professionals in identifying high-risk cases and prioritizing care.

Project summary:
Exploratory Data Analysis:

-Checked for missing values and verified dataset structure.
-Scaled features using StandardScaler for uniformity.
-Visualized feature correlations, distributions, and class balance using heatmaps, boxplots, and count plots.

Machine Learning:

-Random Forest: Captures nonlinear relationships and provides feature importance insights. Achieved ~95.6% test accuracy.
-Support Vector Machine: Used RBF kernel to model complex boundaries; achieved ~98.2% test accuracy.
-Hyperparameter tuning was performed using GridSearchCV and validation curves to optimize model performance.

Deep Learning:

-Built a feed-forward neural network with two hidden layers (32 and 16 neurons) using Keras.
-Applied EarlyStopping to prevent overfitting.

-Tested multiple learning rates; best NN achieved ~95.6% test accuracy.

Evaluation & Visualization:

-Confusion matrices, ROC curves, and feature importance plots were used to interpret model performance.

-Both ML and DL models demonstrated high predictive accuracy and reliability

Conclusion:

This project demonstrates that classical ML methods (RF, SVM) and a simple DL network can effectively classify breast tumors. SVM achieved the highest accuracy, while RF provided interpretable feature importance. The models developed can support early detection, assist clinicians, and reduce diagnostic workload