## Comparative Analysis of ML Models for Breast Cancer Classification

1. Introduction:  
 Breast cancer is one of the most common cancers affecting women worldwide. Early diagnosis significantly increases the chances of successful treatment. This project leverages machine learning algorithms to classify breast tumors as benign or malignant using the Breast Cancer Wisconsin (Diagnostic) dataset.

2. Objective:  
 To build a machine learning model that can accurately predict whether a breast tumor is benign or malignant based on features computed from a digitized image of a fine needle aspirate (FNA) of a breast mass.

## 3. Dataset:

* **Source:** UCI Machine Learning Repository
* **Name:** Breast Cancer Wisconsin (Diagnostic) Data Set
* **Features:** 30 numeric features (e.g., mean radius, texture, perimeter, area, smoothness, etc.)
* **Target:** Diagnosis (0 = Malignant, 1 = Benign)

## 4. Methodology:

## 4.1 Data Loading and Preprocessing:

* Loaded dataset using sklearn.datasets.load\_breast\_cancer()
* Converted data to pandas DataFrame for ease of manipulation
* Scaled features using StandardScaler to normalize input features

## 4.2 Model Building:

* Compared multiple models: Random Forest, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM)
* SVM with a linear kernel gave the best performance
* Split dataset: 80% training, 20% testing

## 4.3 Model Evaluation:

* Evaluated using accuracy score and classification report
* Best model: **SVM** with ~97% accuracy

## 4.4 Model Saving:

* Saved the trained SVM model and the scaler using joblib
* Files created: svm\_model.pkl and scaler.pkl

## 5. Technologies Used:

* Python
* Scikit-learn
* Pandas, NumPy

## 6. Results:

* The SVM model achieved over 97% accuracy on the test set
* The application provides intuitive feedback based on user inputs

7. Conclusion:  
This project demonstrates how machine learning can aid in the early detection of breast cancer. With a clean interface and high-accuracy prediction, this tool can assist medical professionals in making informed decisions.

## 8. Future Work:

* Integrate with real-time medical databases
* Deploy the application online using Streamlit Cloud or other platforms
* Implement additional models like Neural Networks for comparison