# Breast Cancer Classification Project

## 1. Introduction

This project aims to analyze the impact of dimensionality reduction techniques on the classification performance of breast cancer diagnosis data. We apply Principal Components Analysis (PCA), Linear Discriminant Analysis (LDA), and t-Distributed Stochastic Neighbor Embedding (t-SNE), and evaluate their impact using multiple classification algorithms.

## 2. Dataset Description

The Breast Cancer Wisconsin (Diagnostic) dataset contains 569 instances with 30 continuous features. The target variable ('Diagnosis') is binary: M (Malignant) and B (Benign). The features represent various measurements of cell nuclei in digitized images.

Dataset Source: UCI Machine Learning Repository

## 3. Methodology

The methodology involves:  
1. Preprocessing the dataset (label encoding, scaling).  
2. Applying dimensionality reduction techniques (PCA, LDA, t-SNE).  
3. Training classifiers (Logistic Regression, KNN, SVM, Random Forest, Naive Bayes).  
4. Comparing classification performance with and without dimensionality reduction.

## 4. Results

The results of dimensionality reduction and classification will be detailed with visualizations and tables.

## 5. Conclusion

Dimensionality reduction significantly impacts the classification performance. Findings are discussed in detail.

## 6. References

UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/

Scikit-learn documentation: https://scikit-learn.org/stable/