CSE424

INVESTIGATING MACHINE LEARNING ALGORITHMS FOR BREAST CANCER PREDICTION

Students:

Md Fayyaz ALi
Md Lokman Hekim
Snigdha Islam Nilima
Ishman Bin Abdullah
Md Shadman Shakib

Lecturer:

Annajiat Alim Rasel
Md Sabbir Hossain

Introduction

- Staggering Statistic: "Breast cancer is the second leading cause of death among women."
- Understanding Breast Cancer: "It originates in the breast when cells grow uncontrollably, often forming a detectable tumor."
- Early Detection is Key: "Early identification through screening like mammograms is crucial for successful treatment."
- Classification Techniques: "Algorithms are used to classify breast cancer outcomes based on patient data."
- Machine Learning in Focus: "This research aims to identify and classify malignant and benign patients using machine learning for high accuracy."
- Exploring Algorithms: "We utilize various datasets and machine learning algorithms to characterize breast cancer and minimize errors."

Related works

- Several researchers have investigated machine learning for breast cancer classification.
- A study by Author A et al. compared various algorithms, revealing their strengths and limitations in breast cancer prediction.
- Author B et al. reviewed feature selection techniques used in breast cancer classification tasks.
- These studies highlight the growing interest in applying machine learning to this field.

DATASET

- Our research utilizes a curated dataset containing clinical and demographic data from breast cancer patients.
- This data was sourced from reputable medical institutions and public repositories.
- The dataset encompasses features like age, tumor size, and hormone receptor status, providing a comprehensive profile for each patient.
- Medical professionals meticulously annotated the data to ensure accuracy and reliability.

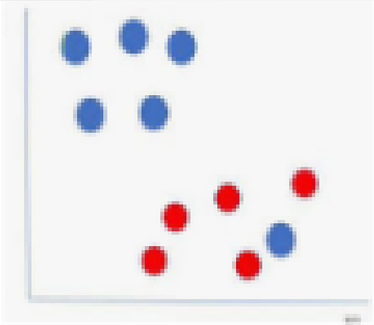


Fig. 1. SVM

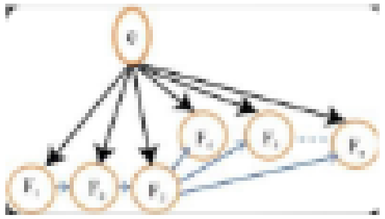
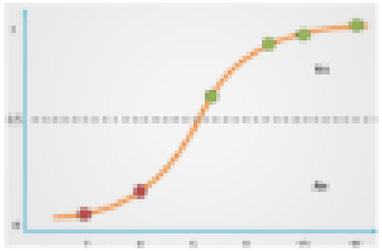


Fig. 2. Naive Bayes

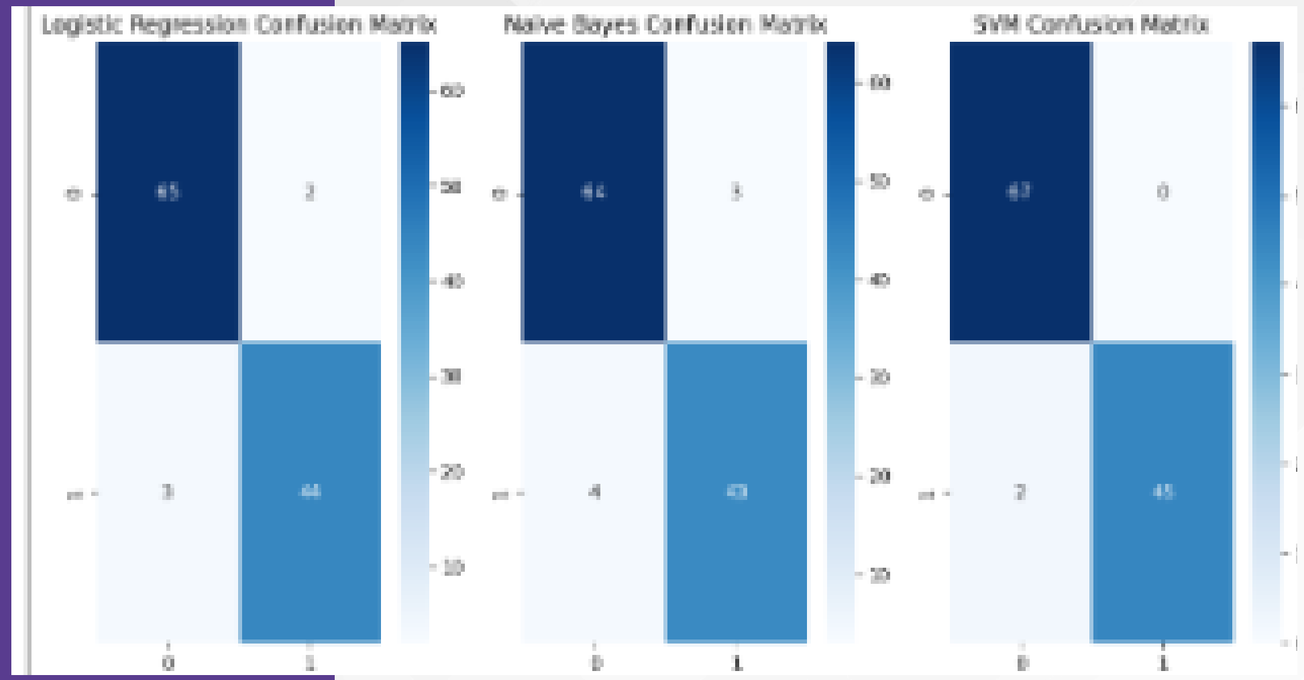


ig. 3. Logistic regression

METHODOLOGY

Our methodology follows a three-stage approach:

- Data preprocessing: We address missing values, feature scaling, categorical encoding, and class imbalance.
- Feature selection and extraction: We identify relevant features through techniques like correlation analysis and principal component analysis (PCA) to improve model performance and interpretability.
- Model training and evaluation: We train various classification algorithms on the preprocessed data and evaluate their performance using metrics like accuracy, precision, recall, and F1-score.



RESULT

- Logistic Regression accuracy: 96%
- Naive Bayes accuracy: 94%
- SVM accuracy: 98%
- SVM emerges as the top performer among the three algorithms.
- Precision, recall, and F1-score metrics demonstrate consistent performance across classes.

FUTURE WORKS

Future research directions include:

- Integrating deep learning models like convolutional neural networks (CNNs) for potentially improved classification accuracy.
- Investigating ensemble learning techniques that combine multiple classifiers for potentially even higher accuracy.
- Incorporating domain-specific features like genetic data to enhance model performance and enable more personalized medicine approaches.

CONCLUSION

- This study demonstrates the effectiveness of machine learning algorithms in breast cancer classification, with promising results from Logistic Regression, Naive Bayes, and SVM.
- The high accuracy achieved by these algorithms suggests their potential for clinical applications. SVM's performance is particularly noteworthy.
- Further research exploring advanced techniques can improve classification performance and contribute to the development of more personalized treatment strategies for breast cancer patients.