



Dayananda Sagar College of Engineering

Department of Computer Science & Engineering

Analysis of Breast Cancer Detection using Ensemble Method

Shalini Singh (1DS19CS146) Kota V Vishnu (1DS19CS723) Reena Jasmine Edwin (1DS19CS738) S Sai Brinda (1DS19CS741)

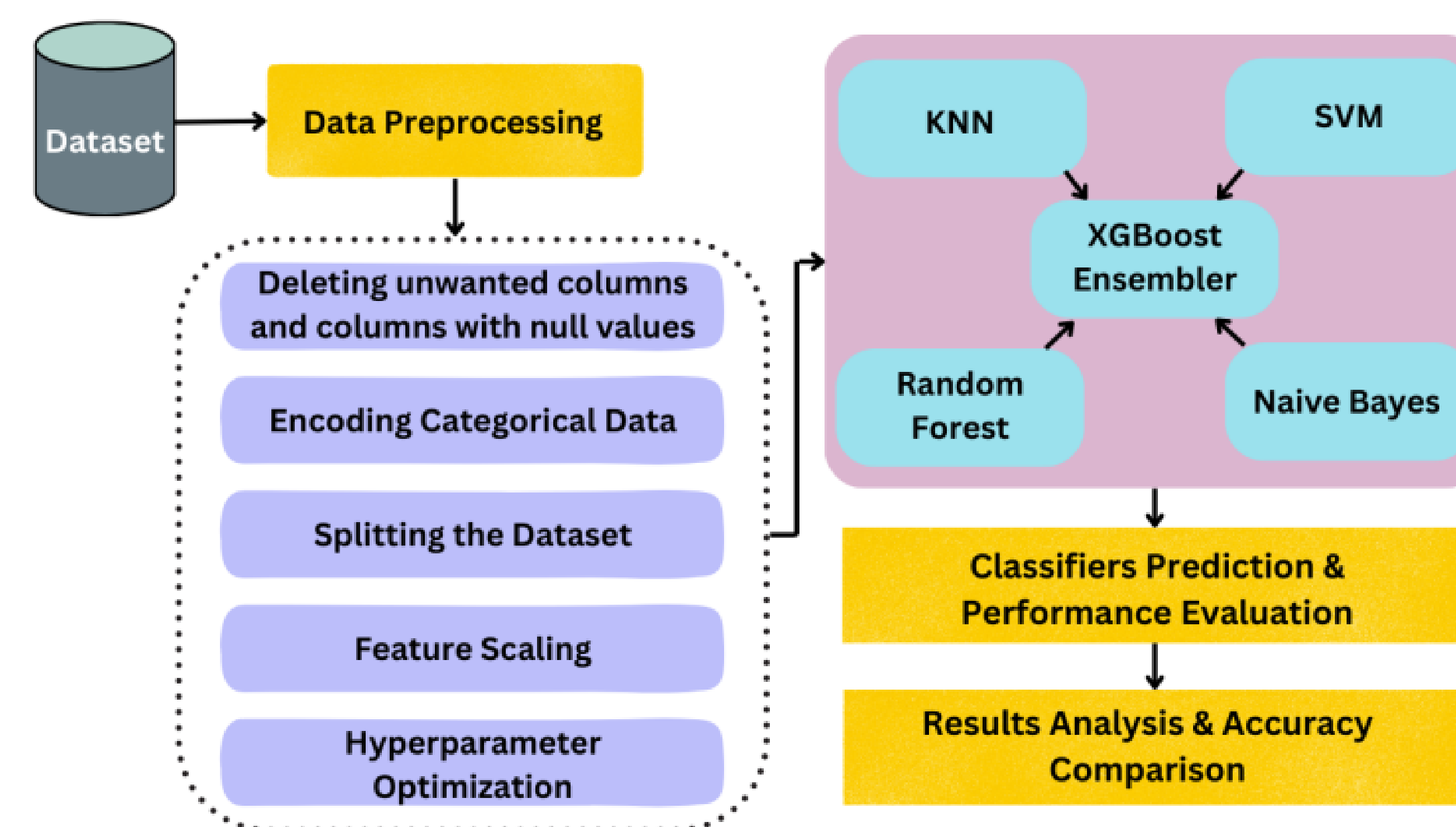
Guide: Mrs. Annapoorna B R

Asst Professor, CSE , DSCE

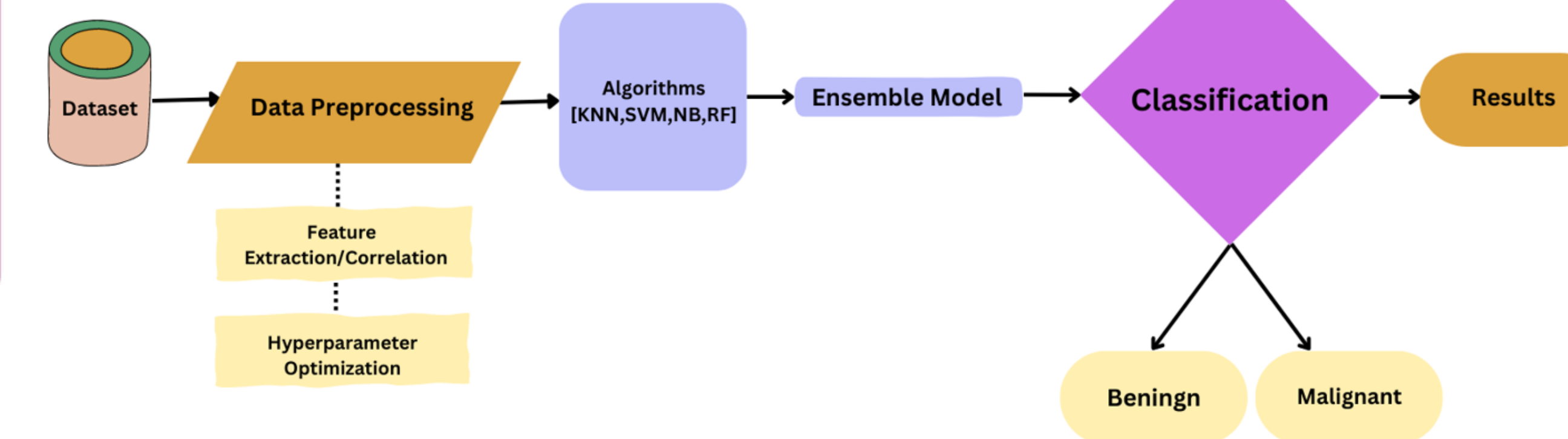
Abstract

Machine learning and artificial intelligence are transforming breast cancer diagnosis by developing reliable diagnostic tools that accurately identify patterns indicative of breast cancer. These tools train on large patient datasets to create an ensemble model that combines machine learning algorithms, providing a significant advancement in accurate breast cancer detection.

Architectural Design



Data Flow Diagram



Objectives

1. Develop high-accuracy ensemble model for breast cancer detection.
2. Compare model performance with traditional ML models.
3. Identify most important features for accurate detection.
4. Optimize model through hyperparameter tuning and improving performance.
5. Validate model on large dataset for reliability and clinical potential.

Hypothesis

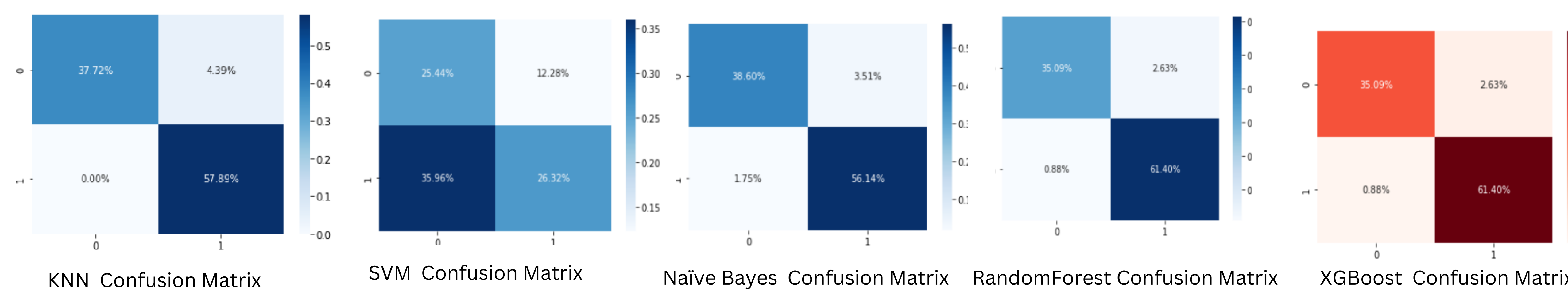
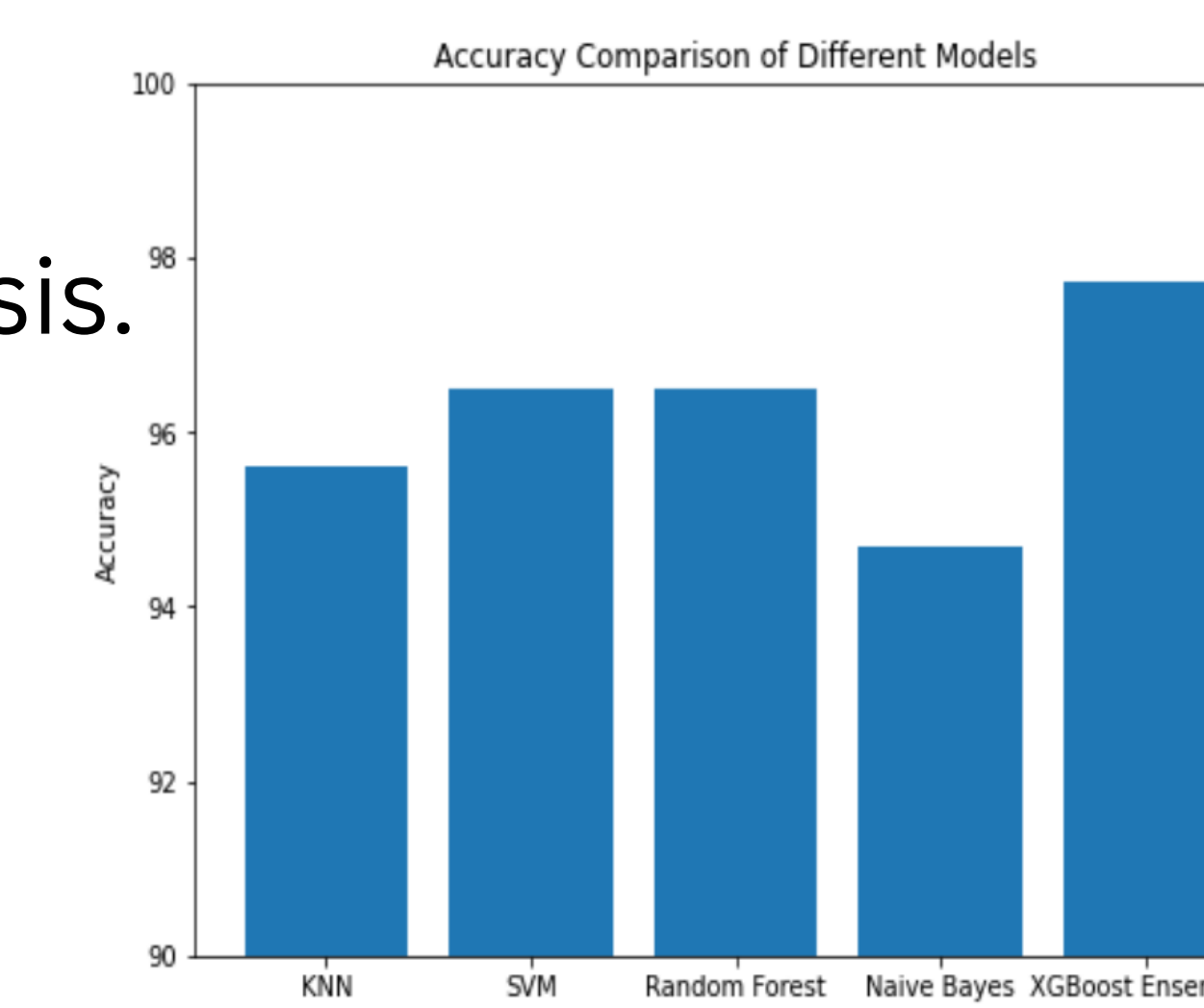
- There are two different hypothesis for this analysis, the null hypothesis which states that there is no co-relation between statistics and observed parameters like genetics, health and the likelihood of getting diagnosed
- The alternative hypothesis says that there is a direct line between inherited diagnosis and specific genes to analyze cancer

Results

- Breast cancer detection using machine learning models - SVM, KNN, Naïve Bayes, Random Forest and XGBoost Ensemble Model.
- Dataset: 569 observations, divided into 80% training set & 20% testing set.
- Performance assessment: Accuracy, Precision, Recall, Confusion Matrix analysis.
- XGBoost Ensemble model outperformed Conventional ML Algorithms, achieving 97.7% Accuracy.

	KNN	SVM	Naive Bayes	Random Forest	XGBoost Ensemble
Accuracy(%)	95.61	96.49	94.7	96.49	97.71
Precision(%)	1.0	1.0	94.11	95.89	95.89
Recall(%)	89.58	91.66	96.96	98.59	98.59

Table 1: Performance Metrics Indices



Conclusion

The Ensemble Model, constructed by combining all the classifiers, outperformed individual models with an accuracy of 97.71%. This emphasizes the significance of machine learning in facilitating early prediction of breast cancer, making it an indispensable tool in healthcare research and medical centers