Machines and Algorithms
http://www.knovell.org/mna

Research Article

Performance Evaluation of Machine Learning Models for Breast Cancer Prediction

Hareem Ayesha[1,*], Laiba Rehman[1]

[1]Institute of Computer Science and Information Technology, The Women University Multan, 60000, Pakistan
[*]Corresponding Author: Hareem Ayesha. Email: hareem.ayesha@wum.edu.pk

Received: 22 December 2023; Revised: 11 January, 2024; Accepted: 26 February 2024; Published: 14 March 2024
AID: 003-01-000034

Abstract: One of the main causes of cancer-related fatalities globally has been breast cancer. The underlying cause of this malady is that it is mostly revealed in late stages after a certain time of its occurrence making it difficult to treat. Another significant characteristic of breast cancer is that it can reoccur after its treatment. Therefore, early prediction of its occurrence and re-occurrence is the best solution to decree the death-rate. This can be achieved through using machine learning based predictive models. This study aims to forecast the breast cancer outcome using machine learning classifiers including Gaussian Naïve Bayes (GNB), Logistic Regression (LR), K-Nearest Neighbor (KNN), Support Vector Machines (SVM), Decision Trees (DT) and Random Forests (RF). The generalization ability and robustness of these distinct classifiers is evaluated on Breast Cancer Wisconsin (Diagnostic) datasets from UCI repository. We analyzed cross-dataset performance in aspects of accuracy, F1 score, precision, and ROC to recognize the most reliable models for accurate breast cancer prediction and to highlight potential dataset-specific biases. The results indicate significant variations in algorithm performance on the dataset. This comparative study not only provides insights into the relative strengths and weaknesses of each machine learning approach but also emphasizes the importance of evaluating predictive models over the dataset to ensure their effectiveness in practical scenarios. Our findings contribute to the expansion of more trustworthy and generalizable breast cancer prediction tools, enhancing early detection and treatment strategies.

Keywords: Cross-Dataset Evaluation; Machine Learning Models; Breast Cancer Detection; Breast Cancer Prediction; Performance Comparison; Predictive Modeling;

1. Introduction

Uncontrollably dividing aberrant cells that have the ability to infect other organs are known as cancers. Breast tumors are caused by abnormal tissue growth in the breast and can be felt as a nipple or discharge, or they can cause a change in the skin's texture surrounding the nipple. Breast cancer has overtaken the lung cancer as most prevalent cancer diagnosed in women around the globe with more than 2.31 million cases in 2022 [1]. The Agency for Research on Cancer states that it is the fourth leading cause of cancer mortality overall. Worldwide, one in five individuals will be diagnosed with cancer at some stage in their lives [2].

Forecasts indicate that the number of cancer diagnoses will increase dramatically over the next several years, increasing by almost 50% between 2020 and 2040. In addition, there have been more cancer deaths—

6.2 million in 2000 compared to 10 million in 2020 [3]. Cancer is the cause of more than one in six fatalities. This emphasizes how important it is to fund both cancer prevention and cancer research.

One of the key components of treating breast cancer is early detection, which increases the likelihood of full recovery. Classifying breast cancer might be challenging because to its wide variety of forms. The most efficient treatment plan is made possible by the precise identification of the breast cancer type. Given the limitations of human classification, automated accurate breast cancer detection could prove advantageous. Over the past 20 years, ML algorithms have been utilized in a wider range of industries, including medicine. Examining medical data is now feasible due to machine learning techniques, which is extremely difficult to analyze manually, with the use of powerful processing units [4]. Over the decay, the number of this research has increased, and every day, new and more efficient methods for analyzing medical data are added to the body of academic literature [5]. Since machine learning mdels are the most accurate and can predict the likelihood of malignancy, they are currently widely used to diagnose breast cancer in women.

Six ML models—KNN, RF, SVM, DT, GNB and LR—for identifying breast cancer are utilized, compared, and demonstrated in this study. The UCI ML library's Wisconsin-Breast Diagnostic Cancer (WDBC) dataset is used in our study [6]. The aim of this study is to demonstrate that machine learning techniques such LR, KNN, RF, SVM, DT, and GNB may be used to answer classification problems. Furthermore, this study helps identify the most effective machine learning method for creating a ML model and offers a framework for contrasting the various strategies. This analysis is critical in the medical field, where the goal is not only high accuracy but also reliability on new data, as models will encounter a wide range of real-world cases. Unlike some previous studies that focus primarily on accuracy, our work highlights the importance of generalization by identifying models prone to overfitting. We recommend that future research or clinical applications consider these aspects of model performance, as high training accuracy without generalizability could lead to incorrect diagnoses when applied to new patient data. Moreover, our work offers novel insights into the specific needs of breast cancer detection by examining how each algorithm performs not only in respect of accuracy but also precision, recall, and F1-score. In a medical diagnostic context, these metrics are crucial because of the high cost associated with both false positives and false negatives. This comparative framework allows researchers and practitioners to make well-informed judgments when choosing models for breast cancer prediction, depending on their specific clinical goals. For example, if high recall is prioritized to avoid false negatives, our results suggest Support Vector Machine as a strong candidate. Alternatively, if interpretability and reduced overfitting are key, Random Forest may be more appropriate. By establishing a standard approach for multi-metric evaluation, our work can serve as a valuable reference point, guiding future studies to apply, refine, or expand upon these methods to improve detection and diagnosis of the breast cancer.

The rest of this article is codified as follow: In Section 2, we review the literature on machine learning techniques for assisting in the diagnosis of breast cancer and provide a number of widely used models and algorithms. The phases and procedures of this experimental investigation are described in Section 3. We report the experiment's findings and compare them with those of other models in Section 4. The research findings are deliberated in Section 5. Lastly, a conclusion along with a discussion on future development is presented in Section 6.

2. Literature Review

The healthcare sector is among the most accurate sectors for data science applications due to the volume of data and the right type of data. The flow of data in hospitals is a continuous process that generally incorporates numerical values. The healthcare system is available to advancements through research on ML and data mining methods. ML techniques help in increasing the effectiveness of a decision support system and automating the decision-making process [7]. Several studies have been conducted to diagnose breast cancer using various ML techniques. In the literature, deep learning approaches such as Convolutional Neural Networks (CNNs) and transfer learning models have also attained extra-ordinary success in breast cancer detection [5]. This study is specifically aimed to explore and compare traditional machine learning

algorithms on the breast cancer dataset, which are often faster to train and easier to interpret. Also, our dataset is comparatively small whereas deep learning models perform better on larger datasets especially in image analysis tasks, and might not have been the best fit for this dataset.

Authors in [8] study evaluated six machine learning algorithms on the WDBC dataset, evaluating classification test sensitivity, specificity and accuracy. The results reveal that all algorithms worked well, with the Multi-layer Perceptron (MLP) method exhibiting the best accuracy at around 99.04%.

Authors in [9] presented a nested ensemble model consisting of two layers for early detection and accurate diagnosis of breast cancer. Using k-fold cross validation, the model classifies tumors with 99.50% accuracy, leaving behind previous models. The K-NN classifier algorithm was used in [10] to gauge the accuracy of breast cancer prediction. It has been demonstrated that supervised ML algorithms can handle incredibly difficult jobs accurately, identifying malignant tumors. The use of this technique may shown value as a significant tool in early detection and treatment of malignant tumors.

An IoT-based diagnostic system for early-stage breast cancer diagnosis is proposed in [11]. They used artificial neural networks and CNNs with hyper parameter optimization for classification. The system uses particle swarm optimization (PSO) feature selection and grid-based search to improve classification performance. Their findings demonstrate that, while the difference is not significant, simple ANNs can nonetheless perform better than CNNs on short datasets.

Authors in [12] analyzed four algorithms on a Breast Cancer dataset: NB, SVM, RF and LR. RF outperformed all others with 99.76% accuracy, making it the optimal choice for disease prediction. In the study [13] The Adaboost algorithm predicted the origins and consequences of breast cancer, together with the cause of mortality. An unassuming Adaboost algorithm was employed.

This research [13] introduced a novel NB (weighted NB) classifier and demonstrated how it can be utilized for breast cancer identification. The efficiency of the weighted NB on the breast cancer database was assessed through a number of trials. The 5-fold cross validation test was used to carry out the tests. Additionally, sensitivity, specificity, and accuracy—three different performance evaluation techniques—are taken into consideration. The weighted NB received the following evaluation values based on the experiments. The determined values for accuracy, specificity, and sensitivity are 98.54%, 98.25%, and 99.11%, respectively.

In [14], RF algorithm was chosen as our main model because it performs better than other algorithms in determining whether breast cancers are benign or malignant. Using a variety of feature selection techniques, it is trained on two distinct subsets of the dataset, each with 16 and 8 characteristics. After hyperparameter adjustment, the RF models are evaluated on a holdout set, yielding 100% and 99.30% accuracy, respectively. Four more ML classification algorithms—SVM, DT, MLP, and KNN—are also used to compare the models. The outcomes demonstrate that Random Forest is the best technique for diagnosing breast cancer.

The table below summarizes the existing work within the specific domain.

Table 1: Literature Review

Authors	Dataset	Algorithm Used	Accuracy
[8]	Wisconsin Diagnostic Breast Cancer (WDBC) dataset	GRU-SVM, Linear Regression, Nearest Neighbor (NN) search, MLP, Softmax Regression and SVM	GRU-SVM (93%), LR (96%), MLP (99%), NN (94%), Softmax Regression (97%) and SVM (96%)
[9]	Wisconsin Diagnostic Breast Cancer (WDBC)	Nested (two-layer) ensemble learners	99.50%
[10]	University of California, Irvine Breast Cancer Dataset	KNN	98%

[11]	Wisconsin Breast Cancer Database	ANN, CNN	ANN (99.2%), CNN (98.5%)
[12]	Wisconsin Breast Cancer datasets	Naïve bayes, Support vector machine (SVM), LR, Random Forest	SVM (98.59%), LR (99.06%), NB (94.83%), Random Forest (99.76%)
[13]	Wisconsin Breast Cancer Database	Naïve bayes, Weighted Naïve bayes	(96.17) NB, (98.54) weighted NB
[14]	Wisconsin Diagnostic Breast Cancer (WDBC)	Random Forest, Support Vector Machine (SVM), Decision Tree, Multilayer Perceptron, and K-Nearest Neighbors	(99.30) RF, (97.90) SVM, (95.80) DT, (96.50) MLP, (93.01) KNN
[15]	Breast cancer database of Srinagarined Hospital in Thailand	Modest Adaboost Algorithm	68.63%
[16]	Wisconsin Breast Cancer Database	Support vector machine (SVM), Naive Bayes, Artificial neural network (ANN), AdaBoost tree	SVM (97.99%), ANN (99.60%), Naïve Bayes (93.32%), AdaBoost (97.19%)
[17]	Wisconsin Breast Cancer Database	Semi-supervised learning (SSL) Co-training	76%
[18]	University of California, Irvine Breast Cancer Dataset	Naive Bayes (NB) classifier and knearest neighbor (KNN)	NB (96.19%), KNN (97.51%)
[19]	SEER database.	Naïve Bayes,NN, C4.5 decision tree	Naïve Bayes (84.5), NN (86.7%), C4.5 decision tree (81.3%)
[20]	General Sample	Back Propagation Neural Network (BPNN) model, Logistic Regression (LR) model	93.7%
[21]	Pubmed	Naïve bayes, SVM	97.3%
[22]	SEER database.	decision tree (C5), ANN, Logistic regression	Decision tree C5 (93.6%), ANN (91.2%), LR (89.2%)

3. Methodology

This section outlines the methodology for evaluating the efficacy of machine learning models by means of data preprocessing and analysis. The steps shown in Figure 1 are used to conduct the investigation. Our research methodology is split up into five sections: Data Acquisition, Data Pre-processing, Classification and Evaluation. These steps are explained in the next sections.

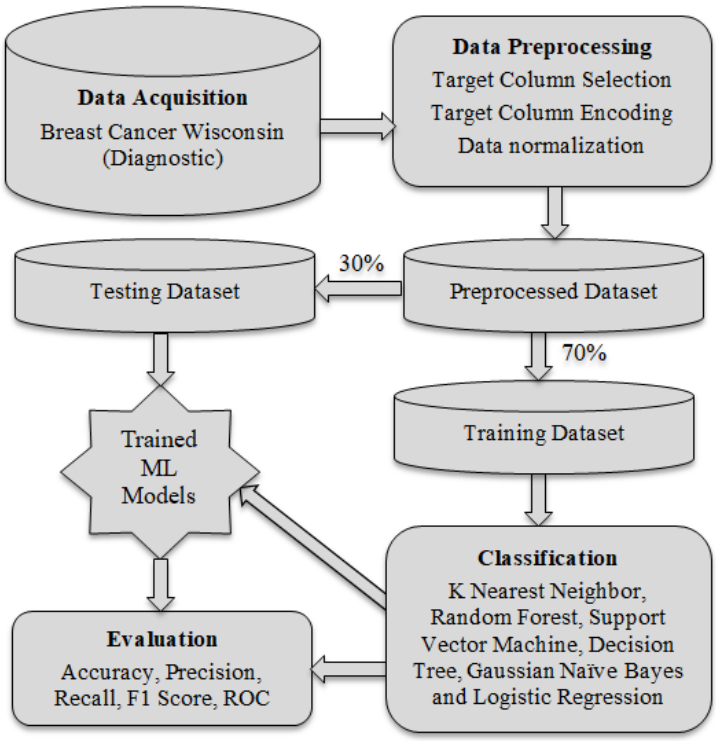


Figure 1: Research Methodology

3.1. Data Acquisition

In our study, we utilized the Breast Cancer Wisconsin (Diagnostic) Data Set available through Kaggle and UCI Machine Learning Repository as our research dataset. The dataset contains 32 columns including 31 features that calculate independent patient characteristics extracted from the digital Fine Needle Aspirate (FNA) test results with the diagnostic classification as either benign (357 cases) or malignant (212 cases) presented in the final 32th feature. Figure 2 illustrates the distribution between these two groups.

Table 2: Features of Breast Cancer Wisconsin (Diagnostic) Dataset

1	id	12	fractal_dimension_mean	23	radius_worst
2	diagnosis	13	radius_se	24	texture_worst
3	radius_mean	14	texture_se	25	perimeter_worst
4	texture_mean	15	perimeter_se	26	area_worst
5	perimeter_mean	16	area_se	27	smoothness_worst
6	area_mean	17	smoothness_se	28	compactness_worst
7	smoothness_mean	18	compactness_se	29	concavity-worst
8	compactness_mean	19	concavity-se	30	concave points_worst
9	concavity-mean	20	concave points_se	31	symmetry_worst
10	concave points_mean	21	symmetry_se	32	fractal_dimension_worst
11	symmetry_mean	22	fractal_dimension_se		

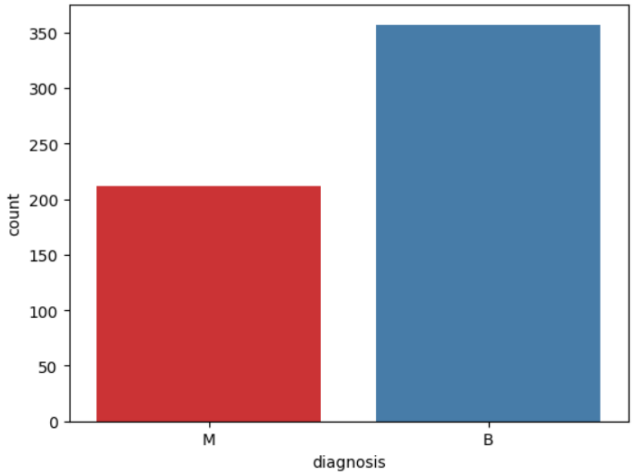


Figure 2: Distribution of Target Variable

3.2. Data Pre-processing

The secondary step is to make the data ready for ML models by applying some preprocessing to data. Following preprocessing is done on the dataset:

Data cleaning: This includes dropping the patient Id column as it does not carry any meaningful relationship with the output feature i.e. diagnosis. Also, it does not contribute to the underlying patterns or correlation in the features.

Separating Output and Input Features: We separated the target feature named as diagnosis from the independent features. The ML models use the independent features to build a relationship with the target feature.

Label Encoding: Finally, we encoded the output categorical feature into binary where 1 denotes malignant and 0 denotes benign. We used LabelEncoder function of sklearn library.

Data Normalization: we applied z-score normalization to input features using StandardScalar function of sklearn library.

3.3. Classification

Six machine learning models—K nearest neighbor, decision tree, Gaussian naïve Bayes, random forest, support vector machine and logistic regression— are utilized to predict breast cancer as malignant or benign. These models are briefly described below.

3.3.1 Logistic Regression

Logistic regression is like linear regression machine learning algorithm that predicts the probability of the class based on dependent features of the dataset. It is widely used for binary classification with a big number of independent features. This statistical algorithm analyzes the relationship between input features by computing the sum of independent features and taking the logistic of the result.

3.3.2 Support Vector Machine

Support Vector Machines (SVMs) demonstrate outstanding performance in their role as classifiers. Sequential Training with PyTorch nymphs the top achievable boundary (hyperplane) among training data classes to achieve separation. Support vectors define an SVM operation that extends distances between these points which represent the closest instances of different classes. SVM achieves optimal performance in high-dimensional spaces because of its maximum boundary. The SVM classifier functions proficiently with linear as well as non-linear data inputs. SVM uses a method named "kernel trick" to apply the data into additional dimension space which enables linear separation of non-linear data. The robustness of SVM

against overfitting reaches its peak in high-dimensional data sets although processing large datasets generates computational expense.

3.3.3 K Nearest Neighbor

K nearest neighbor (KNN) is the simplest yet most important supervised machine learning method, and it is based on a voting system. The concept behind this approach is that the data points that are closest to one another are the most similar data points in the dataset being used. As a result, the values of the data points that are closest to the unseen data point are used to classify it. The value of K indicates the closest neighbors that will be used to make a choice. To determine the nearest points to the supplied unobserved data point, various distance formulas are employed. Euclidean distance, Manhattan distance, and Minkowski distance are among the metrics included.

3.3.4 Gaussian Naïve Bayes

The Gaussian Naive Bayes (GNB) classifier uses Bayes' theorem as its basis to operate as a probabilistic classification model. The model operates under two key assumptions: features DFS independence from each other and distributional values follow Gaussian patterns. Together with estimated mean and variance GNB determines the probability that each data point belongs to a particular class. The model uses prediction time to find the posterior likelihood across different classifications. A data point obtains its assignment by receiving the highest predicted likelihood among all classes. When the independence assumption approximates accuracy GNB performs efficiently on high-dimensional data through its fast processing framework.

3.3.5 Decision Tree

The Decision Tree classifier uses tree-like modeling to perform decisions built around feature value analysis. Each tree node makes decisions based on particular features through specific points which create multiple output routes after the decisions. The model begins its workflow at the root by selecting attributes that provide the optimal split of the data then creates branches that represent different prediction outcomes. The predictive process stops when it reaches leaf nodes which get assigned final class specifications. Both numerical and categorical data thrive under the performance of decision trees which maintain their elegant interpretability. Decision trees display increased susceptibility to overfitting because their growing complexity affects performance accuracy when dealing with new instances.

3.3.6 Random Forest

This classifier functions as an ensemble learning method dedicated to the problem related to classification. Learning algorithms created multiple decision trees through training while merging their predictions to enhance ultimate accuracy and stability. Climate change presents challenges to forest ecology since each tree in the "forest" builds from unique training subsets while choosing random features at nodes to enhance ensemble diversity. Each tree in the "forest" completes its vote for class label during classification before the majority choice becomes the final prediction output. Even when studying large data collections or working with features at varying measurement scales and distributions Random Forest avoids overfitting while simultaneously achieving better generalization performance.

3.4. Evaluation

We used Accuracy, Recall, Precision, F1 Score, and ROC curve metrics to apprise the performance of selected machine learning prediction models.

Accuracy: It is a metric used to measure the correctness of a model's predictions. It is defined as the proportion of accurate predicted instances to the total number of instances.

$$\mathrm{Ac}curacy = \frac{\mathrm{TP+TN}}{\mathrm{TP+FP+TN+FN}} \quad (1)$$

Precision: It is a metric that evaluates the precision of positive predictions generated by a model. It indicates the proportion of accurately anticipated positive instances to the total expected positive cases.

$$Precision = \frac{\text{TP}}{(\text{TP}+\text{FP})} \tag{2}$$

Recall: It (also refer to as sensitivity or true positive rate) quantifies a model's capacity to accurately identify all positive instances within the dataset. It tells us what fraction of actual positive cases was accurately predicted as positive.

$$Recall = \frac{\text{TP}}{(\text{TP}+\text{FN})} \tag{3}$$

F1-score: It is a measure that integrates both recall and precision to deliver a single value of a model's performance, especially useful when the dataset exhibits imbalanced. It is the harmonic mean of recall and precision, balancing the trade-off between them.

$$\text{F1}-Score = \frac{(\text{Precision}+\text{Recall})}{2} \tag{4}$$

Confusion Matrices: A confusion matrix is used as performance summarizer tool for machine learning functions, particularly for classification tasks. It quantifies the efficiency of classification algorithm by matching the predicted classifications to the actual classifications.

4. Experimental Analysis and Results

In our study, we utilized BreastCancer Wisconsin (Diagnostic) dataset available at UCI machine learning repository. All experiments are done in jupyter notebook using python. After applying the necessary preprocessing mentioned in previous section, dataset is divided into testing and training set with 20-80 ratio respectively. All six classification models are trained on training dataset and training accuracies are analyzed. Then testing dataset was fed to trained models and testing accuracies are analyzed. Both testing and training accuracies of all models are shown below in the Table 3. It can be seen that decision tree gave 100% training accuracy while highest testing accuracy is achieved on support vector machine which is 98.25%.

Table 3: Training and Testing Accuracy of ML Models

Model	Training Accuracy	Testing Accuracy
Logistic Regression	98.9	96.49
Support Vector Machine	98.46	98.25
K-Nearest Neighbor	98.02	96.49
Decision Tree	100	93.86
Gaussian Naïve Bayes	94.95	93.86
Random Forest	99.78	97.37

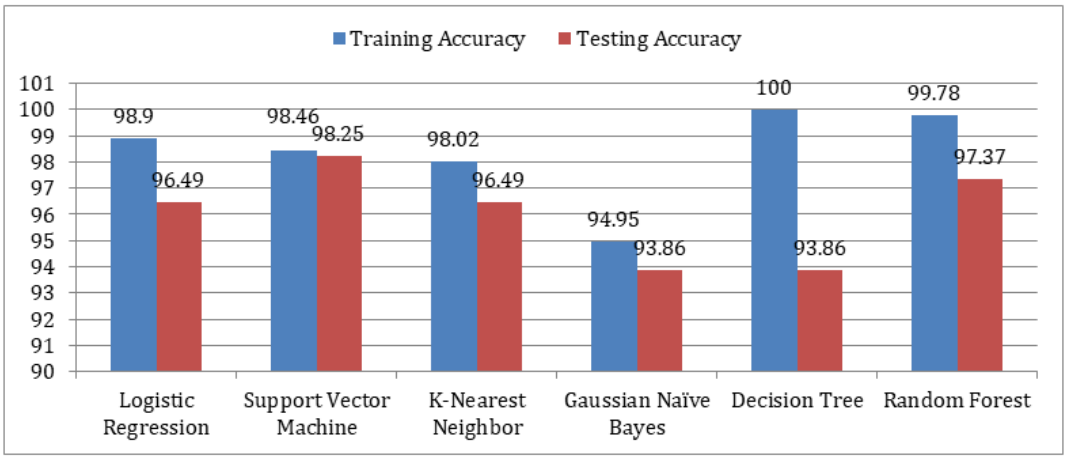


Figure 3: Training and Testing Accuracy of ML Models

Table 4 displays the comparison of the performance of all six models—KNN, RF, SVM, DT, GNB LR —based on accuracy, recall ,precision and F1 score. It is evident that Support Vector Machine achieved better results than the other models in terms of precision, recall, and F1-Score. Logistic regression also produced strong results and came in second place.

Table 4: Precision, Recall and F1-Scoore of Machine Learning Models

Model	Precision	Recall	F1-Score
Logistic Regression	97.78	93.62	95.65
Support Vector Machine	100	95.74	97.83
K-Nearest Neighbor	100	91.49	95.56
Gaussian Naïve Bayes	93.48	91.49	92.47
Decision Tree	93.48	91.49	92.47
Random Forest	100	93.62	96.7

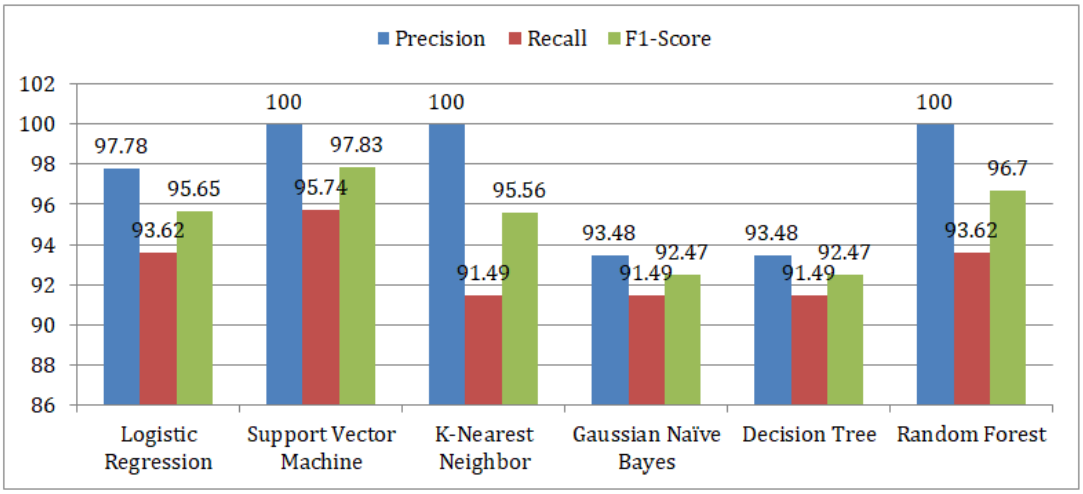


Figure 4: Precision, Recall and F1-Score of ML Models

ROC curve is also be shown in the Figure 5. The model with the highest ROC AUC is support vector machine indicating that it had the best possible balance of sensitivity and specificity, recall and precision.

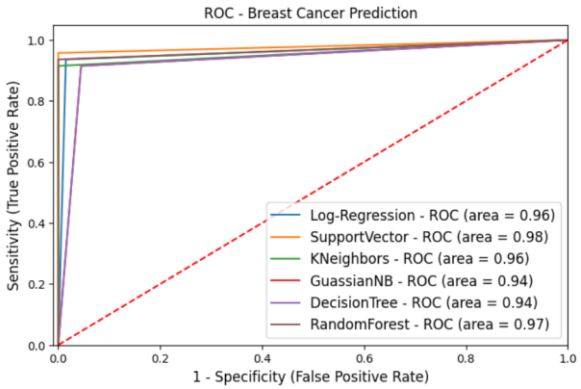
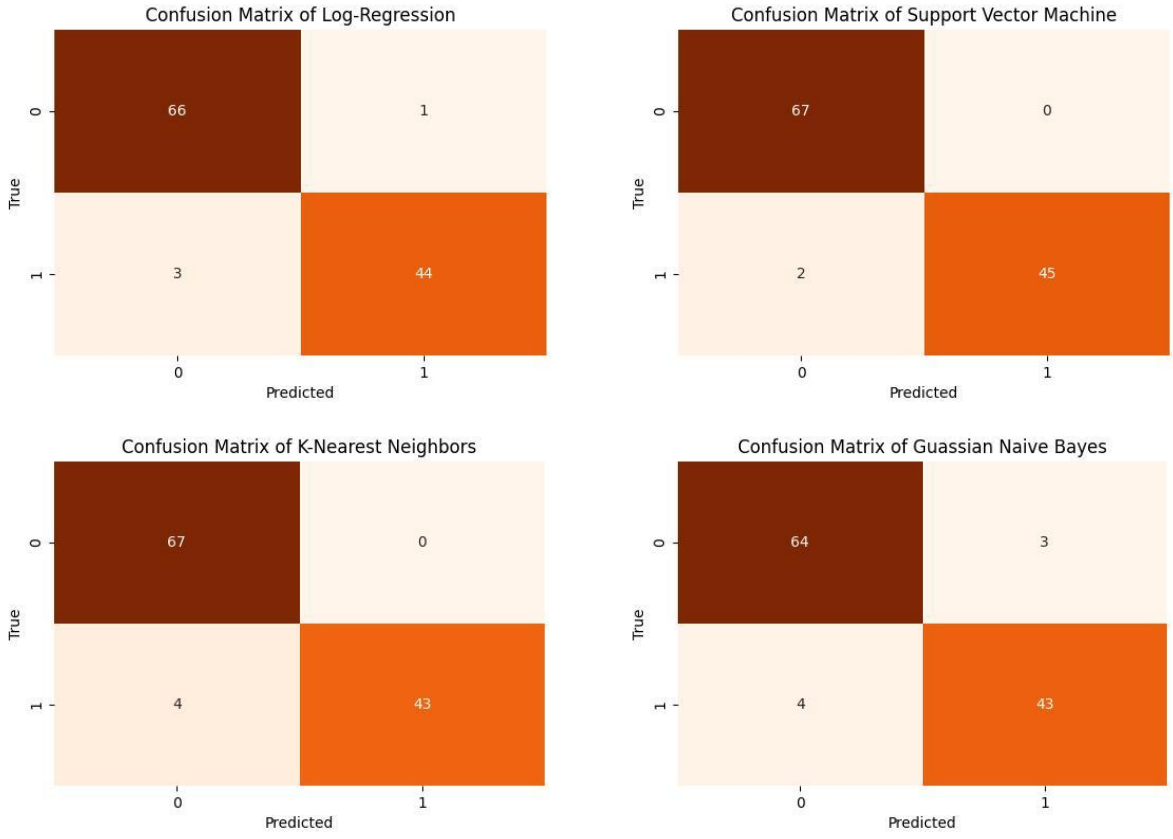


Figure 5: ROC-Curve of ML Models

Confusion matrices of all ML models used in this study to predict the breast cancer are compared below in figure 6. The confusion matrices show the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) for each model on the test dataset. The machine learning systems that missed the fewest cancerous samples had the fewest false negatives. The ML models that misidentified the fewest benign instances as cancer had the lowest number of false positives (FP).



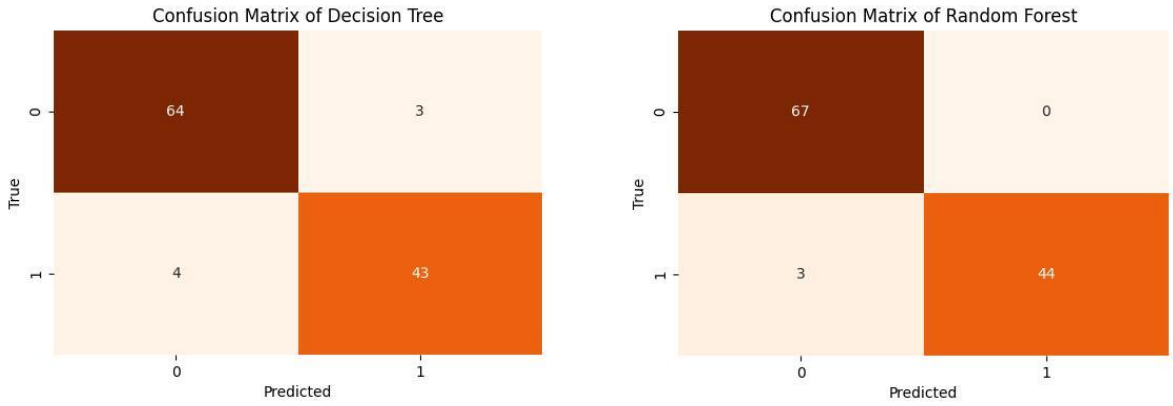


Figure 6: Confusion Matrices of ML Models

5. Discussion

Using a breast cancer dataset, our work provides a broad comparative examination of six ML algorithms: Logistic Regression, KNN, SVM, GNB, RF, and DT. By assessing each model based on four major performance metrics—accuracy, recall, precision and F1-score—we reveal unique strengths and limitations for each algorithm, enabling a more tailored approach to model selection in breast cancer diagnosis.

Logistic Regression: This model showed high precision (97.78%) and a good balance between training and testing accuracy, indicating a low risk of overfitting. Its simplicity and decipherability make it appropriate for cases where model transparency is critical. On the other hand, the limitation is that it assumes linearity, which could limit its ability to capture complex relationships in the data.

Support Vector Machine (SVM): It achieved perfect precision (100%) and high recall (95.74%), making it an excellent choice for minimizing false negatives, which is crucial in breast cancer detection. Its strength lies in its robustness to outliers and ability to work well with complex, non-linear boundaries. However, the model's computational complexity and sensitivity to hyper-parameter choices can be limitations, especially for large datasets.

K-Nearest Neighbors (KNN): KNN demonstrated a strong recall (95.56%), suggesting it is effective at capturing positive cases. Its non-parametric nature enables it to adjust to different data distributions, creating it flexible. However, it is computationally expensive for big datasets and can be delicate to the choice of the number of neighbors (k) and feature scaling, potentially affecting its performance.

Gaussian Naïve Bayes: This model showed reasonable accuracy and precision but had a lower recall (91.49%) compared to other models, which could limit its effectiveness in identifying malignant cases. Its strength is in handling small datasets and performing well with normally distributed data. However, the supposition of feature independence is often unrealistic in complex medical datasets, which may limit its performance.

Decision Tree: The Decision Tree model achieved perfect accuracy (100%) on the training data but showed a significant drop in testing accuracy (93.86%), highlighting its tendency to over fit. Its interpretability and simplicity are valuable for clinical applications, but its limitation is in generalizability; the model might not perform as well on unseen data without regularization techniques like pruning.

Random Forest: Random Forest provided both high training accuracy (99.78%) and strong testing performance (97.37%), indicating a good balance between fit and generalization. Its strength lies in reducing overfitting by averaging multiple decision trees, making it resilient and reliable. However, the model can be computationally intensive and may lack transparency, as the ensemble structure makes it harder to interpret compared to simpler models.

This comparative analysis not only demonstrates each model's effectiveness in terms of accuracy but also provides a practical understanding of how each model handles the specific challenges of breast cancer prediction. By offering insights into these strengths and limitations, our study establishes a benchmark for future research, guiding researchers and clinicians to select models that best align with their specific goals and dataset characteristics.

6. Conclusion

Research primarily examined the extent to which machine learning prediction algorithms identify breast cancer. Our research employed the Breast Cancer Wisconsin (Diagnostic) Data Set for analysis. Six trained ML algorithms worked on an 80% subsampled version of the original dataset. The participated algorithms in the analysis including KNN, RF, SVM, DT, GNB and LR. Our analysis of these predictive models happens through evaluation with testing data obtained from 20% of the original dataset. Three different evaluation methods determined the assessment results including accuracy, recall, precision and F1-score and ROC curve evaluation. SVM produced the highest accuracy during testing yet it shared top performance with linear regression alongside decision trees and random forests for training accuracy. Support vector machine led all prediction models with 100% precision alongside 95.74% recall and 97.83% F1-score and 98% ROC curve accuracy. The evaluation of various machine learning models reveals their predictive abilities toward diagnosing malignancy and benignity in breast cancer cases. Our objective is to evaluate these models through future modifications of their parameter settings.

References

[1] Rasool, Abdur, Chayut Bunterngchit, Luo Tiejian, Md Ruhul Islam, Qiang Qu, and Qingshan Jiang. "Improved machine learning-based predictive models for breast cancer diagnosis." *International journal of environmental research and public health* 19, no. 6 (2022): 3211.

[2] https://www.iarc.who.int/cancer-type/breast-cancer/

[3] Bray, Freddie, Mathieu Laversanne, Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Isabelle Soerjomataram, and Ahmedin Jemal. "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." *CA: a cancer journal for clinicians* 74, no. 3 (2024): 229-263.

[4] Ayesha, Hareem, Sajid Iqbal, Mehreen Tariq, Muhammad Abrar, Muhammad Sanaullah, Ishaq Abbas, Amjad Rehman, Muhammad Farooq Khan Niazi, and Shafiq Hussain. "Automatic medical image interpretation: State of the art and future directions." *Pattern Recognition* 114 (2021): 107856.

[5] Tariq, Mehreen, Sajid Iqbal, Hareem Ayesha, Ishaq Abbas, Khawaja Tehseen Ahmad, and Muhammad Farooq Khan Niazi. "Medical image based breast cancer diagnosis: State of the art and future directions." *Expert Systems with Applications* 167 (2021): 114095.

[6] https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic

[7] Ayesha, Hareem, Mehreen Tariq, and Sundas Israr. "Computer Aided Deep Image Captioning for Medical Images." *Machines and Algorithms* 2, no. 1 (2023): 16-16.

[8] Agarap, Abien Fred M. "On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset." In *Proceedings of the 2nd international conference on machine learning and soft computing*, pp. 5-9. 2018.

[9] Singh, Kuljeet, Sourabh Shastri, Sachin Kumar, and Vibhakar Mansotra. "BC-Net: Early Diagnostics of Breast Cancer Using Nested Ensemble Technique of Machine Learning." *Automatic Control and Computer Sciences* 57, no. 6 (2023): 646-659.

[10] Pandey, Sharma, A. Sharma, M. K. Siddiqui, D. Singla, and J. Vanderpuye-Orgle. "ai3 prediction of breast cancer using k-nearest neighbour: a supervised machine learning algorithm." *Value in Health* 23 (2020): S1.

[11] Ogundokun, Roseline Oluwaseun, Sanjay Misra, Mychal Douglas, Robertas Damaševičius, and Rytis Maskeliūnas. "Medical internet-of-things based breast cancer diagnosis using hyperparameter-optimized neural networks." *Future Internet* 14, no. 5 (2022): 153.

[12] Sivapriya, J., Aravind Kumar, S. Siddarth Sai, and S. Sriram. "Breast cancer prediction using machine learning." *International Journal of Recent Technology and Engineering (IJRTE)* 8, no. 4 (2019): 4879-4881.

[13] Karabatak, Murat. "A new classifier for breast cancer detection based on Naïve Bayesian." *Measurement* 72 (2015): 32-36.

[14] Minnoor, Manas, and Veeky Baths. "Diagnosis of breast cancer using random forests." *Procedia Computer Science* 218 (2023): 429-437.

[15] Thongkam, Jaree, Guandong Xu, Yanchun Zhang, and Fuchun Huang. "Breast cancer survivability via AdaBoost algorithms." In *Proceedings of the second Australasian workshop on Health data and knowledge management-Volume 80*, pp. 55-64. 2008.

[16] Wang, Haifeng, and Sang Won Yoon. "Breast cancer prediction using data mining method." In *IIE Annual Conference. Proceedings*, p. 818. Institute of Industrial and Systems Engineers (IISE), 2015.

[17] Kim, Juhyeon, and Hyunjung Shin. "Breast cancer survivability prediction using labeled, unlabeled, and pseudo-labeled patient data." *Journal of the American Medical Informatics Association* 20, no. 4 (2013): 613-618.

[18] Amrane, Meriem, Saliha Oukid, Ikram Gagaoua, and Tolga Ensari. "Breast cancer classification using machine learning." In *2018 electric electronics, computer science, biomedical engineerings' meeting (EBBT)*, pp. 1-4. IEEE, 2018.

[19] Abdelghani, Bellaachia, and Erhan Guven. "Predicting breast cancer survivability using data mining techniques." SIAM INTERNATIONAL CONFERENCE ON DATA MINING, 2006.

[20] Alarabeyyat, A., & Alhanahnah, M. (2016, August). Breast cancer detection using k-nearest neighbor machine learning algorithm. In *2016 9th International Conference on Developments in eSystems Engineering (DeSE)* (pp. 35-39). IEEE.

[21] Cruz, Joseph A., and David S. Wishart. "Applications of machine learning in cancer prediction and prognosis." *Cancer informatics* 2 (2006): 117693510600200030.

[22] Delen, Dursun, Glenn Walker, and Amit Kadam. "Predicting breast cancer survivability: a comparison of three data mining methods." *Artificial intelligence in medicine* 34, no. 2 (2005): 113-127.