



Analysis Performance Classification of Wisconsin Diagnostic Breast Cancer WDBC Via Supervised Machine Learning Algorithms

Mohammed Al_Duais^{1,2*}, Abdullah Al-Qudaimi³ and Mohammed Nasser Ali Alharasi⁴

¹ Department of Computer science, faculty of Engineering and IT, Amran university, Yemen.

² Department of Artificial Intelligence, Faculty of Engineering and IT, Aljeel Al Jadeed university, sana 'a, Yemen.

³ Department of Information Technology, Faculty of Computing and IT, University of Science & Technology, Sana'a, Yemen

⁴ Master of Information Technology Management Teaching Assistant, Computer Science Department. Faculty of Engineering and information Technology, Amran university.

Received 2 Agu. 2025; accepted in final form 10 Oct. 2025.

Abstract

Breast cancer is one of the leading causes of cancer death in women. It frequently results in deadly outcomes due to delayed identification in advanced stages. Early detection and treatment significantly improve a breast cancer patient's chance of survival. Recent advancements in machine learning have opened up possibilities for early detection. **Objective:** The objective of this study was to apply and compare several machine learning (ML) algorithms to see which performed better for breast cancer prediction. **Methods:** To achieve the study's objectives, several steps were taken. The Wisconsin Diagnostic Breast Cancer (WDBC) has been used. Data preprocessing includes normalizing, removing outliers, and fixing missing values. The data was separated into two parts: training (80%) and a testing set (20%). Several techniques have been utilized, including Multi-Layer Perceptron (MLP), Ada Boost (AB), Gradient Boosting (GB), Support Vector Machines (SVM), K-Nearest Neighbors algorithm, Decision Trees (DT), and Gaussian Naïve Bayes (NB). Accuracy, precision, recall, and F scores were used to evaluate performance. **Result:** The Multi-layer Perceptron (MLP) achieved the highest accuracy of 0.998, compared with other techniques. **Evaluated:** To evaluate the proposed model via comparison with available studies. This study achieved superior performance to existing works for the classification prediction of breast cancer.

Keywords: Breast cancer, prediction, Machine Learning Techniques, classification

المخلص: سرطان الثدي هو أحد الأسباب الرئيسية لوفاة النساء بسبب السرطان. وغالبًا ما يؤدي إلى نتائج مميتة بسبب التأخر في التعرف عليه في المراحل المتقدمة. ويحسن الكشف المبكر والعلاج بشكل كبير من فرصة بقاء مريضة سرطان الثدي على قيد الحياة. وقد فتحت التطورات الحديثة في التعلم الآلي الأبواب أمام إمكانيات الكشف المبكر. الهدف: هدفت هذه الدراسة إلى تطبيق ومقارنة العديد من خوارزميات التعلم الآلي (ML) لمعرفة أيها كان أداءها أفضل للتنبؤ بسرطان الثدي. الطريقة: لتحقيق أهداف الدراسة، تم اتخاذ العديد من الخطوات، بما في ذلك جمع البيانات ومعالجتها مسبقًا، والتي تتضمن تطبيع البيانات وإزالة بيانات الخطوط العريضة وإصلاح القيم المفقودة. تم فصل البيانات إلى قسمين: التدريب (80٪) ومجموعة الاختبار (20٪). تم استخدام العديد من التقنيات، بما في ذلك Multi-Layer Perceptron (MLP) و Ada Boost (AB) و Gradient Boosting (GB) و Support Vector Machines (SVM) و Xوارزمية K-Nearest Neighbors و Decision Trees (DT) و Gaussian Naïve Bayes (NB). استُخدمت Accuracy, precision, recall, and F1 scores لتقييم المقترحات المُقدمة من خلال مقارنة الأداء في دقة التنبؤ مع الدراسات المتاحة. حققت هذه الدراسة أداءً أفضل من الأعمال الحالية في مجال التنبؤ بتصنيف سرطان الثدي.

1- Introduction

Breast cancer is a distinct type of cancer that affects more women than any other demographic. There are two primary causes: modifiable and non-modifiable factors. Environmental and behavioral issues

*Corresponding Author: dauais2025ww@gmail.com

are examples of modifiable characteristics. Gender and family history are two instances of uncontrollable conditions. [1],[2]. Breast cancer is a primary cause of death for women worldwide. The World Health Organization (WHO) estimates that approximately 963,000 women will die in 2021. However, the agency predicts that the figure will climb to 2.9 million globally. Breast cancer can affect both women and men, although relatively rarely. [3,4] Several scientific works on the automated prediction of breast cancer have been presented. Machine learning is another scientific method for extracting hidden knowledge from large volumes of data using a range of tactics and algorithms. [6]–[8]. The goal of this study is to offer a summary of the efficacy of existing classification methods in order to suggest potential avenues for further inquiry. The objectives of this study are to determine the best classification algorithms for properly predicting the malignancy of breast cancer datasets through a comparative analysis of categorized approaches.

Early detection of breast cancer improves the patient's long-term survival and increases the likelihood of a successful treatment strategy. If the disease is detected and diagnosed early, roughly 98% of patients will live for five years or more, but approximately one in every four (26%) will live for five years or more if the condition is discovered and diagnosed later. The "two-week wait" is the most common way of diagnosing breast cancer. Microscopic inspection is commonly used by pathologists to diagnose breast cancer. As a result, the CAD method reduces diagnostic uncertainty while streamlining diagnosis. [9]-[11]. Similarly, breast cancer is the second most frequent cancer worldwide. In 2022, women were diagnosed with 2,296,840 new instances of breast cancer. Men were not reported to have breast cancer [12].

Machine learning is a critical issue for the early diagnosis and detection of breast cancer, and numerous papers have been published in Scopus journals. Nassif et al. (2022) [13] - [15] traced publications in journals in Scopus. The result said that "More than 1,000 papers were found that were published between January 2010 and May 2020". Figure 1 depicts the timeframe.

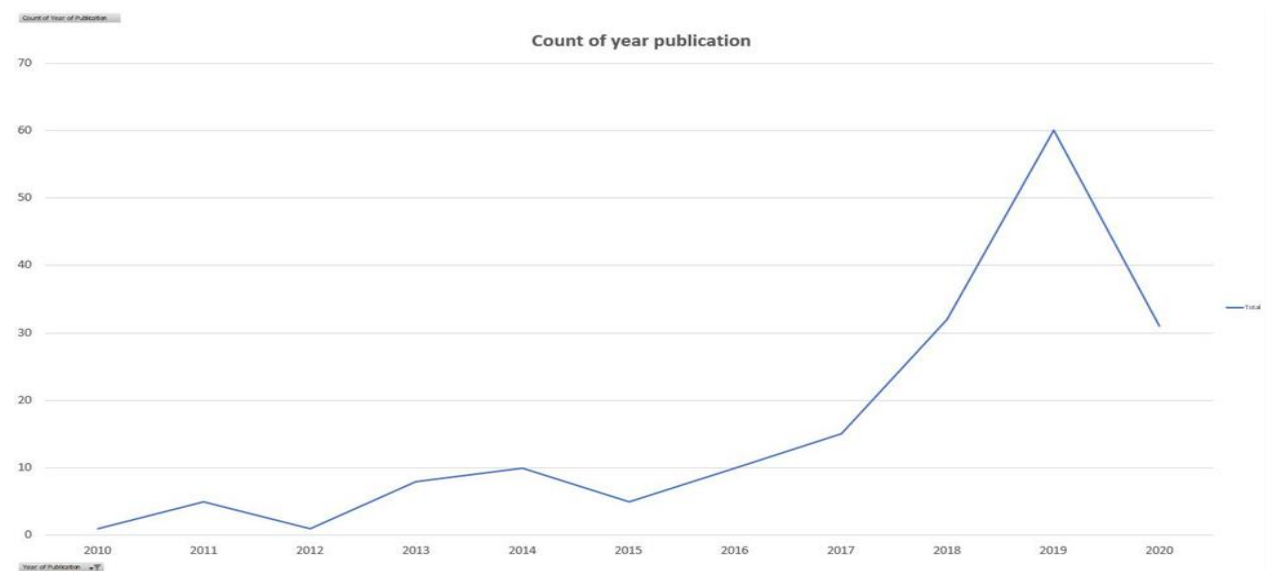


Figure 1: Number of publications between 2010 and 2020 in Scopus journals

As shown in Figure 1, the publication curve is increasing each year, which indicates a powerful motivation for highlighting this kind of topic.

Several traditional approaches have been used in prior publications to address categorization difficulties. These strategies may fail to account for the differential costs of misclassification among different categories, resulting in poor results. This study examined numerous methodologies and assessed both standard procedures and technological demands, as well as their merits and drawbacks. The research gaps in current machine learning approaches for application, as well as future researcher ideas, are then discussed [16]-[18].

Machine learning algorithms have grown into exciting tools in the field of medicine. They are capable of processing and analyzing massive volumes of data in order to gather useful information for decision-making. As a result, they will be ideal for predicting and detecting breast cancer, as well as classifying it as benign or malignant [19], [20]. The highlight of this study is the proposed model. The study has three significant contributions: It examines many Supervised Machine Learning Algorithms, including Multi-Layer Perceptron (MLP), Ada Boost (AB), Gradient Boosting (GB), Logistic Regression (LR), Naive Bayes (NB), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Decision Tree (DT). The performance of the proposed model is examined by using the breast cancer WDBC to determine which classifier is most effective for breast cancer prediction.

The rest of this paper is organized as follows: Section 2 presents Related work. Section 3: describes methods. Section 4: presents Results and Discussions. Section 5: Comparison of the Proposed Model. Section 6: Conclusion.

2- Related works

The previous studies provide a comprehensive summary of the research on data machine learning techniques concerning breast cancer that is currently accessible. The review describes the algorithms that were employed, as well as their scope and limitations. The summary has been recorded in the table (1) as follows:

Rahman et al [22] provided four classification approaches for breast cancer data: Classification and Regression Trees (CART), Alternating Decision Trees (AD Trees), Best First Trees (BF Trees), and j48. The experiment in this study was conducted using the Weka tool. Cross-validation, which uses 10 folds to train each classifier and 1fold for testing, and percentage split, which uses two-thirds of the dataset for training and one-third for testing, are the two test beds for which the classifier has been applied. According to the authors, the decision trees are easily understood and have a common structure, making it possible to extract the rules from them. Additionally, according to the authors, the J48 classifier has the highest accuracy (99%). Liu et al [23] provided four classification approaches for breast cancer data: Classification and Regression Trees (CART), Alternating Decision Trees (AD Trees), Best First Trees (BF Trees), and j48.

Assiri et al [24] offered a data mining diagnostic method based on categorization and association. The authors classified patterns that are commonly identified with benign and malignant cases using Frequent Patterns (FP) in association rule mining. In addition, the authors employed a decision tree algorithm to estimate the likelihood of cancer based on age. The authors created an FP-growth method for creating frequent items without generating candidates, which enhances the program's performance. The prediction of accuracy is 94%.

Islam et al [25] proposed a disease status prediction model using a hybrid approach that combined Decision Trees (DT) and Support Vector Machines (SVM). Two main parts of the system's approach—decision tree, support vector machines, and information treatment and option extraction—are used to alert users to the severity of the illness. By contrasting the results of the proposed model with those of Instance-based Learning (IBL), Sequential Minimal Optimization (SMO), and Nave

Bayes (NB), the author demonstrated that the suggested algorithms performed 91% more accurately than the comparative algorithms.

Srinivas et al [26] investigated the effectiveness of conventional supervised learning algorithms like CART, ID3, C4.5, and Nave Bayes, as well as the usage of decision trees in the prediction of breast cancer. The Weka tool is used to experiment. Five key features from the study were found to be employed in the forecast. The random tree is the most accurate categorization approach for breast cancer, according to the results.

Chaurasia et al. [27] assessed the effectiveness of the classification algorithms for breast cancer data using BF Tree (Best First Tree), IBK (K-nearest neighbor classifier), and SMO (Sequential Minimal Optimization). Three evaluation criteria were used to determine which method was better: accuracy, time, and correctly classified cases. The experiment was conducted using the Weka data mining application. SMO methods fared better than the other two algorithms in terms of accuracy and error rate. The most important factors for raising forecast accuracy were shown by the results, which were 95.46%.

Sumbaly et al. [28] suggest the J48 decision tree classification approach for detecting breast cancer. This study includes breast cancer types, risk factors, symptoms, and therapy. The results demonstrated that the J48 algorithm can achieve 94.5% accuracy with correctly identified samples, suggesting that neural networks and digital mammography could be alternative methods for breast cancer prediction.

3. Research Method

To achieve the objectives of this study, we will be using two techniques: traditional machine learning classifiers (ML) and ensemble machine learning classifiers (ML), all details as follows:

3.1 Dataset collection

The WDBC dataset was used in this investigation. It can be found in the UCI repository [29]. This dataset is used to determine if a cancer is benign or malignant. A digital photograph of an affinogen needle aspirate (FNA) of a mass in the breast is utilized to extract characteristics. It describes the characteristics of the image's nucleus. There are 569 entries in total in this dataset, 212 of which are malignant and 357 of which are benign.

3.2 Data preprocessing

Data preprocessing was used to locate missing values in the data, discover and remove outliers, and resolve inconsistencies. The dataset is missing 16 characteristic values. This study used the standard scaler to normalize the data. To solve the problem of outlier data, this study used the Z-score. The confusion matrix is used to assess accuracy and shows how tuples in training and testing data are correctly identified. The data in this study were divided into two parts: training sets (80%) and testing sets (20%).

3.3 Measurement of the Performance in the Machine Learning Techniques

This section explains all measurements for testing performance. Almost all previous studies used several measurements to evaluate the performance. All formulas were recoded as follows.

- 1- Accuracy = $(TP + TN) / (TP + FP + FN + TN)$ (1)
- 2- Precision: $TP / (TP + FP)$ (2)
- 3- Recall= $TP / (TP + FN)$ (3)
- 4- F1 Score= $2 * (Precision * Recall) / (Precision + Recall)$ (4)
- 5- Specificity = $Tn / (Tn + Fn)$ (5)

3.4 Design the proposed Model

This study presented a way to identify breast cancer by designing the proposed model. Two techniques were applied to breast cancer data, with different classifiers performing differently on the

same data. This proposed model was designed to achieve the main objective of this study: to find the best classification for early detection and diagnosis of breast cancer. All steps of the design of the proposed model are as follows:

- Step 1: Data collection
- Step 2: Pre-processing the dataset
- Step 3: Determining the attributes of the data
- Step 4: Dividing the data into 80% for training and 20 % for testing
- Step 5: Fined the labels malignant and benign
- Step 6: Applying the model with different techniques
- Step 7: Comparing the performance of different classifiers
- Step 8: Evaluation of the model by using some criteria
- Step 9: Selecting the best performance model.

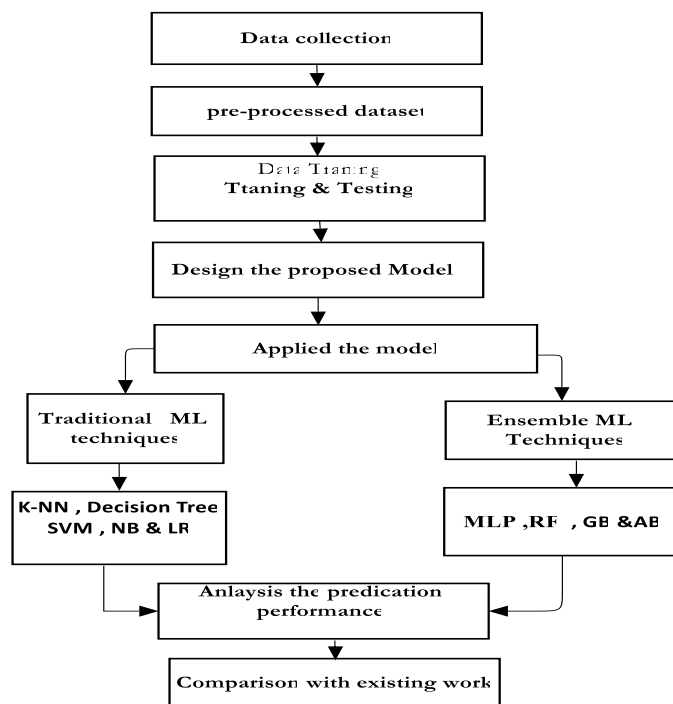


Figure 1: Proposed model for determining the best classifier

3.4 Machine Learning Techniques

This section focuses on several Machine learning algorithms that are used for prediction and classification. All details as follows:

- 1) The multilayer perceptron (MLP) is a supervised, feedforward back-propagation network with three layers: input, output, and hidden. These layers serve a critical function in retrieving key information during learning and altering weights accordingly, as follows formula

$$(xi) = b + \sum_{i=1}^n xi wi \tag{6}$$

where, xi = inputs of incoming layers. wi = weights of hidden layers neurons. b = initial weight

- 2) GB Gradient boosting (GB) is an ensemble supervised machine learning technique that may be applied to both classification and regression tasks. It works sequentially, gradually strengthening weak learners by emphasizing high-value data points. This can be defined as:

$$Y = ax + b + e \tag{7}$$

Where e represents the error and shows the inexplicable data

- 3) Adaboost: AdaBoost(AB) is a boosting-based ensemble method that combines multiple "weak classifiers" to build a "strong classifier" by iteratively updating their weights. Training will continue until the error is reduced to a minimum. AdaBoost is a boosted classifier utilizing the formula.

$$FT(x) = \sum_{t=1}^T f_t x \tag{8}$$

Where each f_t is a weak learner that takes an object as input and returns a value indicating the class of the object

- 4) K-Nearest Neighbors (KNN): It classifies new data points based on their similarity to the entire training set. To label a new data point, it first calculates its distance from separate label points and then locates the nearest neighbor.
- 5) Decision Tree (DT): Decision Tree (DT): In this method, instances are classified according to the value of their attributes. It is divided by the Gini Index and Information Gain. The leaf node indicates the label.
- 6) Nave Bayes(BN): Nave Bayes. This method is founded on Bayes' theorem. It makes predictions based on probabilities.
- 7) Support Vector Machine (SVM): Support Vector Machines (SVM): This classification method requires prior distribution knowledge for categorization. It categorizes data points using hyperplanes.
- 8) Random Forest (RF): Random Forest (RF): This is an ensemble method. It generates majority votes and categorizes various DTs derived from different data sets.
- 9) Logistic Regression: Logistic regression is a linear regression model used to solve classification problems. It provides a probability for two alternative outcomes. This method is beneficial for evaluating the risk of illness in the healthcare profession.

4. Results and Discussion

In this study, two machine learning techniques were used: traditional (or standard) ML techniques and Ensemble ML techniques.

4.1 Experimental Results of ML Techniques

This section is focused on the dedication prediction of several machine learning techniques for both datasets, namely, the testing set and training set.

4.1.1 Analyze the predication performance of ML Techniques with the Training set

This section focuses on the dedication prediction of several machine learning techniques for the training set. The performance of classifiers for traditional machine learning ML algorithms is documented in Table 2 below.

Table 1: Classification Performance of ML Techniques for Training Set

ML Techniques	SVM	DT	NB	K-NN	LR	MLP	GB	AB	Average Performance
Accuracy	0.984	0.992	0.964	0.983	0.983	0.994	0.983	0.984	0.983
Precision	0.974	0.981	0.974	0.963	0.976	0.981	0.964	0.983	0.975
Recall	0.982	0.976	0.967	0.984	0.974	0.991	0.964	0.977	0.977
FI-Score	0.971	0.982	0.965	0.951	0.96	0.991	0.973	0.983	0.972

Table 1 shows the outcomes of several classifiers on the training dataset. The MLP classifier demonstrated the highest performance, achieving an accuracy rate of 0.994. The Decision Trees (DT) achieved the second-highest performance of classification, with an accuracy rate of 0.992. This can be attributed to the decision tree algorithm's suitability for binary classification and its ability to retain all data. In contrast, the Naïve Bayes classifier had the lowest accuracy at 0.964. The average accuracy of machine learning techniques was 0.983. On the other hand, the classification performance using machine learning approaches is shown in Figure 2 below:

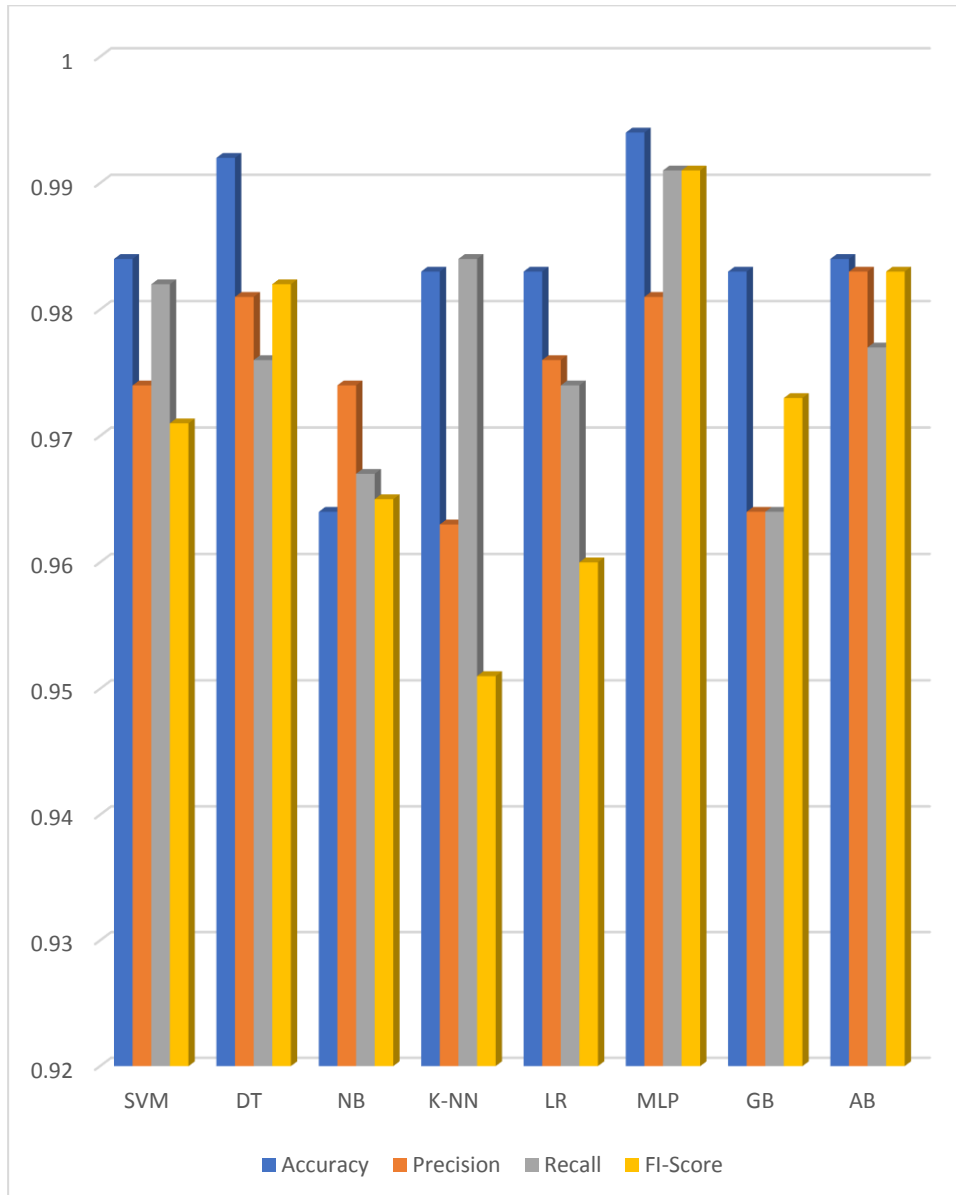


Figure 2. Classifier Performance of ML Techniques on Training Set.

Figure 2 shows the various machine-learning classification approaches used on the breast cancer dataset. From the key mapping, there are four measurements to test performance, each of which has a different color. The high performance of accuracy is achieved with the MLP algorithm. When it

comes to Decision Trees (DT) and the AB algorithm, both were winning the second performance for classification accuracy compared with other machine learning techniques. The lowest Naive Bayes performance of accuracy is Naive Bayes.

4.1.2 Analyze the predication performance of ML Techniques with the Testing set

This section focuses on the dedication prediction of several machine learning techniques for the testing set. The performance of classifiers for traditional machine learning ML algorithms is documented in Table 2 below.

Table 2: Classifier Performance ML Techniques for Testing set

ML Techniques	SVM	DT	NB	K-NN	LR	MLP	GB	AB	Average Performance
Accuracy	0.973	0.988	0.952	0.974	0.983	0.998	0.984	0.971	0.978
Precision	0.934	0.961	0.944	0.983	0.976	0.983	0.962	0.953	0.962
Recall	0.982	0.973	0.977	0.924	0.914	0.984	0.960	0.962	0.959
FI-Score	0.971	0.982	0.985	0.951	0.96	0.981	0.972	0.972	0.972

Table 2 presents the results of different classifiers on the tasting dataset. MLP classifier demonstrated the highest average performance, achieving 0.998 accuracy. The decision tree classifier obtained the second-highest t performance in classification, at 0.988 accuracy, for the testing set. This can be attributed to the decision tree algorithm's suitability for binary classification and its ability to retain all data. In contrast, the Naïve Bayes classifier had the lowest accuracy at 0.952. The average accuracy of machine learning techniques was 0.978. The classifiers' performance using machine learning approaches is shown in Figure 2 below.

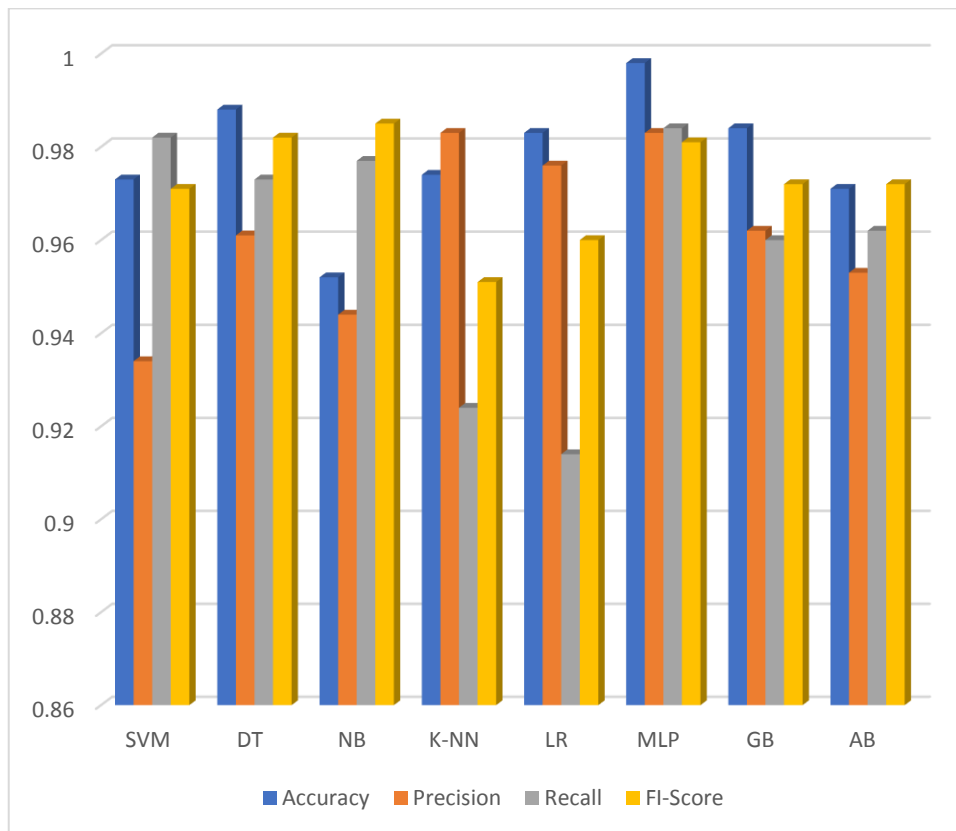


Figure 3. Classifier Performance of ML Techniques on the Testing Set.

As shown in Figure 3, it is clear that the MLP algorithm achieved the highest performance in terms of accuracy compared with other machine learning techniques. This is because the MLP does not lose any data. In contrast, the Naïve Bayes algorithm achieved the lowest performance in terms of accuracy.

5 Comparative Analysis of the proposed Model Classifier with existing models

This section will test the efficacy of the machine learning approaches described in this study by comparing their performance to that of existing techniques in order to identify which provides higher performance. Table 3 shows the results of the comparison.

Table 3: Comparison of the proposed ML Technique with existing work

Ref-N O	K-NN	DT	SVM	NB	LR	MLP	GB	AB
Rahman et al [22]	9708	9532	9708	9357	9825	9825	9532	0.9591
Assiri et al.[24]	0.9181	0.9649	0.9788	-----	0.9825	0.9181	----	-----
Islam et al.[25]	0.93.8	0.938	----	0.95.6	0.956	----	0.938	---
Srinivas et al.[26]	0.95	0.95	0.97	----	0.95	---	0.97	0.95
Our Proposed	0.983	0.992	0.984	0.964	0.983	0.994	0.983	0.984

Table 3. The proposed model technique outperforms existing studies regarding the accuracy of breast cancer. However, in the suggested model, the MLP model outperforms existing studies. The prediction accuracy of the MLP algorithm in this study achieved a rate of 0.994 for breast cancer. While the superior performance of accuracy from existing studies with MLP, which was collected by studies Rahman et al[22] and Islam et al [25], the prediction accuracy rate is 0.9825 and 0.95, respectively, for breast cancer. In addition, in our suggested model, the Decision Trees (DT) algorithm is the second-best performer for classification accuracy. However, the prediction accuracy of the DT algorithm in this study was achieved at a rate of 0.992. while the superior performance of accuracy from existing studies with the DT, which was collected by the study of Assiri et al.[23] , whereas the prediction accuracy rate is 0.9649 for breast cancer.

6. Conclusion

Breast cancer research is a significant issue, and technology is helping to lessen the number of people who die from it. Several studies on medical datasets have recently been published. Breast cancer data must be classified appropriately and effectively in medical diagnosis. Although multiple ways of recognizing breast cancer data have been devised, there are still significant obstacles to overcome, including accuracy. To solve this issue, we created an algorithm for classifying breast cancer. We used nine ML classification algorithms (MLP, AB, GB, DT, KNN, SVM, RF, NB, and LR), as well as three ensemble ML approaches. All these procedures were used on the WDBC dataset. Overall, the findings indicate that ensembling our proposed machine-learning approaches outperforms the existing machine-learning techniques in categorization.

REFERENCES

[1] T R.Mahesh, V Vinoth Kumar, V. Vivek, KM.Karthick Raghunath, G. Sindhu Madhuri, “ Early predictive model for breast cancer classification using blended ensemble learning”. International Journal of System Assurance Engineering and Management. vol26:1-0 , Jun2022 , doi.org/10.1007/s13198-022-01696-0 .

[2] X. Zhou et al., "A Comprehensive Review for Breast Histopathology Image Analysis Using Classical and Deep Neural Networks," in IEEE Access, vol. 8, pp. 90931-90956, 2020, doi: 10.1109/ACCESS.2020.2993788.

- [3] G D Rashmi, A. Lekha, and N. Bawane, "Analysis of efficiency of classification and prediction algorithms (Naïve Bayes) for Breast Cancer dataset," 2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, 2015, pp. 108-113, doi: 10.1109/ERECT.2015.7498997.
- [4] M. Deepika and K. Kalaiselvi, "An Empirical Study on Disease Diagnosis using Data Mining Techniques," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 2018, pp. 615-620, doi: 10.1109/ICICCT.2018.8473185.
- [5] D. S. Jacob, R. Viswan, V. Manju, L. PadmaSuresh, and S. Raj, "A Survey on Breast Cancer Prediction Using Data Mining Techniques," 2018 Conference on Emerging Devices and Smart Systems (ICEDSS), Tiruchengode, India, 2018, pp. 256-258, doi: 10.1109/ICEDSS.2018.8544268.
- [6] A. I. Pritom, M. A. R. Munshi, S. A. Sabab, and S. Shihab, "Predicting breast cancer recurrence using effective classification and feature selection technique," 2016 19th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2016, pp. 310-314, doi: 10.1109/ICCITECHN.2016.7860215.
- [7] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/caac 21492.
- [8] A. B. Nassif, A. Talib, Q. Nasir, Y. Afadar, and O. Elgendy, "Breast cancer detection using artificial intelligence techniques: A systematic literature review," Elsevier, 2022. [Online]. Available: <https://www.cancer.org/>
- [9] Y. Khourdifi and M. Bahaj, "Applying Best Machine Learning Algorithms for Breast Cancer Prediction and Classification," 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Morocco, 2018, pp. 1-5, doi: 10.1109/ICECOCS.2018.8610632.
- [10] Z. Nematzadeh, R. Ibrahim, and A. Selamat, « Comparative studies on breast cancer classifications with k-fold cross-validations using machine learning techniques », in Control Conference (ASCC), 2015 10th Asian, 2015, p. 1–6
- [11] S. Ara, A. Das, and A. Dey, "Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms," 2021 International Conference on Artificial Intelligence (ICAI), Islamabad, Pakistan, 2021, pp. 97-101, doi: 10.1109/ICAI52203.2021.9445249.
- [12] V. Kumar and A. Verma, "Prediction of Malignant & Benign Breast Cancer: A Data Mining Approach in Healthcare Applications." In *Advances in data science and management* (pp. 435-442). Springer, Singapore.
- [13] S. Laghmati, B. Cherradi, A. Tmiri, O. Daanouni, and S. Hamida, "Classification of Patients with Breast Cancer using Neighbourhood Component Analysis and Supervised Machine Learning Techniques," 2020 3rd International Conference on Advanced Communication

- Technologies and Networking (CommNet)*, Marrakech, Morocco, 2020, pp. 1-6, doi: 10.1109/CommNet49926.2020.9199633.
- [14] O. I. Obaid, M. A. Mohammed, M. K. Abd Ghani, S. A. Mostafa, and F. T. Al-Dhief, "Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer," *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 4.36 Special Issue 36, pp. 160–166, 2018, doi: 10.14419/ijet.v7i4.36.23737.
- [15] A. I. Pritom, M. A. R. Munshi, S. A. Sabab, and S. Shihab, "Predicting breast cancer recurrence using effective classification and feature selection technique," *2016 19th International Conference on Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, 2016, pp. 310-314, doi: 10.1109/ICCITECHN.2016.7860215.
- [16] R. Alyami *et al.*, "Investigating the effect of correlation-based feature selection on breast cancer diagnosis using artificial neural network and support vector machines," *2017 International Conference on Informatics, Health & Technology (ICIHT)*, Riyadh, Saudi Arabia, 2017, pp. 1-7, doi: 10.1109/ICIHT.2017.7899011.
- [17] A. Addeh, H. Demirel, and P. ZARBAKHSH, "Early detection of breast cancer using optimized ANFIS and feature selection," *2017 9th International Conference on Computational Intelligence and Communication Networks (CICN)*, Girne, Northern Cyprus, 2017, pp. 39-42, doi: 10.1109/CICN.2017.8319352.
- [18] B. M. Gayathri and C. P. Sumathi, "Mamdani fuzzy inference system for breast cancer risk detection," *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Madurai, India, 2015, pp. 1-6, doi: 10.1109/ICCIC.2015.7435670.
- [19] Rashmi G D, A. Lekha and N. Bawane, "Analysis of efficiency of classification and prediction algorithms (Naïve Bayes) for Breast Cancer dataset," *2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, Mandya, India, 2015, pp. 108-113, doi: 10.1109/ERECT.2015.7498997.
- [20] H. T. Tike Thein and K. M. Mo Tun, "An Approach for Breast Cancer Diagnosis Classification Using Neural Network," *Advanced Computing: An International Journal*, vol. 6, no. 1, pp. 1–11, Jan. 2015, doi: 10.5121/acij.2015.6101.
- [21] E Venkatesan and T. Velmurugan. "Performance analysis of decision tree algorithms for breast cancer classification." *Indian Journal of Science and Technology* **8(29)**, 1-8, 2015.
- [22] T. Srinivas et al., "Novel-Based Ensemble Machine Learning Classifiers for Detecting Breast Cancer," *Math. Probl. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/9619102.
- [23] P. Liu, B. Fu, S. X. Yang, L. Deng, X. Zhong, and H. Zheng, "Optimizing Survival Analysis of XGBoost for Ties to Predict Disease Progression of Breast Cancer," in *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 1, pp. 148-160, Jan. 2021, doi: 10.1109/TBME.2020.2993278.
- [24] A. S. Assiri, S. Nazir, and S. A. Velastin, "Breast Tumor Classification Using an Ensemble Machine Learning Method," *J. Imaging*, vol. 6, no. 6, 2020, doi: 10.3390/JIMAGING6060039

- [25] T. Islam, A. B. Akhi, F. Akter, M. N. Hasan, and M. A. Lata, "Prediction of Breast Cancer using Traditional and Ensemble Technique: A Machine Learning Approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, pp. 867–875, 2023, doi: 10.14569/IJACSA.2023.0140692
- [26] T. Srinivas et al., "Novel-Based Ensemble Machine Learning Classifiers For Detecting Breast Cancer," *Math. Probl. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/9619102.
- [27] V. Chaurasia and S. Pal. "A novel approach for breast cancer detection using data mining techniques." *International Journal of Innovative Research in Computer and Communication Engineering*2(1), 2456-2465, 2014
- [28] R. Sumbaly, N. Vishnusri, and S. Jeyalatha, "Diagnosis of Breast Cancer using Decision Tree Data Mining Technique," *International Journal of Computer Applications*, vol. 98, no. 10, pp. 16–24, Jul. 2014, doi: 10.5120/17219-7456.