

Breast Cancer Diagnosis Through Machine Learning Classification Algorithms: An Experimental Study

Tapas Roy

PhD Scholar, Department of Library and Information Science The University of Burdwan,
<https://orcid.org/0009-0003-0610-5057>, E-mail: tapasroy135013@gmail.com

Mithun Das

PhD Scholar, Department of Library and Information Science, The University of Burdwan,
<https://orcid.org/0009-0007-6482-6425>, E-Mail : 02mithun02@gmail.com

Dr. Swarnali Chatterjee

Assistant University Librarian, Department of Performing Arts and Music, University of Calcutta,
<https://orcid.org/0000-0001-5349-6589>, E-mail: separtm@caluniv.ac.in

Dr. Rajesh Das

Assistant Professor, Dept. of Library and Information Science, The University of Burdwan,
<http://orcid.org/0000-0001-5349-6589>, E-Mail: rajeshdas99@gmail.com

DOI : <https://doi.org/10.5281/zenodo.19414241>

ARTICLE DETAILS

Research Paper

Accepted: 18-03-2026

Published: 10-04-2026

Keywords:

Cancer, Breast cancer, Diagnosis, Machine learning, Algorithms, Logistic regression, Classification.

ABSTRACT

This proposed research study highlights a logical framework for breast cancer diagnosis using machine learning algorithms. In this study, we have evaluated seven supervised machine learning classification algorithms—Logistic Regression, Naïve Bayes, Stochastic Gradient Descent (SGD), Decision Tree (DT), Random, K-Nearest Neighbours (KNN), Forest (RF), and Support Vector Machine (SVM). We have been using the Wisconsin Breast Cancer Diagnostic dataset obtained from the UCI Machine Learning Repository. After that, we selected 10 diagnostic features from the dataset, comprising 569 cases (357 benign and 212 malignant). Cleaning, feature selection, scaling, and train-test splitting (31% test size) were all part of the data preprocessing. Accuracy, precision, recall, F1-score, and confusion matrices were used to assess



the model's performance. The Random Forest classifier, which balanced recall and precision for both benign and malignant classes, had the highest overall accuracy (90%) among the models that were assessed. KNN obtained 88% accuracy, whereas Naïve Bayes, Decision Tree, and Logistic Regression models showed competitive performance (89%). The SGD model has been shown lower recall for malignant cases (63%), indicating higher false-negative rates. The superior diagnostic reliability has been demonstrated by ensemble-based methods, specifically Random Forest. For clinical usefulness, lowering false-negative predictions is still essential.

Introduction

At present, the cancer is one of the most common disease in the world wide. Currently, at least 1 in Every 1000 people is affected by the cancer in a society. In 21st century breast cancer tumor is the most common disease of women. A malignant tumor that starts in breast tissue cells is called breast cancer. The World Health Organization (WHO) reports that a sizable portion of cancer morbidity and mortality among women worldwide is caused by breast cancer. Breast cancer is one of the most prevalent tumors that affect women worldwide. As more and more medical datasets become accessible, machine learning approaches are becoming increasingly powerful in aiding clinical decision-making. In this paper, the data used is the Wisconsin Breast Cancer Diagnostic data available in the UCI Machine Learning Repository. An end-to-end data processing pipeline is used to improve data quality and model performance. Exploratory data analysis is performed to understand the distribution of classes and characteristic behaviour. Several supervised machine learning algorithms are used to carry out binary classification of benign and malignant cases. The performance metrics used for the evaluation include accuracy, precision, recall, and F1-score, all standard metrics. Classification error analysis is performed using a confusion matrix. A comparison between models highlights their strengths and weaknesses. The research will establish the best model for reliable breast cancer classification.

Literature Review

Dutta & Khanna (2020) have discussed in their article how to compare SVM and Random Forest on the Wisconsin Breast Cancer Dataset (WBCD) and reported that SVM achieved higher accuracy.



Chaurasia and Pal (2020) have explored, in their research paper, the diagnosis and prediction of breast cancer using machine learning. This research study include to further advance this area by applying multiple base classifiers (CART, SVM, Naïve Bayes, KNN, Logistic Regression, and MLP) on the WDBC dataset, followed by ensemble techniques such as AdaBoost, Gradient Boosting, Random Forest, Extra Trees, Bagging, and XGBoost. They have also implemented a stacking classifier (Voting Classifier) to combine multiple learners at the meta-level. Their findings indicated that Logistic Regression achieved the highest basic classifier accuracy (approximately 99% AuROC). In comparison, ensemble methods such as Extra Trees and XGBoost achieved around 95.1739% accuracy on reduced feature subsets and 92.9824% in the stack classifier.

Hazra et al. (2016) examined and compared Naïve Bayes, SVM, and ensemble classifiers for breast cancer detection, reporting an accuracy exceeding 97.3978% with only five dominant features, and the time complexity of this algorithm is 0.102023 milliseconds, the least among the three classifiers. Ensemble methods such as Random Forests, AdaBoost, Gradient Boosting, and Extra Trees have consistently improved predictive performance by aggregating multiple weak learners. They also concluded that, for large datasets, how these classification algorithms behave is a future scope of this project.

Yesuf (2019) have discussed in his research paper the breast cancer detection of women using machine learning techniques. In this research study, he compared different machine learning algorithms e.g. including Support Vector Machine (SVM), Artificial Neural Network (ANN), Bayesian Network (NB), and K-Nearest Neighbours, on the Wisconsin breast cancer datasets from the UCI machine learning repository. He also concluded that the experimental results for breast cancer detection using a support vector machine achieved accuracies of 97.6% to 98.8%. The proposed system has significantly improved the accuracy of the diagnostic process. In this regard, he also recommended the growth of identification of breast cancer type by exploring colour and shape features to enhance breast cancer detection.

Jaison et al. (2020) described a machine learning-based method for categorizing breast cancer tumors as benign or malignant. Recognizing that early identification greatly increases patient survival. The study's goal is to create a predictive model that will help doctors diagnose breast cancer more precisely and effectively. They examine, in their paper, five supervised learning models—Gaussian Naïve Bayes, Random Forest, KNN, SVM, and Multilayer Perceptron (MLP)—on a dataset of 683 instances. The study reported that the MLP achieved the highest accuracy (97.08%), slightly outperforming SVM and other classifiers.



According to Ramana, Babu and Venkateswarlu (2011), machine learning techniques have been a widely applied in medical diagnosis, particularly for liver disease detection, where early identification is critical. In this research paper, published in the International Journal of Database Management Systems, evaluates supervised algorithms including Naïve Bayes, C4.5, Back Propagation, K-Nearest Neighbour (KNN), and Support Vector Machines (SVM) using liver patient datasets. They also performed the study, which confirms that classification accuracy improves with relevant biochemical attributes such as bilirubin, SGOT, and SGPT. Comparative results show KNN, Back Propagation, and SVM often outperform others, highlighting the importance of feature selection and dataset quality in predictive medical modelling.

Statement of the problem

Breast cancer diagnosis based on imaging, pathology reports, or clinical features remains time-consuming and subject to human error. Traditional diagnostic methods require extensive expert intervention and can vary based on the expertise level. The research problem is:

QR. Can machine learning classification algorithms accurately distinguish between benign and malignant breast cancer cases using clinical and diagnostic features?

QR. How do different classification algorithms compare with performance metrics such as accuracy, precision, recall, specificity, and F1-score?

QR. How effective are machine learning classification methods in minimizing false-positive and false-negative diagnoses?

Objectives

The primary objectives of this research study are:

1. To diagnosis malignant and benign for breast cancer through ML algorithms.
2. To predict the breast cancer from unknown data.
3. To comparative analysis of different machine learning (ML) algorithms for finding maximum accuracy score.
4. To perform the comparison and evaluation of high accuracy, reliable in benign and malignant patients for breast cancer diagnosis.



Methodology

This research study proposed an experimental design to compare the performance of various machine learning classification algorithms for breast cancer diagnosis. Firstly, we have collected multiple data on breast cancer, which is called benign and malignant, from many female patients. The experiment involves systematic data preprocessing, model training, testing, and performance evaluation. We have drawn a suitable framework from the breast cancer dataset for step-by-step work (see Figure 1).

Data Collection

To conduct this proposed research work, we downloaded a 50.1 KB dataset from the UCI machine learning repository: the breast cancer Wisconsin (diagnostic) dataset (<https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>). The dataset was multivariate, the subject area was health and medicine, the associated task was classification, the feature type was real-valued, and the number of instances was 569. The number of feature columns was 32, of these instances 569 and the 10 columns were selected for the research work.

Table- 1: Description of the ten variables

No.	Variable	Description
1	Radius	Radius is the average distance from the centre of the object (cell or tumour) to its boundary. It represents the object's overall size.
2	Texture	Texture measures variations in pixel grey levels, indicating surface roughness or uniformity.
3	Perimeter	Perimeter is the sum of the total length of the boundary of the object. It reflects boundary complexity.
4	Area	Area is the total number of pixels contained within the object and represents its size.
5	Smoothness	Smoothness measures the local variation in radius lengths. Higher variation indicates a rough boundary.
6	Compactness	Compactness indicates the degree of closeness or density of an object's shape. Formula (conceptual): $\text{Compactness} = (\text{Perimeter}^2 / \text{Area})$. Higher compactness indicates a more irregular shape.
7	Concavity	Concavity measures the depth of indentations along the boundary of the



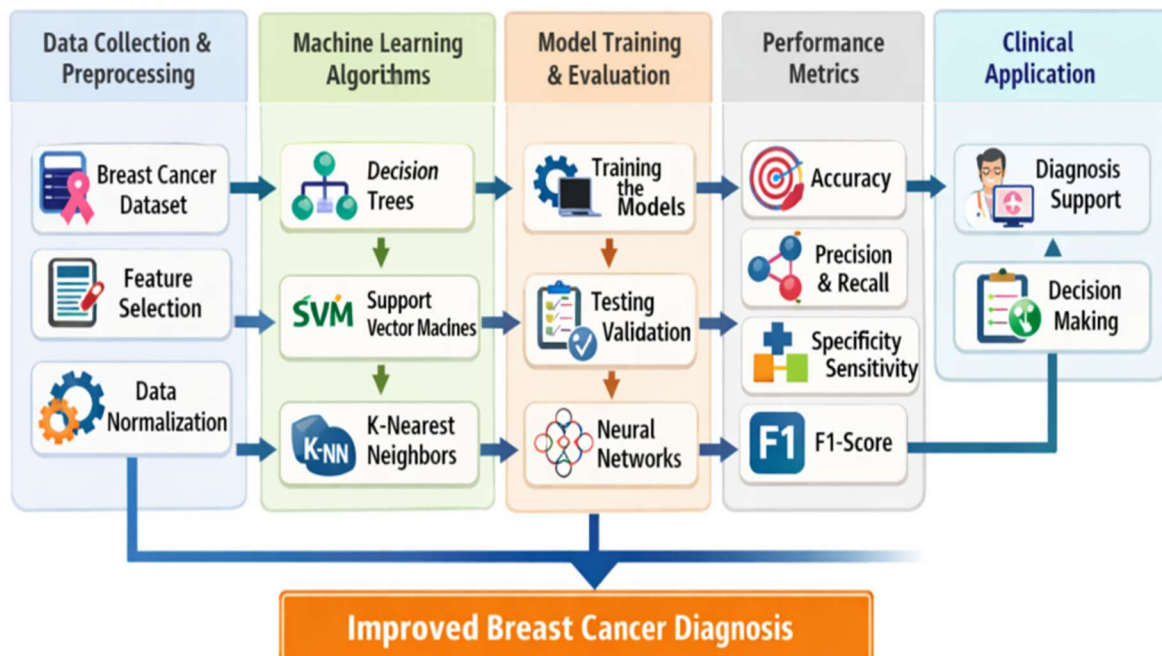
No.	Variable	Description
		object.
8	Concave Points	Concave points are the number of inward curves along the boundary. Each inward dent counts as 1 point. More concave points indicate a less regular margin.
9	Symmetry	Symmetry measures the similarity between two halves of the object. High symmetry → regular shape Low symmetry → distorted, irregular shape
10	Fractal Dimension	Fractal dimension measures the complexity of the boundary. Values typically range from 1 to 2. Higher values indicate more complex and jagged edges.

Data-preprocessing

Machine learning involves data preprocessing, the preparation of raw data. It includes data cleansing, handling missing data, eliminating noise and outliers, feature scaling, and assigning categorical data to numerical values. Learning is further enhanced with the feature selection and data balancing. When properly preprocessed, machine learning models are more accurate, reliable, and efficient.

Figure-1: A logical framework of breast cancer diagnosis classification algorithms through a machine learning model.

Table-1: Different machine learning algorithms



Algorithms	Strength
Logistic Regression	Baseline performance and high interpretability.
Random Forest	Handles non-linear data well and reduces overfitting.
Support Vector Machines (SVM)	Effective in high-dimensional spaces.
K-Nearest Neighbors (KNN)	Simple, instance-based learning for similarity.
Neural Networks (MLP)	Capable of capturing complex, hidden patterns.

Exploratory data analysis (EDA)

To determine the number of benign and malignant cases in the data, a chi-squared test was used to analyse the class distribution of the diagnosis variable. As shown, the dataset contains 357 benign (B) samples and 212 malignant (M) samples, resulting in a relatively balanced class distribution. To provide more insight into this distribution, the visual analysis used a method that provides a straightforward graphical comparison of the two classes. Such a visualisation quickly identifies the dominance of classes and the representation of minorities. Past studies have demonstrated that class imbalance can negatively affect the performance of machine learning models, leading to biased predictions. Thus, before model training, it is necessary to comprehend the distribution of diagnosis classes. The analysis allows making informed choices about data preprocessing techniques, such as resampling and class weighting. Altogether, the class distribution analysis is an essential preliminary step in the research on breast cancer diagnosis using machine learning techniques.

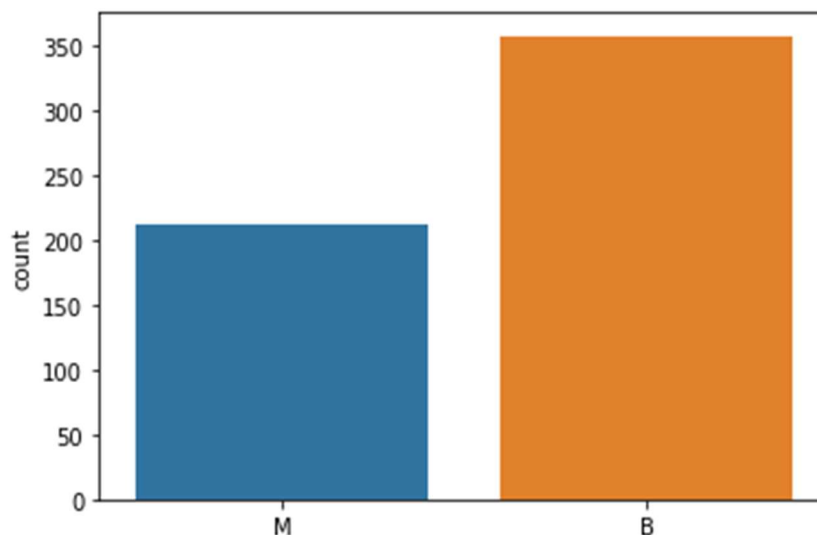


Figure- 2: Value of count "Benign" and "Malignant"



Splitting of dataset into Traing and test data

The purpose is to split the dataset into training and testing sets. Such a separation enables the model to be trained on a subset of the data and tested on previously unseen samples. It helps measure the model's generalisation capability and reduces overfitting. Correct data splitting will provide a fair and reliable system for assessing model performance.

Training and testing data

The data is split into training and test sets at a 31% test size, ensuring sufficient data to evaluate the model. The rest is devoted to the training of the machine learning model. Fixing a random state ensures reproducibility of results. This is a data partitioning method that helps perform a sound evaluation of the model's generalisation.

Table-2: Classification of datasets of benign and malignant in breast cancer diagnosis through different ML algorithms

Sl. No.	Classifier	Precision		Recall		Accuracy (f1- score)
		Benign	Malignant	Benign	Malignant	
1	Logistic Regression	0.89	0.88	0.93	0.82	0.89
2	Naïve Bayes	0.89	0.88	0.93	0.82	0.89
3	SGD	0.81	0.93	0.97	0.63	0.84
4	KNN	0.85	0.94	0.97	0.72	0.88
5	Decision Tree	0.89	0.88	0.93	0.82	0.89
6	Random Forest	0.90	0.89	0.94	0.84	0.90
7	SVM	0.89	0.82	0.89	0.82	0.86

Figure-3: Classifier Performance Distribution based on Precision (Benign)

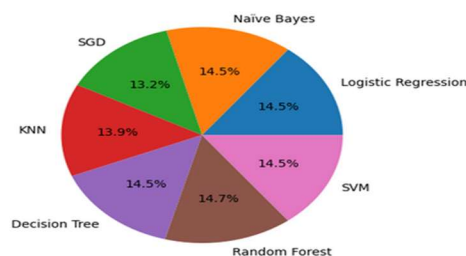


Figure-4: Classifier Performance Distribution based on Accuracy/F1 Scores

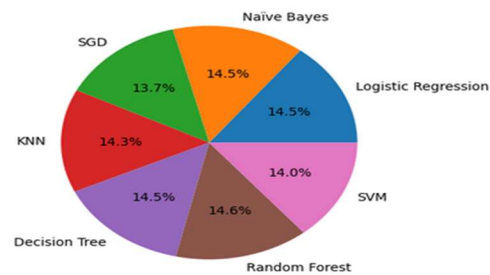


Figure-5: Classifier Performance Distribution based on Recall (Malignant)

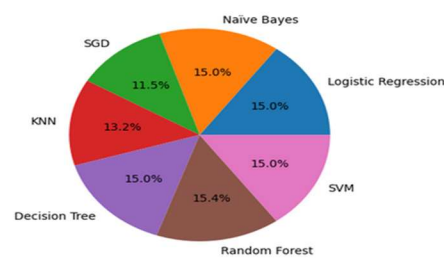
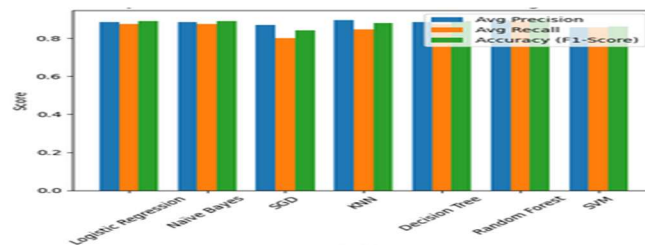


Figure-6: Comparative Performance of Machine Learning Classifier



Classification Report of logistic regression

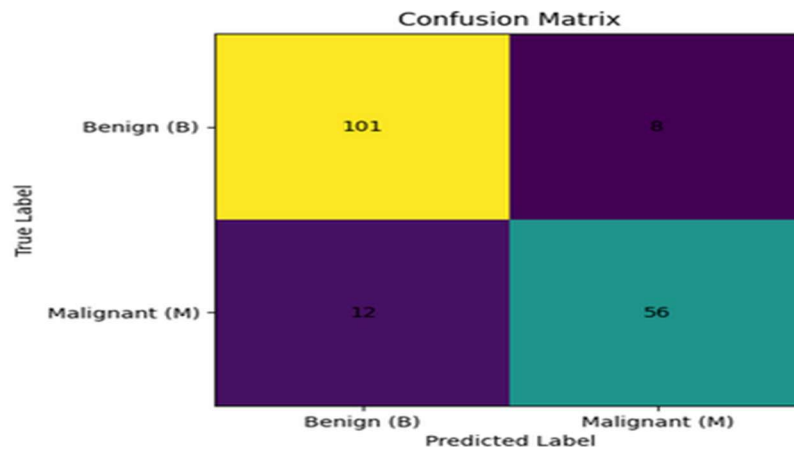
The classification report assesses the model's accuracy, precision, recall, and F1-score. The accuracy of the predictions is 0.89 (89), indicating that the majority are accurate. The model's high precision indicates a low false-positive rate. Recall values indicate the model's ability to identify malignant cases correctly. The F1-score validates an equal trade-off between recall and precision.

Confusion matrix of logistic regression

The confusion matrix provides an in-depth evaluation of false and true classifications. The model correctly classifies 101 benign and 56 malignant cases, indicating strong performance. The misclassifications are 8 false positives and 12 false negatives, which are not that big. False negatives

have important implications for medical diagnosis and must be given due attention. In general, the confusion matrix demonstrates the model's usefulness for breast cancer classification.

Figure-7: Confusion matrix of logistic regression



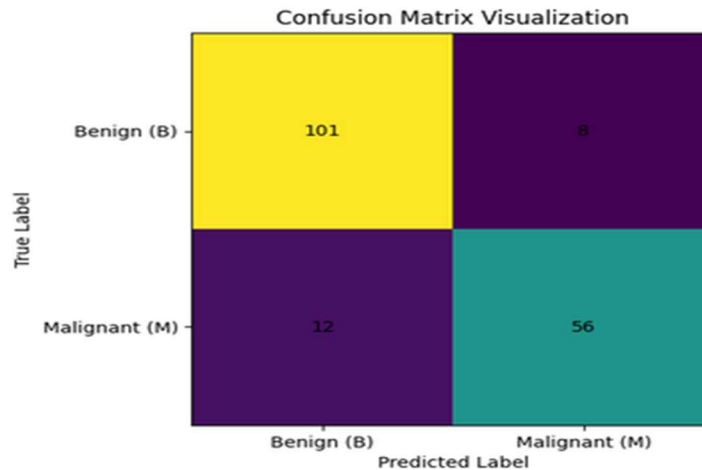
Naive Bayes classification report

The classification report shows good, balanced performance across both classes, with an accuracy of 0.89, indicating that the model generalises successfully to unseen data. In the case of the benign (B) class, the high precision (0.89) and recall (0.93) indicate that the model can recognise benign cases and also reduce false negatives. The malignant (M) category has a precision of 0.88 and a recall of 0.82, indicating a slightly lower sensitivity, a characteristic of imbalanced medical data. B and MF1-scores of 0.91 (B) and 0.85 (M) have established a good trade-off between precision and recall in both categories. The fact that the macro and weighted averages (0.88–0.89) are similar and do not favour one of the classes justifies the strength of the classifier.

Confusion matrix of Naive Bayes

According to the confusion matrix, the true negative and actual positive values are high, with 101 benign samples and 56 malignant samples correctly identified. The false-positive rate was low, with only 8 of the benign cases mistakenly diagnosed as malignant, which is essential for minimising unwarranted clinical interventions. The model misclassified 12 malignant cases as benign, an area that needs improvement in terms of clinical risk factors for false negatives. On the whole, the matrix indicates an appropriate balance between sensitivity and specificity, which corresponds to the accuracy and F1 Scores. The graphical representation of the predictions confirms that misclassifications are minimal and that the model does not have unstable decision boundaries.

Figure- 8: Confusion matrix of naive Bayes



Stochastic gradient descent classification report

The classification report shows an overall accuracy of 0.84, indicating reasonable predictive performance for the SGD-trained model. In the benign (B) class, the model has a high recall (0.97). The malignant (M) group, however, has a lower recall (0.63), indicating that a significant proportion of malignant cases are misclassified. The malignant class has a relatively high precision (0.93), suggesting that when the model predicts a malignant case, it tends to be accurate. The F1-score weighted (0.83) proves the existence of a moderate precision and recall of both classes.

Confusion matrix of stochastic gradient descent

The confusion matrix indicates that 106 cases were correctly classified and 43 cases were correctly identified as benign or malignant. And there are only 3 false-positive cases that are misclassified as malignant cases. But 25 malignant cases are misdiagnosed as benign, a critical error in medical diagnosis. This imbalance shows that the model is biased toward predicting the benign class with high confidence. To enhance clinical reliability, false negatives in cases of malignancy should be reduced.

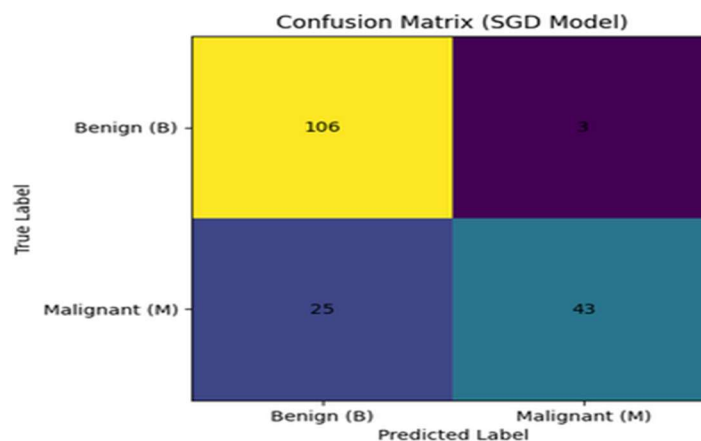


Figure- 9: confusion matrix of stochastic gradient descent

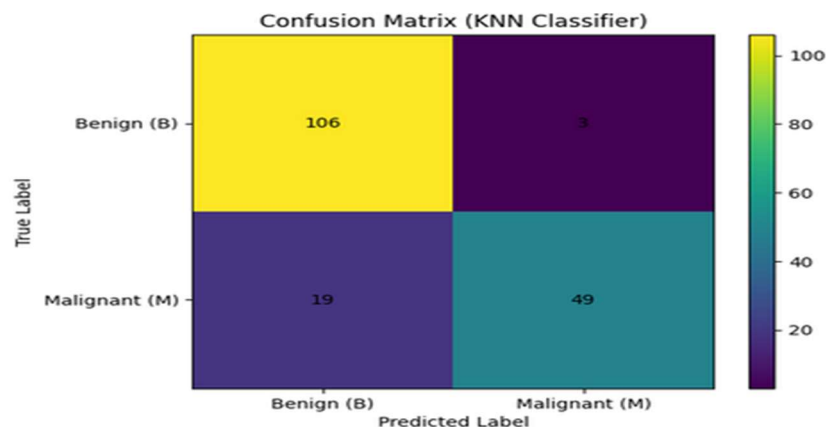
On K-Nearest Neighbours Classification report

The overall accuracy of the KNN classifier was 88%, indicating that it can accurately predict on the test data. With the Benign (B) model, the recall (0.97) and the F1-score (0.91) are high, indicating the ability to identify non-malignant cases. The Malignant (M) category has high precision (0.94), suggesting that it is a good predictor, with most predictions being correct. The recall of the malignant class is, however, relatively lower (0.72), meaning that not all malignant cases have been identified. The overall weighted average F1-score of 0.87 demonstrates balanced performance between the two classes, although there is variation between them.

Confusion matrix of K-Nearest Neighbours

The confusion matrix indicates that 106 Benign cases were correctly identified, and only 3 were misclassified as Malignant. In the Malignant category, 49 instances are correctly predicted, and 19 are incorrectly predicted as Benign. The prevailing diagonal figures indicate that most predictions align with the actual class occurrences. Misclassification of malignant samples is more likely, which is essential in a medical diagnosis setting. These findings indicate that although KNN is a good concept in general, it is necessary to increase malignant recall to make the process clinically reliable.

Figure- 10: Confusion Matrix of the K-Nearest Neighbours (KNN) Classifier



Decision Tree Model Classification Report:

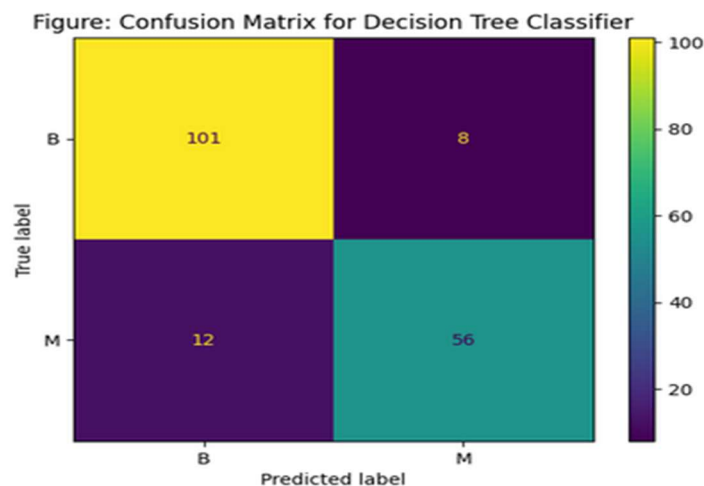
According to the classification report, the Decision Tree model has a high predictive performance with a general accuracy of 0.89. In the benign category (B), the model has high precision (0.89) and recall

(0.93), so it is effective in identifying non-malignant issues. The least recall (0.82) is exhibited by the malignant class (M), signifying that a few malignant samples were missed. An F1-score of 0.88, when averaged across the two classes, indicates balanced performance. All in all, the weighted-average scores validate the model's performance on the skewed dataset.

Decision Tree Model Confusion matrix

The confusion matrix also describes the model's classification behaviour. There were 101 benign samples and 56 malignant samples, and this is the 151st correct classification in the diagonals. But 8 benign samples were mistakenly labelled as malignant, and 12 malignant samples were mistakenly labelled as benign. The false-negative rate in the malignant cases indicates a moderate limitation in sensitivity. Notwithstanding this, the matrix indicates reliable discrimination between the two classes.

Figure- 11: Confusion Matrix for Decision Tree Classifier



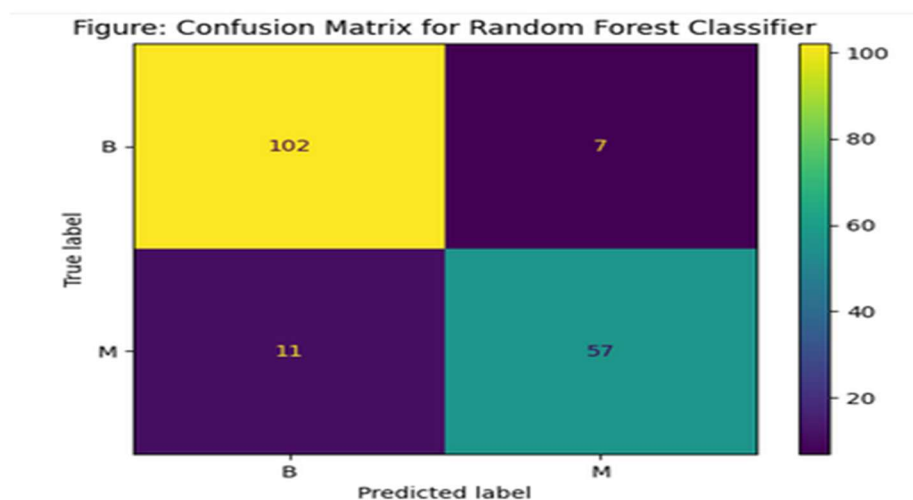
The classification model is a random forest

According to the classification report, the overall accuracy of the Random Forest model is 0.90, indicating strong predictive power. The benign B category has a high precision (0.90) and recall (0.94), and it identifies the non-malignant samples with great accuracy. For the malignant (M) class, the model has a precision and recall of 0.89 and 0.84, respectively, which are better than those of the simple models. The balanced performance across both classes is confirmed by the macro-averaged F1-score of 0.89. In general, the weighted averages highlight the ensemble method's strength on the dataset.

Random Forest Model confusion matrix

The confusion matrix provides an overview of how the model has been performing. There are also 102 benign and 57 malignant samples that are properly classified, resulting in good diagnostic dominance. There are 7 benign samples misclassified as malignant and 11 malignant samples misclassified as benign. The low rate of false negatives in malignant cases indicates better diagnostic reliability. This supports the claim that the Random Forest model is suitable for minimising classification errors through ensemble learning.

Figure- 12: Confusion Matrix for Random Forest Classifier



The Classification Report of the SVM Model

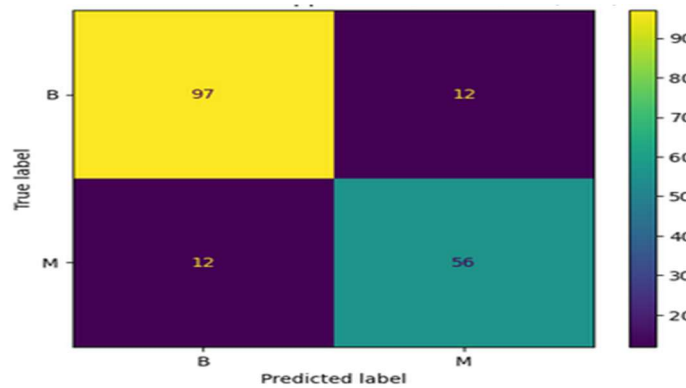
According to the classification report, the Support Vector Machine (SVM) model achieves a total accuracy of 0.86, indicating that it is stable and moderately accurate. In the benign class (B), the model consistently identifies benign samples, with precision, recall, and F1-score all at 0.89, which represent balanced values. The malignant type (M) has slightly lower scores (precision, recall, and F1-score of 0.82), indicating that it is less receptive to malignant cases. The macro-analysed scores show consistent performance among classes. Overall, the weighted averages indicate that the SVM has acceptable classification performance, but it can be further refined.

Confusion Matrix of SVM Model

The confusion table shows the finer details of the SVM classifier's prediction behaviour. The total number of correctly classified samples is 97 benign and 56 malignant, which constitute the diagonal elements. But 12 benign samples will still be misclassified as malignant, and 12 malignant samples will

be misclassified as benign. The fact that both false positives and false negatives are equal points to a balanced distribution of errors. Despite these misclassifications, the model continues to exhibit acceptable discrimination between the two classes.

Figure- 13: Confusion Matrix for Support Vector Machine (SVM) Classifier



Conclusion

This paper is a comparative study of seven machine learning methods used for breast cancer diagnosis, detailing the applications of each. Optimal data preprocessing also contributed to a high degree of reliability and efficiency of the models. The exploratory analysis found that a manageable proportion of classes had no extreme imbalance. Random Forest and other ensemble-based models, including Hungarian Forest and very random forests, were the most effective overall. Competitive accuracies were also achieved, with moderate misclassification rates, in tree-based and instance-based models. The models based on linear and margin features demonstrated consistent but relatively lower sensitivity to malignant cases. The analysis of confusion matrices showed that false negatives are a significant issue in medical diagnosis. Such errors need to be reduced to enhance clinical safety. Overall, the research indicates that machine learning is effective for medical classification. Future research can focus on more sophisticated feature engineering and deep learning methods.

References

- Abunasser, B. S., Al-Hiealy, M. R. J., Zaqout, I. S., & Abu-Naser, S. S. (2023). Literature review of breast cancer detection using machine learning algorithms. *AIP Conference Proceedings*, 2808, 040006. <https://doi.org/10.1063/5.0133688>



- Chaurasia, V., & Pal, S. (2020). Applications of machine learning techniques to predict diagnostic breast cancer. *SN Computer Science*, 1(5). <https://doi.org/10.1007/s42979-020-00296-8>
- Hamed, S., Mesleh, A., & Arabiyyat, A. (2021). Breast cancer detection using machine learning algorithms. *International Journal of Computer Science and Mobile Computing*, 10(11), 4–11. <https://doi.org/10.47760/ijcsmc.2021.v10i11.002>
- Hassan, M. M., Yasmin, F., Khan, M. A. R., Zaman, S., Galibuzzaman, Islam, K. K., & Bairagi, A. K. (2023). A comparative assessment of machine learning algorithms with the least absolute shrinkage and selection operator for breast cancer detection and prediction. *Decision Analytics Journal*, 7, 100245. <https://doi.org/10.1016/j.dajour.2023.100245>
- Hazra, A., Kumar, S., & Gupta, A. (2016). Study and analysis of breast cancer cell detection using naïve Bayes, SVM and ensemble algorithms. *International Journal of Computer Applications*, 145(2), 39–45. <https://doi.org/10.5120/ijca2016910595>
- Hussain, L., et al. (2018). Automated breast cancer detection using machine learning techniques by extracting different feature extracting strategies. In *Proceedings of IEEE* (pp. 327–331). <https://doi.org/10.1109/trustcom/bigdatase.2018.00057>
- Islam, M. M., Haque, M. R., Iqbal, H., Hasan, M. M., Hasan, M., & Kabir, M. N. (2020). Breast cancer prediction: A comparative study using machine learning techniques. *SN Computer Science*, 1(5). <https://doi.org/10.1007/s42979-020-00305-w>
- Jain, B., & Singla, N. (2023). Breast cancer detection using machine learning algorithms. *Journal of Computers, Mechanical and Management*, 2(6), 30–35. <https://doi.org/10.57159/gadl.jcmm.2.6.230109>
- Khourdifi, Y., & Bahaj, M. (2018). Applying best machine learning algorithms for breast cancer prediction and classification. In *Proceedings of IEEE* (pp. 1–5). <https://doi.org/10.1109/icecocs.2018.8610632>
- Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2014). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13, 8–17. <https://doi.org/10.1016/j.csbj.2014.11.005>



- Naji, M. A., Filali, S. E., Aarika, K., Benlahmar, E. H., Abdelouhahid, R. A., & Debauche, O. (2021). Machine learning algorithms for breast cancer prediction and diagnosis. *Procedia Computer Science*, 191, 487–492. <https://doi.org/10.1016/j.procs.2021.07.062>
- Nemade, V., & Fegade, V. (2023). Machine learning techniques for breast cancer prediction. *Procedia Computer Science*, 218, 1314–1320. <https://doi.org/10.1016/j.procs.2023.01.110>
- Omondiagbe, D. A., Veeramani, S., & Sidhu, A. S. (2019). Machine learning classification techniques for breast cancer diagnosis. *IOP Conference Series: Materials Science and Engineering*, 495, 012033. <https://doi.org/10.1088/1757-899X/495/1/012033>
- Ramana, B. V., Prasad Babu, M. S., et al. (2011). A critical study of selected classification algorithms for liver disease diagnosis. *International Journal of Database Management Systems*, 3(2), 101–114. <https://doi.org/10.5121/ijdms.2011.3207>
- Sharma, S., Aggarwal, A., & Choudhury, T. (2018). Breast cancer detection using machine learning algorithms. In *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)* (pp. 114–118). <https://doi.org/10.1109/ctems.2018.8769187>
- Sherina Sara, J., Nayana, R., Mounusha, S., & Mallikarjuna, K. (2020). Machine learning techniques to detect breast cancer. *International Journal of Advanced Research in Computer Science*, 11(1), 58–62.
- Sinha, N., Sharma, P., & Arora, D. (2020). Prediction model for breast cancer detection using machine learning algorithms. In *Advances in Intelligent Systems and Computing* (pp. 431–440). https://doi.org/10.1007/978-981-15-6876-3_33
- Tahmooresi, M., Afshar, A., Rad, B. B., Nowshath, K. B., & Bamiah, M. A. (2018). Early detection of breast cancer using machine learning techniques. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 10, 21–27. <http://journal.utem.edu.my/index.php/jtec/article/view/4706/3462>
- Telsang, V. A., & Hegde, K. (2020). Breast cancer prediction analysis using machine learning algorithms. In *Proceedings of IEEE* (pp. 1–5). <https://doi.org/10.1109/c2i451079.2020.9368911>



- Yesuf, S. H. (2019). Breast cancer detection using machine learning techniques. *International Journal of Advanced Research in Computer Science*, 10(5), 27–33.
<https://doi.org/10.26483/ijarcs.v10i5.6464>