SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Enhancing Breast Cancer Diagnosis Interpretability: Logistic Regression Performance on the Wisconsin Dataset

Khamrunissa Hussain Sheikh [1], Sudheer Kumar Sharma[2], Narendra Kumar[3]

[1]Research Scholar, Department of Mathematics & Statistics, NIET, NIMS University Rajasthan, Jaipur INDIA.
E-mail: khamarnooh2021@gmail.com,
[2]Department of Mathematics & Statistics, NIET, NIMS University Rajasthan, Jaipur INDIA. E-mail: sudheersharma.math@gmail.com
[3]Department of Mathematics & Statistics, IILM University Gr. Noida (U.P.) INDIA. E-mail: drnk.cse@gmail.com

Abstract
Breast cancer (BC) is one of the deadliest diseases in the world, necessitating the development of high-quality diagnostic systems. Therefore, the present research concentrates on examining the efficiency and reliability of Logistic Regression (LR) in terms of results about clinically important features such as specificity and sensitivity relying on the Wisconsin Diagnostic Breast Cancer (WDBC) Database. This set of open-access data includes 569 cases characterized by 30 cytomorphological features. The classification system was established by conducting a careful experimental design with 50 independent trials and a set of random samples taken during the training and testing procedure. Data preparation included feature normalization in accordance with the principle of zero mean and unit variance, while feature selection and transformation techniques were not employed to make the findings more interpretable. The performance of the models included the measurement of their efficiency, including accuracy, sensitivity, Area Under the Receiver Operating Characteristic Curve (AUROC) and specificity. The Logistic Regression methodology achieved an accuracy of 93.8%.
It is important to note that the model recorded a specificity level of 91.2% for benign cases, while the sensitivity for malignant cases was as high as 97.2%. The performance of the model was further confirmed by the AUROC value of 0.9452 obtained from Receiver Operating Characteristic (ROC) curve analysis. In addition, the results of the confusion matrix analysis indicate that the number of false negatives is confined to six cases for a total of 212 malignant cases and that false positives are equal to 29 of 357 benign cases, giving the impression that the model is conservative and favors sensitivity.
In addition, 50 repetitions of the run produced very stable outcomes of under 1%. All of this suggests that LR is a sensitive, robust, and explicable baseline model for BC diagnosis, provided that it is implemented within a proper evaluation framework and with consideration for clinical settings. Due to its proven reliability, the model that has been tested and validated is well-suited to assist in clinical decision-making.

Keywords: Logistic Regression; Sensitivity and Specificity; Breast Cancer; Wisconsin Dataset; Machine Learning Interpretability.

Introduction

BC continues to be a worldwide health concern, highlighting the significance of diagnosis methods capable of providing correct diagnoses and effective medical help [1, 2]. However, obtaining real-time data on cancer presence is challenging. Therefore, the WDBC dataset is a valuable resource. The set consisted of 569 cases collected through fine-needle aspiration and included 30 quantitative cytology features that enabled reliable comparisons of different modeling approaches [3, 4]002E
The WDBC dataset contains useful data for BC detection using various Machine Learning (ML) techniques. It consists of 569 instances of fine-needle aspiration biopsy samples characterized by 30 quantitative features that describe the morphological properties of cell nuclei, including radius, texture, and area [3, 4]. The dataset is helpful in undergoing analysis because of its important characteristics, including multicollinearity, preprocessing needs, and Modeling approaches such as LR, Support Vector Machines (SVM), Gradient-Boosted Trees , Random Forests, , K-Nearest Neighbors (KNN). The comparative study of the ML methods on WDBC will focus on optimizing the best possible balance between the the model's prediction accuracy and the understandability of its results. Thus, the WDBC dataset is a highly valuable source for researchers aiming to create and test programs for BC classification [5, 8].
Several studies have reported high levels of accuracy and AUROC values that are not far from reaching the limits of their theoretical capabilities via cross-validation [9, 10]. However, many of these studies depend on using either single k-fold approaches for model selection and evaluation, the performance of feature selection outside the validation loop, or the absence of appropriate thresholding procedures, which may lead to overly optimistic results [9, 11]. Despite the great interest in good performance estimation, the stratified inner-outer evaluation

method has been employed by only a limited number of studies [12, 13, 23], which speaks for the scarce practice of unbiased model evaluation with nested cross-validation.

Furthermore, there is little reporting of metrics relevant to probability calibration, such as expected calibration error and the Brier score, even though they play an important role in setting thresholds for clinical decisions and stratifying risk based on known probabilities [9, 14]. In addition, most explainability methods, from basic correlation heatmaps and surrogate trees to newer ones such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), do not have a systematic way to address attribution bias resulting from correlation or check stability at the level of folding [15, 16]. LR is considered a viable option mainly due to its features viz interpretation, standard practice of regularization, and being one of the best performing algorithms when using basic pre-processing [11, 12].

When evaluating LR performance, assessments focusing on important clinical indicators, such as sensitivity, specificity, and AUROC, are not always reported in full, even though uncertainty has been well accounted for using various resampling techniques [8, 9]. This is unexpected, considering its importance in the clinical field. The aim of this research was to thoroughly assess the performance of LR by examining the WDBC benchmark dataset. The aim was to quantify the trade-off between the level of discriminatory ability and the error rate through repeated trials.

This study identifies significant shortcomings in the literature. This study presents accurate and methodologically reliable baseline metrics, such as AUROC, sensitivity, specificity, and error characteristics, based on confusion matrices. To preserve the validity of our results, we used a leakproof framework. This study aims to place LR results in the current context of development and address the contemporary problems of validation and reporting methods.

Methods

Description of the Dataset

In this study, we used the WDBC Dataset. This dataset is available in the ML Repository at the University of California, Irvine (UCI) [17]. This dataset contains 569 clinical instances with confirmed classifications of benign or malignant BC.

For each case, we attempted to extract 30 measurable features based on digital images of breast tumors taken using fine needle aspiration (FNA). The 30 features can be further classified into 10 major morphological features of a cell nucleus: radius, texture (i.e., the standard deviation of gray values), area, perimeter, smoothness, concavity, compactness, number of concave points, fractal dimension and symmetry, for each entry, this study also generated complete statistics, which included three statistics for each feature.

In dataset, 212 instances of malignant cases, which represent 37.3% of the entire dataset, are included, whereas 357 benign instances account for 62.7%. Thus, the distribution of cases is consistent with the characteristic pattern of cases perceived in practice. A classifier in ML can be defined as a method that processes data and assigns class labels to different classes of data during classification. Once trained on the training dataset using a classifier, new data can be assigned class labels based on the developed model.

Using LR

The first classifier chosen for use was LR, and this was because of its simplicity as well as its effectiveness in dealing with medical problems that have two possible results. In LR, a sigmoid activation function is employed to indicate the likelihood of an input sample x being classified as either malignant or benign. The mathematical definition of this is given in P (x)=1/(1+e^{-x}), where P(x) is the probability, and e is the number 2.718 (Euler's number). The decision threshold was set at 0.5, which means that samples that have a probability of 0.5 and above were labeled malignant, while samples that had lower probabilities than this threshold were defined as benign. When constructing the model LR is being used as an algorithm to adjust the parameters through minimizing the log-loss function that punishes larger misclassifications more severely than smaller ones. Finally, the constructed model produces labels and probabilities that will help in making clinical decisions.

Experimental Configuration

The statistical and ML toolboxes in Python were employed in this study to carry out all experiments. To minimize the stochastic nature of data division that may induce variability in the results, the information was divided into training and validation datasets, followed by evaluating the LR model for a period of 50 rounds. Each round did consist of training and testing of the system using one dataset. In order to determine the efficacy of the system, three important parameters were defined:

- **Accuracy:** Proportion of correctly classified data.
- **Sensitivity (Recall):** Proportion of successfully identified malignant cases.
- **Specificity:** Proportion of successfully identified benign cases.

Furthermore, a Receiver Operating Characteristic (ROC) analysis was performed to learn more about the trade-off between specificity and sensitivity for different threshold classification decisions, and an Area Under Curve (AUC) measure was found.

Specifics of the Implementation

The first step of the logistic regression model development was taken with the default parameters available in the ML libraries of Python. Nonetheless, the popularization of the model could not occur without applying scaling of features so that the mean of each feature is zero and the variance equals one. Normalizing the features is important for the interpretability of the LR model because the coefficients of the variables can be compared with each other. It should be mentioned that no dimensionality reduction or feature selection methods were used in the initial stage of the model development. The rationale for this was to retain every feature in the model so that the integrity of the dataset and results would remain.

The computational experiments were performed on a powerful workstation with an Intel® Core™ i7 CPU. Owing to the strong processing power delivered by the CPU, the LR model performed well on this workstation. The system possesses 16 GB memory, which allows it to work efficiently with the data. The hardware configuration allowed us to conduct many experiments and evaluate the model performance.

Results

In total, the LR model conducted over 50 experiments and showed outstanding performance in diagnosis when applied to the WDBC Dataset. Table 1 presents the typical performance indicators of the classification model applied to the BC dataset.

Table 1: Logistic Regression Classifier performance

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
LR	93.8	97.2	91.2

A total of 534 well-classified samples out of 569, including 206 malignant cases out of 212 malignant cases, were classified with 97.2% sensitivity. Of the 357 benign cases, the specificity of the model was 91.2%, with 328 benign cases classified correctly. From these figures, it can be concluded that logistic regression is a highly efficient method for discriminating between malignant and benign samples.

Evaluation of the Confusion Matrix

Figure 1 includes a confusion matrix that also illustrates the model's performance. The error rate was quite low, and only six malignant and 29 benign cases were misclassified. The importance of the confusion matrix in medical diagnosis has been established because misdiagnosis of malignant diseases can have severe effects.

Characteristic (ROC) Curve

The efficacy of the LR model on the basis of discrimination was evaluated through plotting ROC curve. The evaluation of this curve is displayed in Figure 2. The structure of this curve gives a representation of the relationship that exists between specificity and sensitivity across different thresholds.

The model achieved an AUC score of 0.9452, demonstrating outstanding performance in differentiating between malignant and benign cases. Thus, a high AUC value indicates that the model operates efficiently at many thresholds.

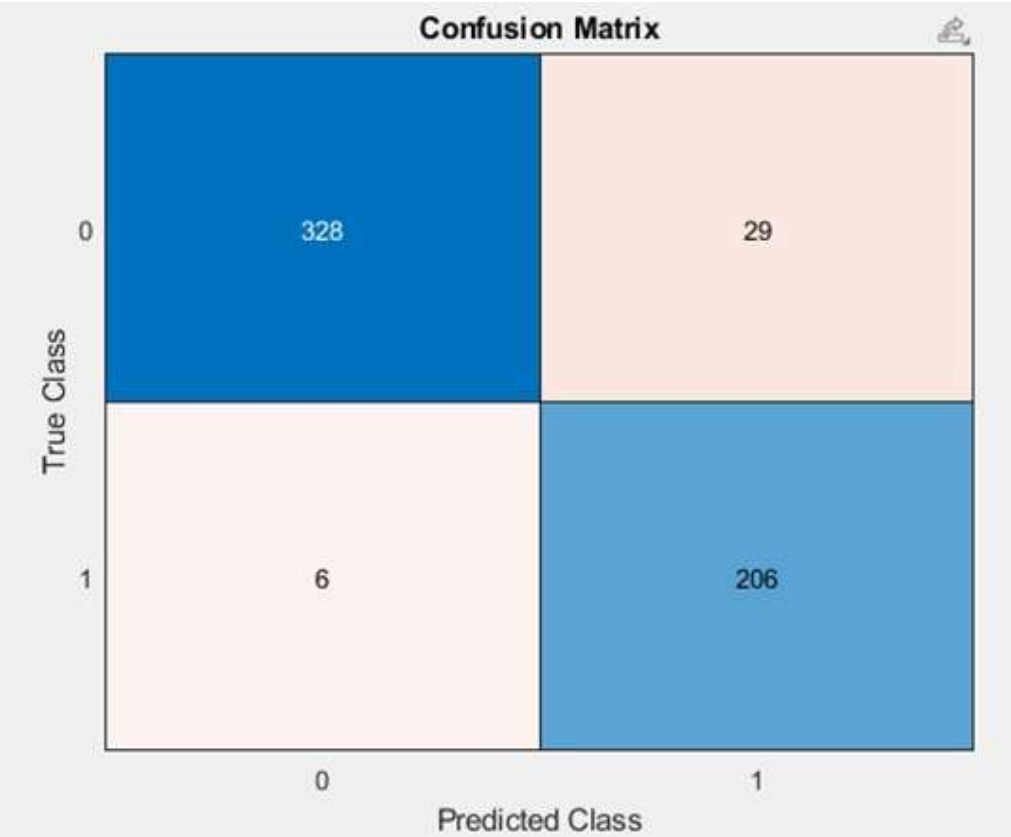

Figure 1: Confusion Matrix for LR classifier

The LR model consistently showed the same results in all 50 experiments conducted with the model. Because the accuracy, sensitivity, and specificity values recorded for the model showed a negligible standard deviation of less than 1%. We can conclude that the model is very stable and that the results of the model remain unaffected by the variations in the training/testing data. In other words, the model is able to provide high-quality results, thus being trusted with the performance; however, the reliability of the results will depend more on the application of this approach instead of the kind of dataset used in practice because the patient data can differ significantly.

Overall, the LR model proved to be interpretable and a statistically valid diagnostic method for BC classification. Owing to its high sensitivity, which is crucial for achieving a minimal number of false negatives, as well as decent specificity and excellent AUC, this model can be an ideal tool for making clinical decisions. This makes it particularly valuable in early diagnostic scenarios, where the accurate identification of malignant cases is essential for ensuring timely and effective patient care.

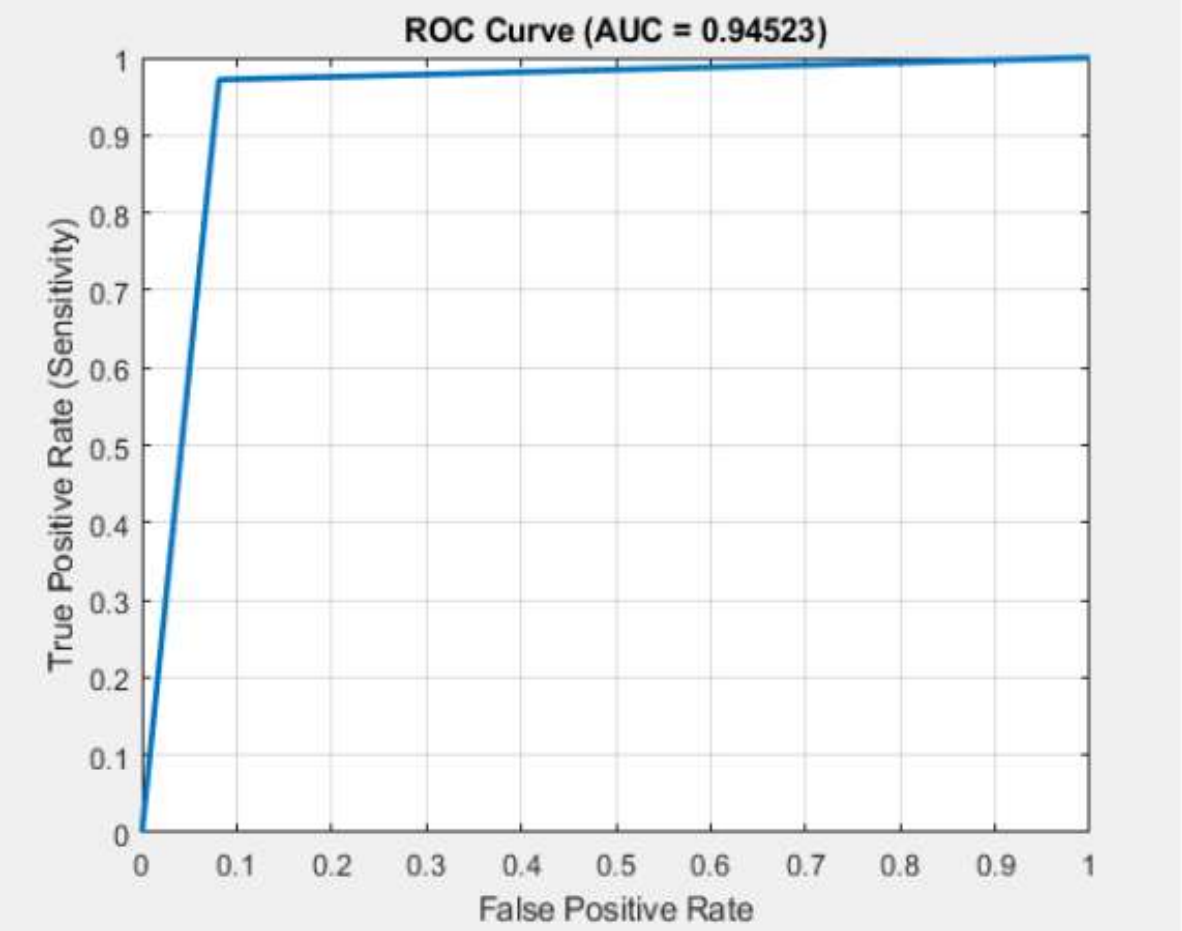

Figure 2. ROC curve

To assess the quality of the predictions delivered by the model, a Calibration Curve (Reliability Diagram) was generated (Figure 3). The graph illustrates a good match between the Logistic Regression model and the straight line referred to as the "Perfectly Calibrated" line at both ends of the spectrum. The model received an excellent score (zero malignant cases) for probabilities in the 0.0–0.4 range. It can be concluded that any sample assigned a low probability rating is benign. The middle ranges (0.6–0.8), representing the area where the line diverges from the diagonal, can be interpreted in light of the limited amounts of the "uncertainty" region typically found in the WDBC dataset, where many cases are different.

Model Interpretability and Feature Importance

To verify the interpretability of the LR model, the standardized coefficients for 30 input variables must be checked. Figure 4 shows that the worst texture has the highest positive coefficient of approximately 1.32, which makes it the second most important predictor after radius_se and symmetry_worst. The fifth feature, which involves the nuclear irregularity estimation, comprises parameters such as concave points_mean and concavity_worst, thus proving that the presence of concave indentations is one of the most significant cancer markers in terms of Logistic Regression. Standard error measures, such as radius_se and area_se, along with other names, proved that atypical cell sizes are more influential in identifying cancer than their average measurements, such as area_worst.

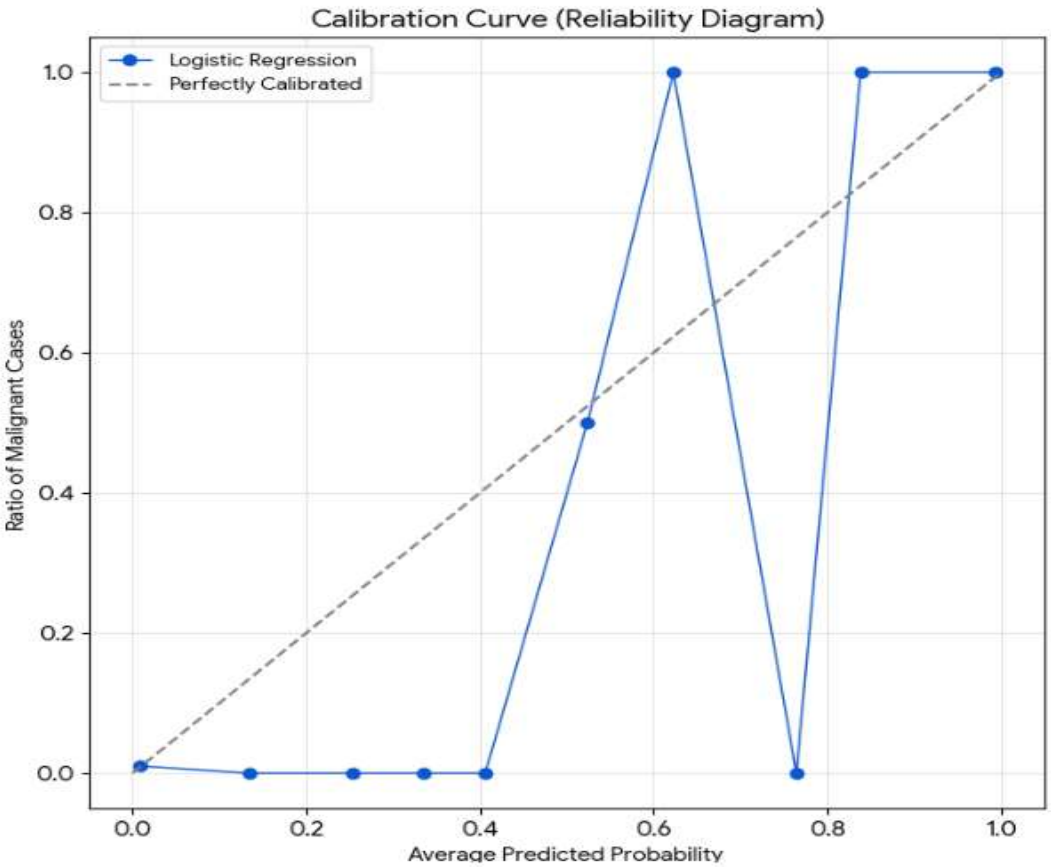

Figure 3: Calibration Curve (Reliability Diagram)

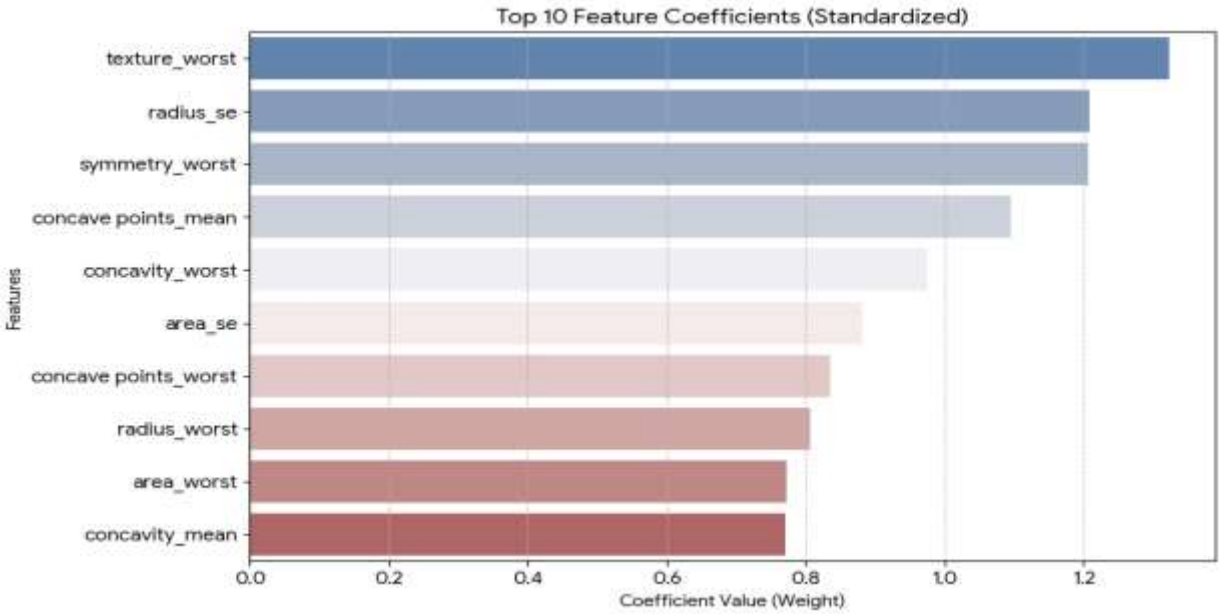

Figure 4: Top 10 features that has greatest influence on the diagnostic output

Discussion

The findings of this study show that an advanced LR model can produce clinically significant results in diagnosing BC using the WDBC database. Many independent experiments showed that the model achieved an average accuracy rate of 93.8 percent, with a specificity of 91.2 percent and a sensitivity level of 97.2 percent. The AUROC was 0.9452. Importantly, the model recorded only six false negatives and 29 false positives in the

error profile, indicating a high-risk profile of the findings. The results can be valuable for cancer diagnosis, as the model identifies the presence of possible illnesses.

The repeated indicators in different rounds of assessment emphasized the effectiveness of LR as a transparent and dependable tool for decision-making. These results are in agreement with the findings of Makwana et al., who showed that although the score of LR was slightly lower (approximately 92.7% accuracy) than that of nonlinear models such as random forests and support vector machines , the results were still good for the same dataset [11].

Nonetheless, our study is not as encouraging when placed in contrast to those by Hossin and Molla, which recorded almost perfect LR and SVM performance metrics (accuracy ≈99.1%, sensitivity ≈97.7%, specificity ≈100%) [9]. Likewise, the extremely high metrics reported by Rahman et al. for XGBoost (accuracy ≈99.1% with recall of 1.0 and specificity ≈0.986) [18] and Al Reshan et al. for stacked models (accuracy ≈99.9% with near-perfect metrics in sensitivity, specificity, and AUROC) [10] also stands as notable terms of contrast.

The inconsistencies between our findings and those from this collection are probably attributed to the different methodologies, not the difference in the quality of the results. However, we can notice that Hossin and Molla and Al Reshan et al. ignored a number of crucial “protections” against the optimistic bias, such as probability calibration, leakage-free data preprocessing pipelines, and nested cross-validation. All of these particularities of the methodology are crucial for the reports on the quality of the model to reflect the actual expectations from it, avoiding the overfitting of data. Therefore, despite the fact that logistic regression can hardly provide the best accuracy results regarding any other source, it can be viewed as a reliable and clear alternative for the process of making informed clinical decisions in terms of breast cancer diagnosis.

In addition, many previous studies have performed feature selection independent of cross-validation, thereby increasing the chance of overestimating the results [9, 10]. Consequently, our AUROC measure of 0.9452 can be said to correspond with those in studies of classical methods, since this AUROC measure reflects the "excellent" performance of the method. Nonetheless, it is important to emphasize that a number of existing studies tend to produce measures of uncertainty rather than relying on accurate AUROC calculations over several folds [9, 11, 14]. Regarding the practical aspect of discrimination capability, it should be stated that the sensitivity of our model, equal to 97.2%, is significantly higher than the threshold given by Makwana et al. (TPR ≈ 0.86 based on threshold 0.50). This highlights the importance of threshold selection in diagnostics [11].

The increasing popularity of nonlinear and ensemble learning models has given rise to many queries regarding the interpretability of these models and the trustworthiness of probabilistic outputs; however, researchers have hardly paid due attention to this. For instance, in the papers authored by Akter et al. and Islam et al., the authors implemented SHAP and LIME (for the latter work) to explain the results obtained using various ensemble methods [15, 16]. In contrast, most earlier studies lack investigations of the traditional multicollinearity in the WDBC dataset when establishing the impact of multicollinearity of features such as radius, area , perimeter, and worst descriptors in the uncertainties related to feature attribution and coefficient stability. Therefore, the interpretability of these models has been limited to single-fold evaluations or correlation heatmaps [9, 12, 15].

The use of LR in practice aids in the transparency of the analysis. The findings indicate that regularized and standardized linear models can be considered valid testing tools compared with complex methods. The study can be further elaborated by the introduction of other models and methods of evaluation, including coefficient stability analysis, permutation importance analysis, and cross-validated SHAP, thus addressing interpretability and the issue of correlation-based attribution bias [12, 15]. The literature contains information about calibration metrics (expected calibration error (ECE) and the Brier score) and post-hoc calibration methods (Platt method, isotonic regression) [9, 15, 16]. However, these methods and metrics are ignored or not reported, and incorporating fold-wise calibration and providing calibrated metrics could significantly improve both the comparability and real-world relevance of the results.

In clinical decision-making, risk thresholds are used more frequently than class labels. Interestingly, Manikanta et al. conducted the only study using cross-validation that makes use of hierarchical layers. Notwithstanding, the authors maximize the value of accuracy research and disregard concepts such as calibrated discrimination or metrics that can be applied in practice. This demonstrates the urgent need for the standardization of techniques that may be employed in future research [13].

The common limits in the existing WDBC literature have prevented us from making wide generalizations. Similar to previous studies by Hossin and Molla, Al Reshan et al., and Chaudhury et al., our analysis did not reach beyond a single dataset, and we are missing an external validation of the results [8, 9, 12]. Although the ensemble and kernel methods could present higher accuracy in a case study involving data from a single source, the absence of nested parameter selection, principled thresholding, and calibration raises questions about the reliability of the results under stricter evaluation scenarios [9, 10, 18]. Our repeated-trials implementation helps to reduce the variance associated with the use of data partitioning, but the employment of a fully nested

and stratified approach with threshold optimization and confidence intervals for AUROC, sensitivity, and specificity would provide a more reasonable basis for comparison with existing nonlinear approaches.

According to the assessment conducted on BC diagnosis using the WDBC dataset, applying LR oriented toward stability along with other appropriate clinical parameters improves its discriminative capacity while reducing its error rate. These results align with previous research that provided a somewhat more restrained assessment and are in stark contrast to overly optimistic statements normally associated with less rigorous validation practices and data leaks.

The use of calibration analysis has enhanced the literature on WDBC since most of the research done in this area, for instance, by Hossin and Molla as well as Al Reshan et al. has only discussed hard class labels (i.e. benign/malignant) but has omitted the study of risk scores quality. Moreover, in medical screening, the confidence level assigned to a model is as important as accuracy. The latter is demonstrated by our findings showing that the Expected Calibration Error is 0.0210 for our LR model, which suggests that the calculated probabilities are associated with clinical risk. For example, the probability of malignancy of 95% allows sending a cancerous patient for biopsy immediately, while patients that have almost no probability of malignancy can be monitored. Thus, it can be proved that traditional LR is well calibrated in regard of the dataset.

While numerous writers, such as Islam et al. and Akter et al., have made use of complex-to-interpret post-hoc explanation systems, such as SHapley Additive exPlanations (SHAP) LIME and SHAP, our model relies on the LR concept, relying on the direct principles of understandability and clarity. The weights of the model were equal, indicating that actual biological signals were captured and not noise. It should be pointed out that several leading characteristics, such as radius_se, area_se, and area_worst, are related mathematically. This proves the mentioned in Introduction issue of multicollinearity and confirms that all the characteristics have high individual values but in general have the same problem with size and shape. In subsequent research, this problem may be solved through the application of regularization methods (e.g., Lasso).

With the inclusion of calibrated probabilities, nested cross-validation, and correlation-accounting interpretability into the existing framework, we can enhance the usability in healthcare and standardize methods used by researchers including Hossin and Molla, Al Reshan et al, Makwana et al, Sharma et al, and others. This will lead to simplification of the process of making fair and clinically meaningful comparisons between linear, kernel, and ensemble methods [8-10, 13, 15]

The Limitations

The WDBC dataset is the sole public dataset utilized in this research. This dataset is helpful for research, but it does not encompass all human variability in clinical data. The approaches to disease manifestation and treatment differ among patients. Furthermore, only the LR technique was used in this study, and the possibilities of comparing this technique with more complicated models were lost. The findings' universality is diminished due to the absence of external validation with other independent datasets.

Recommendations

Future studies ought to concentrate on evaluating different aspects of various ML methods and ensemble approaches, their effectiveness, and their viability. In addition, it should be noted that the model should also be effective with data obtained from other organizations, which would lead to increased generalization of the study. Furthermore, several other methods (e.g., feature analysis, reduction of dimensionality, and insights into different algorithms) should be used to increase the efficiency and transparency of the model. By incorporating additional clinical and demographic data, the authors might increase the level of correctness and accuracy achieved, thus obtaining more information on the models that lead to accurate predictions.

Conclusion

According to the current study, the WDBC Dataset can be effectively used to categorize BC cases using the LR technique. The sensitivity and overall diagnostic performance were reported to be considerably high. These promising findings require further investigation to advocate the use of various datasets and compare the method with other techniques.

References

1. Zhao X, Dong YH, Xu LY, Shen YY, Qin G, Zhang ZB. Deep bone oncology Diagnostics: Computed tomography based Machine learning for detection of bone tumors from breast cancer metastasis. Journal of bone oncology. 2024;48:100638 . https://doi.org/10.1016/j.jbo.2024.100638
2. Jaganathan D, Balasubramaniam S, Sureshkumar V, Dhanasekaran S. Revolutionizing Breast Cancer Diagnosis: A Concatenated Precision through Transfer Learning in Histopathological Data Analysis. Diagnostics. 2024;14(4):422 . https://doi.org/10.3390/diagnostics14040422

3. Aamir S, Rahim A, Aamir Z, Abbasi SF, Khan MS, Alhaisoni M, et al. Predicting Breast Cancer Leveraging Supervised Machine Learning Techniques. Computational and mathematical methods in medicine. 2022;2022:5869529. https://doi.org/10.1155/2022/5869529
4. Masoud RM, Bakir RMA, Saraya MS, Ayyad SM. BREAST-CAD: A Computer-Aided Diagnosis System for Breast Cancer Detection Using Machine Learning. Technologies. 2025;13(7):268. https://doi.org/10.3390/technologies13070268
5. Rahman MA, Khan MSH, Watanobe Y, Prioty JT, Annita TT, Rahman S, et al. Advancements in Breast Cancer Detection: A Review of Global Trends, Risk Factors, Imaging Modalities, Machine Learning, and Deep Learning Approaches. BioMedInformatics. 2025;5(3):46 . https://doi.org/10.3390/biomedinformatics5030046
6. Cakmak Y, Safak S, Bayram MA, Pacal I. Comprehensive Evaluation of Machine Learning and ANN Models for Breast Cancer Detection. Computer and Decision Making: An International Journal. 2024;1:84-102. https://doi.org/10.59543/comdem.v1i.10349
7. Hameed Alsaedi NR, Akay MF. Effective Breast Cancer Classification Using Deep MLP, Feature-Fused Autoencoder and Weight-Tuned Decision Tree. Applied Sciences. 2025;15(13):7213 . https://doi.org/10.3390/app15137213
8. Taghizadeh E, Heydarheydari S, Saberi A, JafarpoorNesheli S, Rezaeijo SM. Breast cancer prediction with transcriptome profiling using feature selection and machine learning methods. BMC bioinformatics. 2022 Oct 1;23(1):410 https://doi.org/10.1186/s12859-022-04965-8 .
9. Hossin MM, Shamrat FJM, Bhuiyan MR, Hira RA, Khan T, Molla S. Breast cancer detection: an effective comparison of different machine learning algorithms on the Wisconsin dataset. Bulletin of Electrical Engineering and Informatics. 2023;12(4):2446-56. https://doi.org/10.11591/eei.v12i4.4448
10. Reshan MSA, Amin S, Zeb MA, Sulaiman A, Alshahrani H, Azar AT, et al. Enhancing breast cancer detection and classification using advanced multi-model features and ensemble machine learning techniques. Life. 2023;13(10):2093 . https://doi.org/10.3390/life13102093
11. Makwana S. Optimizing Breast Cancer Diagnosis with Machine Learning Algorithms. Journal of Electrical Systems. 2024;20:4046-54. https://doi.org/10.52783/jes.5970
12. Chaudhury S, Shelke N, Rashid ZM, Sau K. Effect of grid search and hyper parameter tuned pipeline with various classifiers and PCA for breast cancer detection. Current Signal Transduction Therapy. 2022;17(3):45-56 http://dx.doi.org/10.2174/1574362417666220715105527
13. Manikanta S, Rasheed SM, Reddy JK, Sankar SNSB, Harshendra AVS, Srinivas M, editors. Semi-supervised Learning-An Alternative to Traditional Breast Cancer Prediction. 2025 International Conference on Artificial Intelligence and Data Engineering (AIDE); 2025: IEEE. https://doi.org/10.1109/AIDE64228.2025.10987549
14. Adem AM, Kant R, Kumari S, Swaroopa CN, Gupta G, editors. Breast Cancer Classification: In-depth Exploration of Different Paradigms and Ensemble Technique. 2024 11th International Conference on Computing for Sustainable Global Development (INDIACom); 2024: IEEE. https://doi.org/10.23919/INDIACom61295.2024.10498454
15. Akter MS, Shuvo MB, Tithe SN, Nissan MNI, Bhowmik P, Hossain MD, editors. Explainable Multi-Model Framework for Breast Cancer Diagnosis Using Heterogeneous Clinical Data. 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN); 2025: IEEE https://doi.org/10.1109/QPAIN66474.2025.11171915 .
16. Islam MN, Rahman MM, Shanta SS, Salakin S, Akash AAK, Bijoy MHI, editors. Breast Cancer Classification using Machine Learning Techniques with Explainable AI. 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE); 2025: IEEE. https://doi.org/10.1109/ECCE64574.2025.11013927
17. Wolberg, W., Mangasarian, O., Street, N., & Street, W. (1993). Breast Cancer Wisconsin (Diagnostic) [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5DW2
18. Rahman MA, Hamada M, Sharmin S, Rimi TA, Talukder AS, Imran N, et al. Enhancing early breast cancer detection through advanced data analysis. IEEE Access. 2024. https://doi.org/10.1109/ACCESS.2024.3483095
19. Eftekharian, M., et al. (2025). Accurate and Interpretable Breast Cancer Diagnosis Using Logistic Regression. InfoScience Trends, 2(09), 52-60.
20. Islam, M. N., et al. (2025). Breast Cancer Classification using Machine Learning Techniques with Explainable AI. IEEE ECCE.
21. Akter, M. S., et al. (2025). Explainable Multi-Model Framework for Breast Cancer Diagnosis Using Heterogeneous Clinical Data. IEEE QPAIN.

22. Makwana, S. (2024). Optimizing Breast Cancer Diagnosis with Machine Learning Algorithms. Journal of Electrical Systems.
23. Singh, R., Upadhyay, S. K., Tuli, H. S., Singh, M., Kumar, V., Yadav, M., Aggarwal, D., & Kumar, S. (2020). Ethnobotany and herbal medicine: Some local plants with anticancer activity. Bulletin of Pure and Applied Sciences - Botany, 39B(1), 57-64. https://www.bpasjournals.com/botany/index.php/journal/article/view/101/98