© December 2024 | IJIRT | Volume 11 Issue 7 | ISSN: 2349-6002

Optimizing Breast Cancer Prediction and Identifying a Machine Learning Model using a Data – Driven Approach

Ishita Roy[1], K Vaishnavi[2], J Mahesh Babu[3], B Mojesh[4], Mr. T Raghavendra Gupta[5]
[1,2,3,4,5] Department of Computer Science and Engineering/Hyderabad Institute of Technology and Management/India

Abstract—**Breast cancer is among the most common cancers that have affected men and women in the world. Identifying it promptly is essential to improve patient outcomes. Beginning with the preprocessing of a dataset related to tumor characteristics, the project involves cleaning the data, encoding categorical variables, and standardizing features. It tests a variety of machine learning algorithms, ranging from basic models such as Gradient Boosting, Neural Networks, K-Nearest Neighbors (KNN), Decision Trees, Random Forests, Support Vector Machine (SVM), and Logistic Regression, alongside advanced models like XGBoost and LightGBM. Evaluation of each model in terms of metrics including accuracy, precision, recall, F1-score, and ROC-AUC score with hyperparameter tuning is conducted for selected models like Random Forest to enhance performance. The results go through comparison and documentation in order to identify the optimal model so that future work recommendations can be provided, including potential dataset enhancements and model refinements.**

Index Terms—**Data driven approach, Model Evaluation, Gradient Boosting, Neural Networks, K-Nearest Neighbors, Decision Trees, Random Forests, Support Vector Machine (SVM), Logistic Regression, LightGBM, XGBoost.**

I. INTRODUCTION

Amongst other prevalent cancers around the world, breast cancer tops the list and claims the lives of millions worldwide, with women being the highest affected population. It is reported that the number of deaths from breast cancer worldwide has increased by 14% since 2008. Early detection is critical to improving survival rates, as timely intervention can significantly enhance treatment outcomes and increase survival rates.

Machine learning offers a promising solution by automating and optimizing predictive models for identifying breast cancer at earlier stages. This project leverages a data-driven approach to evaluate various ML algorithms for breast cancer prediction, with preprocessing tumor characteristic datasets through steps like data cleaning, encoding categorical variables, and standardizing features to ensure high-quality model input. The project tests both basic and advanced algorithms which include Gradient Boosting, Neural Networks, K-Nearest Neighbors (KNN), Decision Trees, Random Forests, Support Vector Machine (SVM), Logistic Regression, XGBoost and LightGBM. Each model is tested through a critical performance measure that includes ROC-AUC curve, recall, accuracy, F1 Score, and precision.

This paper discusses the critical requirement for better diagnostic tools and does so by looking into the potential of machine learning in the detection of breast cancer. Large patient information datasets make it possible to train and validate these models. The present work focuses on the comparison of the performance of a variety of machine learning algorithms from traditional methods to more complex ensemble techniques. The ultimate goal here would be to assist in creating newer, more efficient screening processes that are easier for patients to access, thereby impacting overall patient outcomes and reducing deaths due to breast cancer. The comparison of the algorithms here will help to bring better understanding of the strengths and weaknesses associated with each algorithm toward better prediction of breast cancer

II. RELATED WORK

The paper [1] " Prediction of Breast Cancer using Machine Learning Approaches" discusses the evolution and application of machine learning models in analyzing breast cancer data, focusing on both traditional and modern algorithms. It also emphasizes the importance of feature selection, data preprocessing, and model evaluation metrics in achieving accurate predictions with models such as Random Forest, Support Vector Machine, decision trees and neural networks. The paper [2] " Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms" is used for predicting breast cancer survival outcomes using several machine learning algorithms, to determine their effectiveness in forecasting patient survival based on clinical and genomic data. The authors focus on the performance of these models in terms of accuracy, precision, and recall. It provides insights into these approaches can be utilized to improve survival prediction and personalized treatment strategies in breast cancer care. The project [3] " Breast Cancer Detection and Prevention Using Machine Learning " by Hiba Masood reviews techniques applied to the Wisconsin Breast Cancer dataset. The proposed system employs image preprocessing techniques like adaptive mean filtering, k-means clustering, and Hidden Markov Models (HMRF) to increase detection accuracy. Despite its promise, the system faces limitations such as the complexity of training large datasets, potential overfitting, and the need for improved accuracy in real-world applications. The paper [4] " Accurate Prediction of Neoadjuvant Chemotherapy Pathological Complete Remission (pCR) for the Four Sub-Types of Breast Cancer" investigates the potential for predicting pathological complete remission (pCR) in breast cancer patients who neoadjuvant chemotherapy (NAC) using only the nodal sizes from the first three treatment cycles. It demonstrates that a predictive model can achieve an average accuracy (Avc) of 0.7594 across four breast cancer subtypes. Challenges identified include the need for larger datasets to validate the models and ensure they are robust against overfitting. Future directions involve conducting multi-center clinical trials to enhance model reliability and exploring additional predictive factors to improve accuracy further, ultimately aiming to assist in treatment decision-making after initial NAC cycles. The paper [5] " Breast cancer detection using machine learning approaches: a comparative study” explores the algorithms for breast cancer classification. The researchers compared eight commonly used algorithms, including Multilayer Perceptron, and Logistic Regression, using the Wisconsin Breast Cancer Database (WBCD). Their findings suggest that these two algorithms had high accuracy (around 98%) in differentiating benign and malignant tumors. Future directions, as proposed by the study, could involve incorporating additional data sources (e.g., genetic information) and exploring more sophisticated algorithms like deep learning for better accuracy. The paper [6] called “Prediction of benign and malignant breast cancer using data mining techniques” emphasizes the effectiveness of combining multiple algorithms to improve classification accuracy compared to single models. It identifies key features from datasets that contribute to accurate predictions of malignancy and benignity in these cases. Challenges highlighted in the research include the need for diverse and large datasets to train models effectively and the potential for overfitting when using complex algorithms. The paper [7] titled " Prediction of Cancer Disease using Machine Learning Approach" by F.J. Shaikh and D.S. Rao focuses on leveraging machine learning techniques to enhance the accuracy of diagnosis. It discusses various algorithms, including decision trees and neural networks, that can analyze clinical data to identify patterns indicative of cancerous conditions. It highlights the potential of these methods to improve early detection rates and personalize treatment plans. The paper [8] " Machine Learning Model for Breast Cancer Detection " focuses on extracting features from a dataset sourced from the Wisconsin database, containing information from Fine Needle Aspirate (FNA) images of breast masses. The model aims to classify tumors as either benign or malignant, achieving an accuracy of 94.3%. Challenges identified in the study include the need for high-quality data and the complexity of feature selection, which can impact model performance. The future scope suggests further research into improving model accuracy through advanced algorithms, like Convolutional Neural Networks (CNNs), and exploring additional datasets to enhance the applicability of the model in clinical settings. The paper [9] "Predicting diagnosis and survival of bone metastasis in breast cancer using machine learning" by Xugang Zhong, Yanze Lin, Wei Zhang and Qing Bi identified key risk and prognostic factors and developed highly accurate diagnostic and prognostic models using the XG Boost algorithm by analyzing the SEER database. Surgery emerged as a critical factor for both diagnosis and prognosis. The study [10] titled

" AI models predicting breast cancer distant metastasis using LightGBM with clinical blood markers and ultrasound maximum diameter" addresses the critical challenge of predicting distant metastasis in breast cancer. By integrating clinical blood markers and ultrasound data, the researchers developed predictive models utilizing the LightGBM machine learning algorithm, which demonstrated superior accuracy compared to traditional imaging methods. Despite its promise, the study highlights significant challenges, including the difficulty in distinguishing between benign and malignant nodules in imaging and the need for extensive validation across diverse patient populations to ensure generalizability and clinical applicability. The paper [11] " Cancer Detection and Prevention Using Machine Learning" by Khalid A introduces a new method of breast cancer detection using advanced machine learning techniques. The authors of this study suggest an efficient deep learning model that can identify breast cancer in computerized mammograms with varying densities. It emphasizes the importance of early detection and treatment for improved outcomes among patients since breast cancer is still the leading cause of death among women worldwide. The paper [12] " Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI" classifies breast cancer among Bangladeshi patients by the use of different machine learning algorithms while keeping the decisions of the model interpretive through explainable AI techniques is developed. The classification accuracy of the study proved to be high, showing the potential of improving clinical decision-making using machine learning in local health care settings. The paper [13] " Machine Learning Algorithms for Breast Cancer Prediction And Diagnosis" investigates the strength of algorithms applied in predicting and diagnosis of breast cancer. Applying five algorithms on the Breast Cancer Wisconsin Diagnostic dataset, results show SVM outperforms others by 97.2%, implying its potentiality to develop breast cancer diagnosing strategies and treatment. The paper [14] " Deep learning algorithms for the early detection of breast cancer: A comparative study with traditional machine learning" by by Rolando Gonzales Martinez and Daan-Max van Dongen studies the efficiency of deep learning methods versus traditional machine learning approaches in early breast cancer detection. It is argued that deep learning algorithms are mostly superior in accuracy and performance than traditional approaches, meaning that there is a promise in their improvement of the early diagnosis and, consequently, outcomes of breast cancer screening. The paper [15] " Understanding breast cancer as a global health concern" Louise Wilkinson and Toral Gathani presents the increasing incidence of breast cancer, which is now the most frequently diagnosed cancer in the world, with an estimated 2.26 million new cases in 2020. The authors underscore the survival gap between developed and developing regions, attributing this gap to delayed diagnosis and limited access to effective treatment, yet also mentioning the Global Breast Cancer Initiative of the World Health Organization to improve outcomes through health promotion, timely diagnosis, and comprehensive care. The paper titled [16] "Comparison of Machine Learning Methods for Breast Cancer Diagnosis" compares algorithms and finds that SVM was the most accurate method; thus, the conclusion reached is that it may help to enhance the diagnosis outcome in the case of breast cancer. The paper [17] "Gaussian Dropout Based Stacked Ensemble CNN for Classification of Breast Tumor in Ultrasound Image" is one of the latest innovates-presenting deep architecture with an idea of Gauss drop in stack ensemble with CNNs toward the goal of classification of breast ultrasonography tumor. The study results report 92.15% accuracy, thus evidencing this approach as effectively improving model generalization, reducing overfitting, and enhancing the efficacy of detecting breast cancer in ultrasound imaging. Research [18] "Breast cancer detection using deep learning: Datasets, methods, and challenges ahead" presents a broad overview of the application of deep learning techniques in the detection of breast cancer. Authors discussed various datasets used to train deep learning models, discussed the diverse methodologies adopted in the domain, and outlined the challenges encountered when implementing such technologies, especially underlining the requirement of better model interpretability and generalization in order to ensure higher accuracy in diagnoses. A. Sharma, S. Kulshrestha, and S. Daniel present the analysis of various machine learning techniques in the diagnosis and prognosis of breast cancer in their paper [19] titled "Machine learning approaches for breast cancer diagnosis and prognosis". It indicates that these are indeed effective algorithms in achieving a very high accuracy rate-the highest of which was provided by Logistic Regression at 98.1%. These methods may therefore be applicable to improving clinical decision-making capabilities in the management of breast cancer. The paper [20] "A Comprehensive Review for Breast Histopathology Image Analysis Using Classical and Deep Neural Networks" provides an extensive overview of both classical and deep learning techniques applied to the analysis of breast histopathology images. It groups the methods, explains their strengths toward better accuracy in diagnostics, and mentions the limitations present in the field of classification, indicating the importance of strong models that would efficiently deal with the intricacy of histopathological data and yet generalize to others. The paper [21] "A Brief Survey on Breast Cancer Diagnostic With Deep Learning Schemes Using Multi-Image Modalities" by T. Mahmood, J. Li, Y. Pei, F. Akhtar, A. Imran and K. U. Rehman pointed out the advantages of multi-modal data toward more accurate diagnosis, along with the challenges that integrate and train models for prediction and consequently advocate the design of further robust deep learning frameworks that improve outcomes both on detection and treatment of the breast cancer. The paper [22] "Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis" providing a broader view of the diversity of techniques in terms of applicability with machine learning approaches toward the prediction and diagnosis of breast cancer, discussing how performance varies in the execution of diversified algorithms for accomplishing high rates in accuracy along with a special emphasis on feature selection and interpretability of the model upon its outcome in diagnosis. The paper [23] titled “An Automatic Computer-Aided Diagnosis System for Breast Cancer in Digital Mammograms via Deep Belief Network” presents an automatically computed system of computerized aid in the diagnosis by employing DBN to automatically analyze a digital mammogram regarding the diagnosis of breast cancer. This work has highlighted the efficiency of the proposed system and has shown a mean AUC of 0.952 that indicates very good accuracy in malignancy identification across the range of breast densities. Thus, deep learning-based methods reveal promise for further improvement in mammography screening.

III. PROBLEM STATEMENT

Breast cancer represents a major global health challenge, with a high mortality rate in men and women. Despite advances in diagnostic techniques, the early detection of breast cancer remains complex and often unreliable, leading to challenges like delayed diagnoses, misinterpretation of tumor characteristics, and consequently, poorer treatment outcomes. The advanced data-driven method to improve the accuracy and efficiency of the breast cancer prediction problem requires development. The objective is to evaluate machine learning algorithms, including both basic models (such as Decision Trees and Logistic Regression) and advanced models (such as XGBoost and LightGBM), while optimizing their performance to achieve the highest possible accuracy and reliability. The aim is to determine the model that achieves the maximum possible accuracy and reliability in breast cancer prediction. By identifying the optimal algorithm, this project seeks to contribute to the development of more precise and timelier in developing prediction tools even more precise for healthcare professionals in making better-informed decisions, ultimately leading to improved patient outcomes.

IV. PROPOSED METHODOLOGY

The methodology is structured into key stages, beginning with data pre-processing, and leading to model evaluation and optimization. The process begins with exploratory data analysis (EDA), where the dataset is examined for missing values, distribution of features, and any outliers. Key tumour-related features are analyzed, and missing values are handled using imputation techniques. The data is then standardized to ensure uniformity across features, which is especially important for models such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVM). Several machine learning algorithms are implemented, including Neural Networks, Decision Trees, Random Forests, Gradient Boosting, SVM, KNN, Logistic Regression, XGBoost, and LightGBM. Every model is trained using the dataset that has been pre-processed. Key performance metrics, such as accuracy, precision, recall, F1-score, and ROC-AUC, are used to compare model effectiveness. The project concludes by comparing the performance of all models and selecting the one with the best balance of accuracy and interpretability for breast cancer prediction.

A. Block Diagram

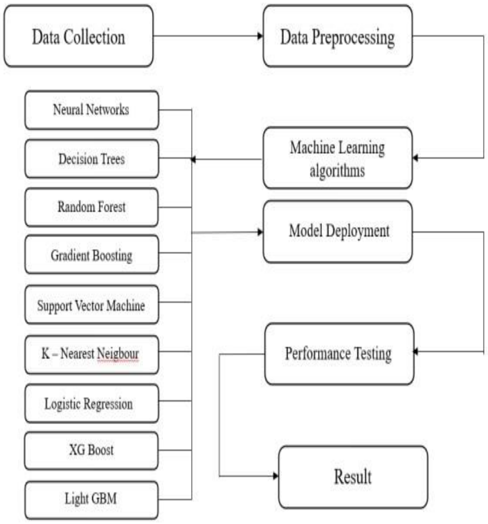

Fig 4.1 Block Diagram

B. Data Pre-Processing

Data preprocessing is helps to safeguard the quality of the data used for training any machine learning model. The following actions were taken:

- The breast cancer dataset consisting of additional features and diagnosis labels in a CSV file, was brought into the Python environment using the Pandas library.
- The dataset was streamlined to focus only on important features by discarding some columns. For instance, 'id' and 'Unnamed: 32' columns were removed from the dataset.
- The 'diagnosis' target variable, which indicates the status of tumor whether benign or malignant, was converted into a number using the function LabelEncoder. This step is important since most machine learning models work with numerical values.
- The data was split into independent features (X) and the dependent variable that is to be predicted (Y).
- To make an effective evaluation, the dataset was divided into the ratio of 70% for training and 30% for test sets, to enable unbiased analysis of the performance of the model on unseen data.
- Standardization of the feature set was done to make the performance better. The StandardScaler function from scikit-learn was utilized to standardize the features by removing the mean and scaling to unit variance to ensure that all features contribute equally to model training.

C. Model Selection

A few machine learning algorithms were selected to identify the most suitable model for breast cancer prediction:

Basic Algorithms: Neural Networks, Decision Trees, Random Forests, Gradient Boosting, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Logistic Regression

Advanced Algorithms: XGBoost and LightGBM

D. Model Training and Evaluation

Each model was trained on the standardized training set and evaluated using various metrics:

- Each classifier was fitted to the training data, followed by predictions on the test set.
- Performance Metrics: The accuracy score, ROC AUC score, classification report, and confusion matrix were computed for each model. These metrics reveal observations into model performance in terms of both classification accuracy and error analysis.

E. Prediction on New Data

After evaluation, the most accurate machine learning model is used to predict new patient data. Finally, the model is used to give predictions about whether the tumor is benign or malignant along with the probability of malignancy.

V. RESULT AND DISCUSSION

This research tested several machine learning algorithms in terms of predicting outcomes of breast cancer patients; it used accuracy, precision, recall, F1-score, and ROC AUC metrics. The results are summarized in the table below, which highlights the accuracy of each model:

Table 5.1 Accuracy of the models (in %)

Model	Accuracy (in %)
Neural Network	97.66
Decision Tree	94.15
Random Forest	97.07
Gradient Boosting	95.90
Support Vector Machine	97.66
K – Nearest Neighbor	95.90
Logistic Regression	98.24
XGBoost	97.07
LightGBM	95.90

Among the models evaluated, Table 5.1 shows that Logistic Regression achieved the highest accuracy of 98.25%, followed by both Neural Network and SVM, which recorded accuracies of 97.66%. The Random Forest and XG Boost models also performed well, each achieving an accuracy of 97.08%. In contrast, the Decision Tree classifier exhibited the lowest accuracy at 94.15%, indicating that it may not be the best choice for this dataset.

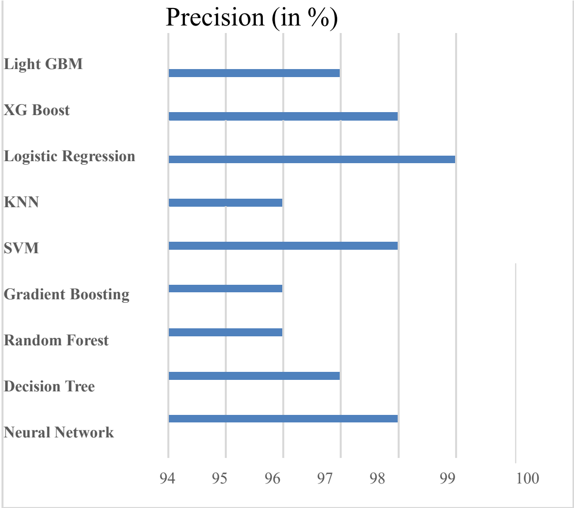

Fig 5.1 Precision Bar Chart

Figure 5.1 presents precision scores for each model, showing that Logistic Regression has the maximum precision value, indicating a strong ability to correctly identify positive cases.

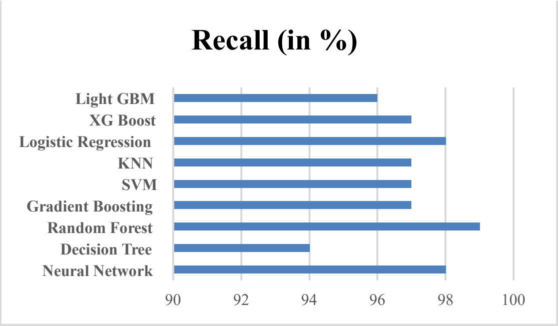

Fig 5. 2 Recall Bar Chart

Figure 5.2 indicates how well each model identifies actual malignant cases, with Random Forest achieving the highest recall score.

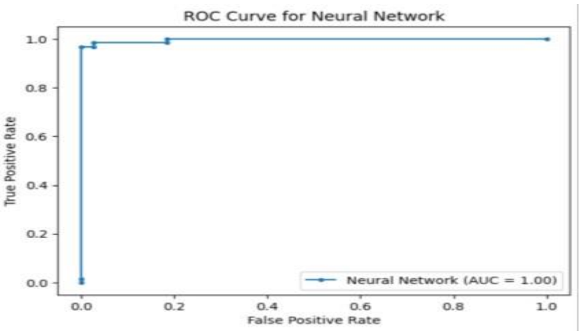

Fig 5.3 ROC Curve for Neural

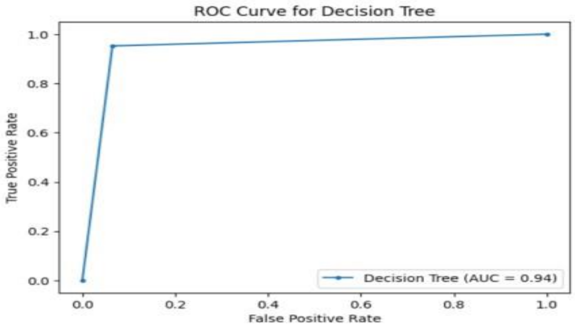

Fig 5.4 ROC Curve for Decision Tree

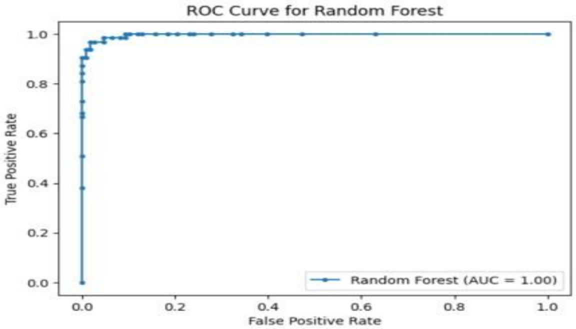

Fig 5.5 ROC Curve for Random Forest

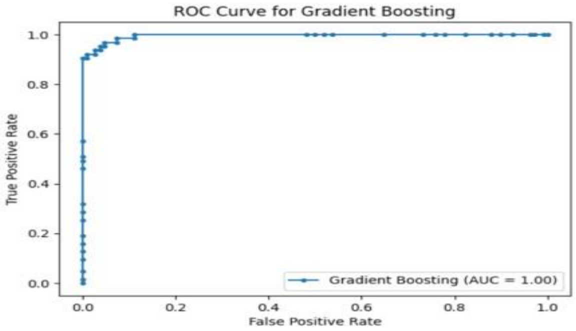

Fig 5.6 ROC Curve for Gradient Boosting

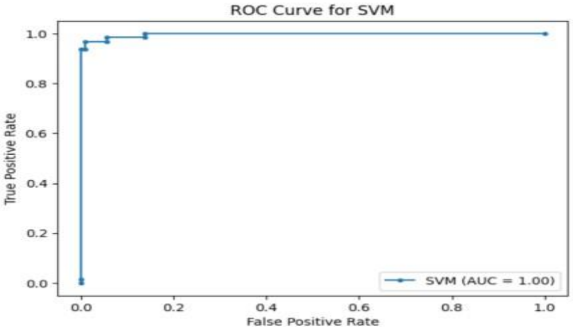

Fig 5.7 ROC Curve for Support Vector Machine

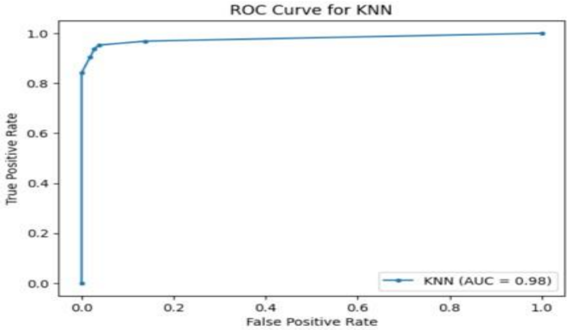

Fig 5.8 ROC Curve for K-Nearest Neighbor

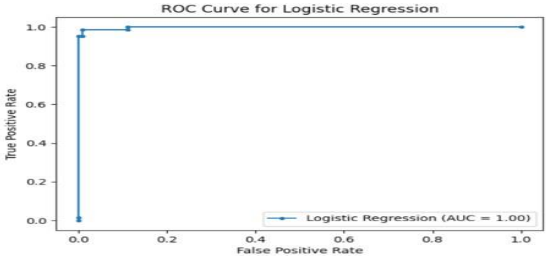

Fig 5.9 ROC Curve for Logistic Regression

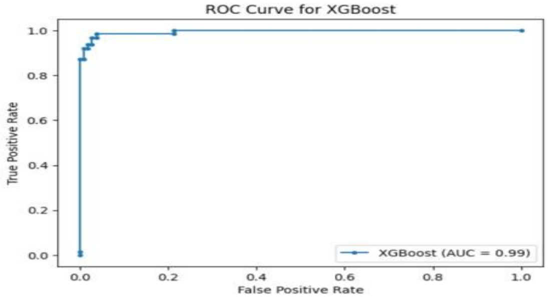

Fig 5.10 ROC Curve for XGBoost

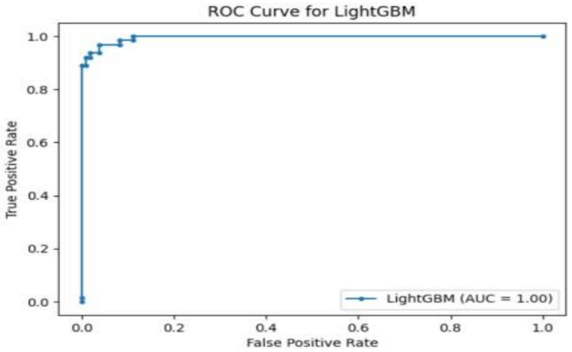

Fig 5.11 ROC Curve for LightGBM

The above graphs display the ROC curves for each model, with Logistic Regression showing the highest area under the curve (AUC), which signifies its model's ability to distinguish between benign and malignant tumors efficiently.

VI. CONCLUSION

In conclusion, our analysis demonstrates that Logistic Regression best captures the breast cancer class predictor with the highest scores on accuracy, precision, and recall. The high ROC AUC values across multiple models indicate robust performance in distinguishing between classes. These findings suggest that implementing machine learning algorithms like Logistic Regression can significantly enhance diagnostic processes in clinical settings for the detection of cancer. Future work could explore ensemble methods or deep learning approaches to further improve predictive performance and generalizability across different datasets.

REFERENCES

[1] Rabiei R, Ayyoubzadeh SM, Sohrabei S, Esmaeili M, Atashi A. Prediction of Breast Cancer using Machine Learning Approaches. J Biomed Phys Eng. 2022 Jun 1;12(3):297-308. doi: 10.31661/jbpe. v0i0.2109-1403. PMID: 35698545; PMCID: PMC9175124.

[2] Botlagunta M, Botlagunta MD, Myneni MB, Lakshmi D, Nayyar A, Gullapalli JS, Shah MA. Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms. Sci Rep. 2023 Jan 10;13(1):485. doi: 10.1038/s41598-023-27548-w. PMID: 36627367; PMCID: PMC9831019

[3] Hiba Masood - Breast Cancer Detection using Machine Learning Algorithm

[4] Xin Feng, Lelian Song, Shaofei Wang, Haoqiu Song, Hang Chen, Yuxuan Liu1, Chenwei Lou, Jian Zhao, Quewang Liu, Yang Liu, Ruixue Zhao, Kai Xing, Sijie Li, Yunhe Yu, Zhenyu Liu, Chengyang Yin, Bing Han, Ye Du, Ruihao Xin, Lan Huang, Zhimin Fan, And Fengfeng Zhou - Accurate Prediction of Neoadjuvant Chemotherapy Pathological Complete Remission (pCR) for the Four Sub-Types of Breast

[5] Muawia Abdelmagid Elsadig, Abdelrahman Altigani, Huwaida Tagelsir Elshoush - Breast cancer detection using machine learning approaches: a comparative study

[6] Vikas Chaurasia, Saurabh Pal, BB Tiwari -Prediction of benign and malignant breast cancer using data mining techniques

[7] F.J. Shaikh, D.S. Rao, Prediction of Cancer Disease using Machine learning Approach,Materials Today: Proceedings,Volume 50, Part 1,2022,Pages 40-47,ISSN 2214-7853

[8] Malik Adeiza Rufai, Ahmad Shehu Muhammad, Garba Suleiman, Audu Lovingkindness - Machine Learning Model for Breast Cancer Detection

[9] Zhong, X., Lin, Y., Zhang, W. et al. Predicting diagnosis and survival of bone metastasis in breast cancer using machine learning. Sci Rep 13, 18301 (2023).

[10] Tan, Y., Zhang, Wh., Huang, Z. et al. AI models predicting breast cancer distant metastasis using LightGBM with clinical blood markers and ultrasound maximum diameter. Sci Rep 14, 15561 (2024).

[11] Khalid A, Mehmood A, Alabrah A, Alkhamees BF, Amin F, AlSalman H, Choi GS. Breast Cancer Detection and Prevention

[12] Using Machine Learning. Diagnostics (Basel). 2023 Oct 2;13(19):3113. doi: 10.3390/diagnostics13193113. PMID: 37835856; PMCID: PMC10572157.

[13] Islam, T., Sheakh, M.A., Tahosin, M.S. et al. Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI. Sci Rep 14, 8487 (2024). https://doi.org/10.1038/s41598024-57740-5

[14] Mohammed Amine Naji, Sanaa El Filali, Kawtar Aarika, EL Habib Benlahmar, Rachida Ait Abdelouhahid, Olivier Debauche,Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis,Procedia Computer Science,Volume 191,2021,Pages 487-492,ISSN 1877-0509.

[15] Rolando Gonzales Martinez, Daan-Max van Dongen, Deep learning algorithms for the early detection of breast cancer: A comparative study with traditional machine learning,Informatics in Medicine Unlocked,Volume 41,2023,101317,ISSN 23529148.

[16] Louise Wilkinson, Toral Gathani, Understanding breast cancer as a global health concern, British Journal of Radiology, Volume 95, Issue 1130, 1 February 2022, 20211033.

[17] E. A. Bayrak, P. Kırcı and T. Ensari, "Comparison of Machine Learning Methods for Breast Cancer Diagnosis," 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey, 2019, pp. 13, doi: 10.1109/EBBT.2019.8741990.

[18] R. Karthik, R. Menaka, G.S. Kathiresan, M. Anirudh, M. Nagharjun, Gaussian Dropout Based Stacked Ensemble CNN for Classification of Breast Tumor in Ultrasound Images, IRBM,Volume 43, Issue 6,2022,Pages 715-733,ISSN 1959-0318

[19] Nusrat Mohi ud din, Rayees Ahmad Dar, Muzafar Rasool, Assif Assad, Breast cancer detection using deep learning: Datasets, methods, and challenges ahead, Computers in Biology and Medicine,Volume 149,2022,106073,ISSN 0010-4825

[20] A. Sharma, S. Kulshrestha and S. Daniel, "Machine learning approaches for breast cancer diagnosis and prognosis," 2017

[21] International Conference on Soft Computing and its Engineering Applications (icSoftComp), Changa, ndia, 2017, pp. 1-5, doi: 10.1109/ICSOFTCOMP.2017.8280082

[22] 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

[23] T. Mahmood, J. Li, Y. Pei, F. Akhtar, A. Imran and K. U. Rehman, "A Brief Survey on Breast Cancer Diagnostic with Deep Learning Schemes Using Multi-Image Modalities," in IEEE Access, vol. 8, pp. 165779-165809, 2020, doi:

[24] 10.1109/ACCESS.2020.3021343.

[25] N. Fatima, L. Liu, S. Hong and H. Ahmed, "Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis," in IEEE Access, vol. 8, pp. 150360-150376, 2020, doi: 10.1109/ACCESS.2020.3016715

[26] Al-antari, M.A., Al-masni, M.A., Park, SU. et al. An Automatic Computer-Aided Diagnosis System for Breast Cancer in Digital Mammograms via Deep Belief Network. J. Med. Biol. Eng. 38, 443–456 (2018).