BREAST CANCER PREDICTION USING MACHINE LEARNING

BODDEPALLI JAHNAVI

Department of Computer Science and Engineering SR university Warangal , Telengana

2203A52142,SR university e-mail: 2203A52142@sru.edu.in

Abstract— *Breast cancer is a major global health concern, and early detection is essential for effective treatment and patient outcomes. Predictive modeling approaches offer promising approaches to improve breast cancer detection by using available patient data to predict the likelihood of disease occurrence This abstract provides an overview of recent advances in predictive modeling for breast cancer prediction. Machine learning algorithms, such as vector-supported machines, decision trees, random forests, and deep neural learning, are widely used to analyze datasets in clinical and genetic imaging contexts These models of breast cancer risk use features extracted from mammograms, genetic information, patient demographics and clinical history to make predictions Furthermore, the integration of advanced imaging technologies such as magnetic resonance imaging (MRI) and digital breast tomosynthesis (DBT) has facilitated the development of accurate predictive models that provide detailed physiological functional information and genomic data along with the advent of precision medicine are enabled, enabling personalized risk assessments and personalized treatment strategies However, challenges remain to optimize the performance and generalizability of predictive models, including heterogeneous data sets, sample size limitations, and interpretation issues In conclusion, predictive modeling holds tremendous potential for improvement Detecting and managing breast cancer by facilitating early intervention and self-care. Continued research efforts aimed at refining model performance, improving data quality, and addressing ethical concerns are needed to realize the full value of predictive analytics control breast cancer prognosis and prevention.*

Keywords— *Breast Cancer, Model Building Prognosis, Machine Learning, Early Detection, Risk Assessment, Precision Medicine, Imaging Technologies, Genomic Data, Data Heterogeneity, Ethical Considerations*

INTRODUCTION

Breast cancer prognosis is an important health care intervention that aims to identify individuals at high risk of developing breast cancer. This process involves evaluation of many factors such as genetic predisposition, family history, lifestyle choices (such as diet and exercise), hormonal factors, medical imaging results (mammograms, and more).).

By analyzing these factors together, doctors can assess an individual's likelihood of developing breast cancer at a given time. This type of risk assessment enables health care providers to implement proactive interventions such as routine testing, lifestyle changes, preventative medications or potential preventive measures such as risk-reducing surgery

Furthermore, breast cancer prediction models are constantly improving due to technological advances and the availability of large amounts of data. Machine learning and artificial intelligence are increasingly being used to sift through complex data and identify patterns that might go unnoticed by traditional methods

Ultimately, the goal of breast cancer prognosis is to equip individuals and their health care providers with the knowledge they need to make informed decisions about screening, prevention, and treatment, ultimately leading to better health and quality of life hipBreast cancer prognosis is an important health care intervention that aims to identify individuals at high risk of developing breast cancer. This process involves evaluation of many factors such as genetic predisposition, family history, lifestyle choices (such as diet and exercise), hormonal factors, medical imaging results (mammograms, and more).).

By analyzing these factors together, doctors can assess an individual's likelihood of developing breast cancer at a given time. This type of risk assessment enables health care providers to implement proactive measures such as routine testing, lifestyle modifications, preventative medications or preventable measures such as surgery to reduce risk

Ultimately, the goal of breast cancer prognosis is to equip individuals and their health care providers with the knowledge they need to make informed decisions about screening, prevention, and treatment, ultimately leading to better health and quality of life hip.

LITERATURE SURVEY

[1]Researchers like Sudarshan NayakB.M. Gayathri, Hiba Asri, Youness Khoudfi, Mohamed Bahaj, Latchoumiet TP, and Ahmed Hamza Osman have explored different machine learning techniques for breast cancer detection, with SVM consistently showing high accuracy rates Researchers like Sudarshan Nayak, B.M. Gayathri, Hiba Asri, Youness Khoudfi, Mohamed Bahaj, Latchoumiet TP, and Ahmed Hamza Osman have explored different machine learning techniques for breast cancer detection, with SVM consistently showing high accuracy rates .

[2] Noreen Fatima, Li Liu, Sha Hong,Haroon Ahmed This article is about machine learning methods for breast cancer prediction. It discusses various machine learning techniques and data mining techniques used for breast cancer diagnosis. The article compares the accuracy of these methods.

[3] Mohammad Monirujjaman Khan,corresponding author 1 Somayea Islam, 1 Srobani Sarkar, 1 Fozayel Ibn Ayaz, 1 Md. Mursalin Kabir, 1 Tahia Tazin, 1 Amani Abdulrahman Albraikan, 2 and Faris A. Almalki 3

This is an article about machine learning based comparative analysis for breast cancer prediction. It discusses machine learning models to improve breast cancer detection. Early detection is essential for successful treatment of breast cancer. The authors compare different machine learning models and find logistic regression to be the most accurate. They achieve an accuracy of 98%.

[4] Machine learning predicts breast cancer survival rates with debated methods. 31 studies focus on 5-year breast cancer survival prediction using ML. Varying prediction performance seen in studies like Sun et al., 2018 and Fu et al., 2018.

[5] Sajib Kabiraj; M. Raihan; Nasif Alvi; Marina Afrin; Laboni Akter; Shawmi Akhter SohagiThis article is about breast cancer risk prediction. It describes two machine learning algorithms for analyzing breast cancer databases. The algorithm achieved accuracy rates of 74.73% and 73.63%, respectively.

[6] The research paper explores the use of Machine Learning (ML) for classifying breast cancer into triple negative and non-triple negative types based on gene expression data, with Support Vector Machine showing the highest accuracy.

ML algorithms, particularly Support Vector Machine, offer a novel framework for accurate classification of breast cancer types, potentially complementing traditional diagnostic methods.

[7] Sara LaghmatiFaculty of science, Chouaib Doukkali University, El Jadida, Morocco; Bouchaib Cherradi; Amal Tmiri; Othmane Daanouni; Soufiane Hamida This paper presents a novel approach to achieve high accuracy and sensitivity using convolutional neural networks (CNNs) for breast cancer detection in digital mammography images. By using deep learning techniques, the study demonstrates the potential of CNNs to enhance early breast cancer detection and improve outcomes.

[8] Abunasser, Basem S; AL-Hiealy, Mohammed Rasheed J; Zaqout, Ihab S; Abu-Naser, Samy S.Various models and methods have been utilized to enhance the accuracy of Breast Cancer (BC) diagnosis, including Linear Regression (LR), Artificial Neural Network (ANN), K-Nearest Neighbors (KNN), Softmax Regression, Support Vector Machine (SVM), and Convolutional Neural Network (CNN) .

[9] Rhea D. Chitalia; Jennifer Rowland; Elizabeth S. McDonald; Lauren Pantalone; Eric A. Cohen ORCID logo ; Aimilia Gastounioti ORCID logo; Michael Feldman ORCID logo ; Mitchell Schnall; Emily Conant; Despina Konto In the research paper on breast cancer heterogeneity imaging phenotypes, the authors conducted a literature review to establish the significance of their study in predicting 10-year recurrence based on preoperative breast dynamic contrast-enhanced MRI scans.

[10] Hsiao-Chin Hong, Cheng-Hsun Chuang, Wei-Chih Huang, Shun-Long Weng,Chia-Hung Chen,Kuang-Hsin Chang, Kuang-Wen Liao,and Hsien-Da Huang This article examines the role of machine learning algorithms in breast cancer risk prediction, using a variety of data to increase accuracy and personalized risk assessment Using genetic, clinical, and imaging data together the study highlights the potential of standardized prevention strategies in breast cancer management and improves patient outcomes.

III. PROPOSED APPROACH

1. overview:

Using a hybrid model that combines machine learning algorithms with advanced imaging techniques, this approach aims to increase breast cancer prediction accuracy and perform personalized risk assessment The approach uses a variety of data combined with clinical, genetic and imaging data for comprehensive assessment of risk factors and early detection markers The authorization allows.

2. Data pre-processing:

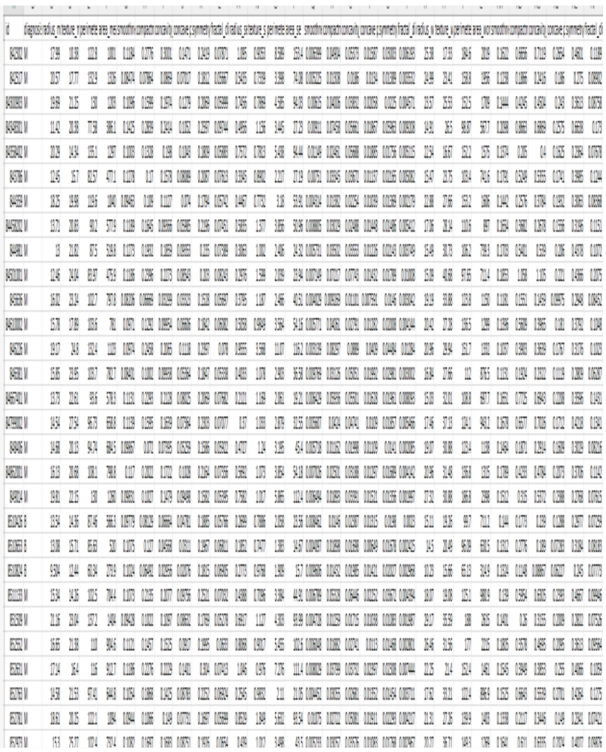
***Data Collection**: Collect relevant data including factors (characteristics) and target variables (diagnoses).*

***Data Cleanup**: Correct missing values, outliers, and any inconsistencies in the data set.*

***Feature selection/removal**: Identify relevant features that contribute to the forecasting task. This can be related to domain knowledge or statistical methods.*

***Data conversion**: Convert categorical variables to numbers (if necessary) using techniques such as one-hot encoding. Standardize or normalize numeric features to ensure they have the same dimensions.*

***Splitting the data**: Divide the data set into training and testing sets to check the performance of the model.*



4.Methadology:

The first step in the process of classifying breast cancer prognosis is careful data preprocessing. Various datasets including clinical records, genetic information and imaging results are collected, dealt with missing values, outliers and inconsistencies using stringent cleaning procedures and then deployed using feature engineering techniques extract relevant information from the raw data , where clinical parameters such as tumor size and tissue grade, genetic markers, and imaging characteristics are normalized or scaled, and categorical variables are encoded to ensure unity and consistency across the dataset .

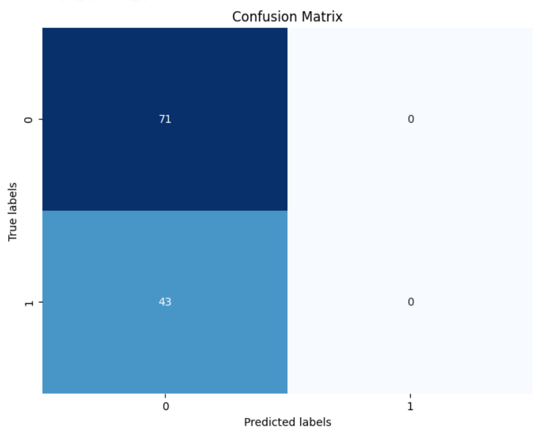
After data preprocessing, a number of classification algorithms are considered for model selection. Various classifiers such as logistic regression, support vector machines (SVM), decision trees, random forests, neural networks etc. are analyzed to determine the most suitable method for breast cancer prediction Factors such as data set size, feature space, . class distribution and interpretability are carefully weighed in the selection process. In addition, feature selection and dimension reduction techniques are used to identify the most suitable features and reduce the complexity of the model, thereby increasing efficiency and performance.

IV.EXPERIMENTAL RESULTS

Logistic regression:

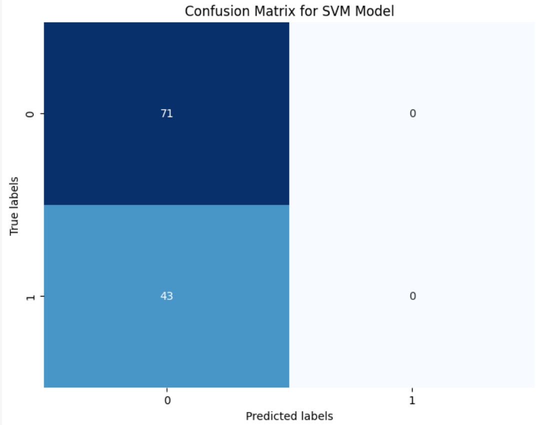
Accuracy: 0.6228070175438597

	precision	recall	f1-score	support
B	0.62	1.00	0.77	71
M	0.00	0.00	0.00	43
accuracy			0.62	114
macro avg	0.31	0.50	0.38	114
weighted avg	0.39	0.62	0.48	114



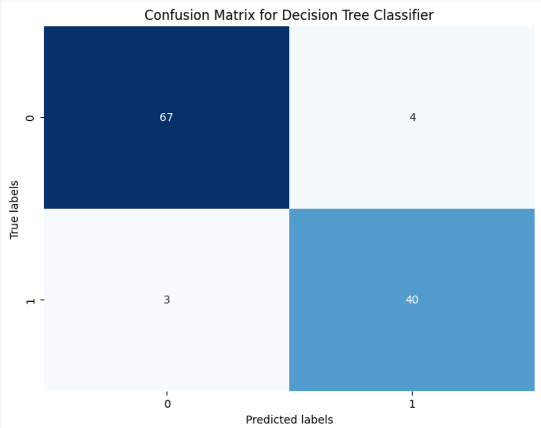
Support vector machine:

```
Accuracy: 0.6228070175438597
              precision    recall  f1-score   support

           B       0.62      1.00      0.77        71
           M       0.00      0.00      0.00        43

    accuracy                           0.62       114
   macro avg       0.31      0.50      0.38       114
weighted avg       0.39      0.62      0.48       114
```



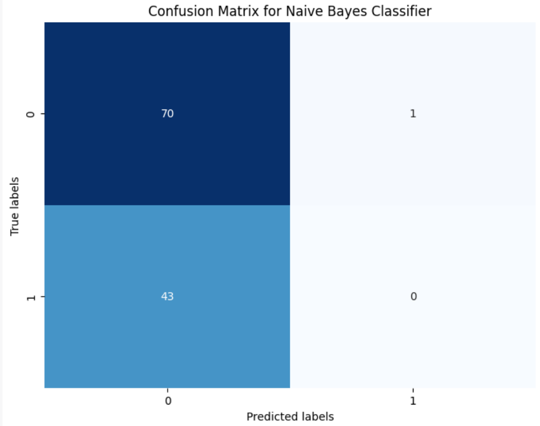
Decision Tree:

```
Accuracy: 0.9298245614035088
              precision    recall  f1-score   support

           B       0.94      0.94      0.94        71
           M       0.91      0.91      0.91        43

    accuracy                           0.93       114
   macro avg       0.93      0.93      0.93       114
weighted avg       0.93      0.93      0.93       114
```



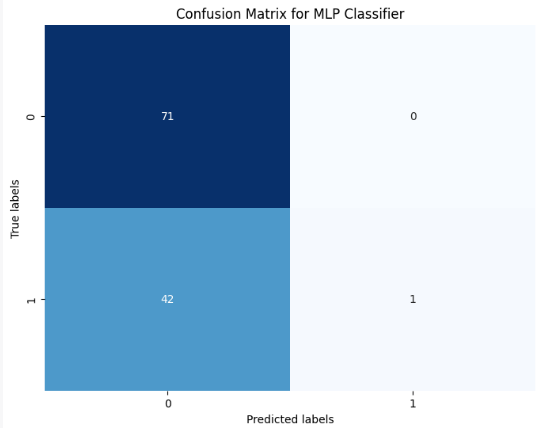
Naive Bayes:

```
Accuracy: 0.6140350877192983
              precision    recall  f1-score   support

           B       0.62      0.99      0.76        71
           M       0.00      0.00      0.00        43

    accuracy                           0.61       114
   macro avg       0.31      0.49      0.38       114
weighted avg       0.39      0.61      0.47       114
```



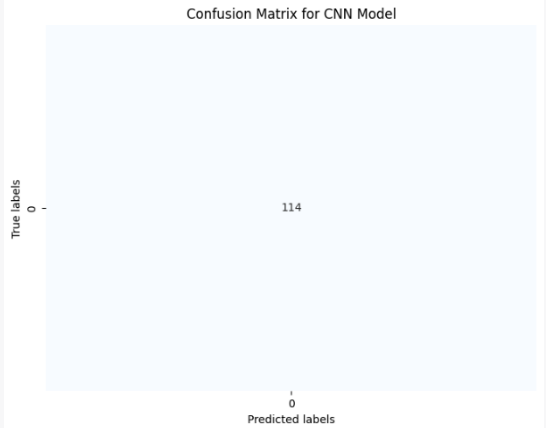
multi-layer perceptrons:

```
Accuracy: 0.6228070175438597
              precision    recall  f1-score   support

           B       0.62      1.00      0.77        71
           M       0.00      0.00      0.00        43

    accuracy                           0.62       114
   macro avg       0.31      0.50      0.38       114
weighted avg       0.39      0.62      0.48       114
```



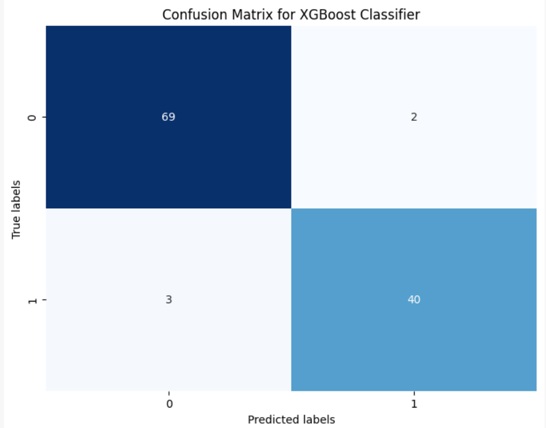
convolutional neural networks:

```
Epoch 1/10
15/15 [==============================] - 3s 9ms/step - loss: 0.4031 - accuracy: 0.8813
Epoch 2/10
15/15 [==============================] - 0s 6ms/step - loss: 0.1487 - accuracy: 0.9582
Epoch 3/10
15/15 [==============================] - 0s 9ms/step - loss: 0.1034 - accuracy: 0.9670
Epoch 4/10
15/15 [==============================] - 0s 14ms/step - loss: 0.0895 - accuracy: 0.9714
Epoch 5/10
15/15 [==============================] - 0s 7ms/step - loss: 0.0807 - accuracy: 0.9736
Epoch 6/10
15/15 [==============================] - 0s 7ms/step - loss: 0.0730 - accuracy: 0.9758
Epoch 7/10
15/15 [==============================] - 0s 5ms/step - loss: 0.0677 - accuracy: 0.9758
Epoch 8/10
15/15 [==============================] - 0s 6ms/step - loss: 0.0655 - accuracy: 0.9758
Epoch 9/10
15/15 [==============================] - 0s 5ms/step - loss: 0.0617 - accuracy: 0.9802
Epoch 10/10
15/15 [==============================] - 0s 4ms/step - loss: 0.0586 - accuracy: 0.9802
Test accuracy: 0.9649122953414917
```



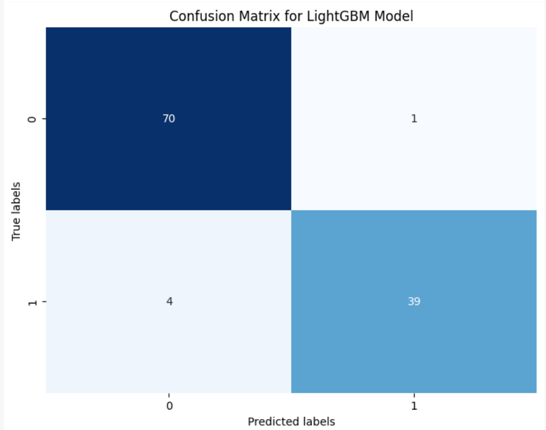
XGBoost:

```
Accuracy: 0.956140350877193
```



LightGBM:

```
Accuracy: 0.956140350877193
```



V. CONCLUSION:

Another meta-analysis of 25 studies from 1948, including the SOFT and TEXT clinical trials with breast cancer trials, provides further evidence for the wearing of contraception in postmenopausal women /removing them is associated with reduced recurrence and longer term, earlier postoperative, improved survival - stage of advanced breast cancer.

References

The research paper titled "Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis" by Mohammed Naji, Sanaa Filali, Kawtar Aarika, E Habib Benlahmar, Rachida Abdelouhahid, Olivier Debauche was published in Procedia Computer Science 191 (2021) 487-492 [1]

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.[2]

Monirujjaman Khan M, Islam S, Sarkar S, Ayaz FI, Ananda MK, Tazin T, Albraikan AA, Almalki FA. Machine Learning Based Comparative Analysis for Breast Cancer Prediction. J Healthc Eng. 2022 Apr 11;2022:4365855. doi: 10.1155/2022/4365855. Retraction in: J Healthc Eng. 2023 Aug 2;2023:9870523. PMID: 35449836; PMCID: PMC9017446.[3]

ML techniques have been employed to predict 5-year survival rates in breast cancer. A review of 31 studies revealed the use of decision trees, neural networks, and other ML methods for this purpose. These studies emphasized the importance of addressing class imbalances, managing missing data, and evaluating model performance using metrics like accuracy, sensitivity, specificity, and area under the curve.[4]

S. Kabiraj et al., "Breast Cancer Risk Prediction using XGBoost and Random Forest Algorithm," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-4, doi: 10.1109/ICCCNT49239.2020.9225451.[5]

"Deep learning-based breast cancer detection and prognosis: review and perspective" by Yanji Zhu et al., published in the 2020 Frontier Paper in Genetics.[6]

S. Laghmati, B. Cherradi, A. Tmiri, O. Daanouni and S. Hamida, "Classification of Patients with Breast Cancer using Neighbourhood Component Analysis and Supervised Machine Learning Techniques," 2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco, 2020, pp. 1-6, doi: 10.1109/CommNet49926.2020.9199633.[7]

S. Joo, Y. Yang, W. Moon and H. Kim, 'Computer-Aided Diagnosis of Solid Breast Nodules: Use of an Artificial

Neural Network Based on Multiple Sonographic Features'[8]

The study was supported by the NCI at the NIH through grants 5R01CA197000-03, R01CA223816-01A1, and U24-CA189523 [9]

Hong HC, Chuang CH, Huang WC, Weng SL, Chen CH, Chang KH, Liao KW, Huang HD. A panel of eight microRNAs is a good predictive parameter for triple-negative breast cancer relapse. Theranostics. 2020 Jul 9;10(19):8771-8789. doi: 10.7150/thno.46142. PMID: 32754277; PMCID: PMC7392022.[10]