International Journal on Engineering Technologies and Informatics

A Quick Review on Breast Cancer Prediction Using Machine Learning

Review Article

Volume 5 Issue 1- 2024

Author Details

*Jabeen Sultana**

Department of Computer Science, Majmaah University, Kingdom of Saudi Arabia

*Corresponding author

Jabeen Sultana, College of Computer and Information Sciences Majmaah University, Kingdom of Saudi Arabia

Article History

Received: December 23, 2023 Accepted: January 02, 2024 Published: January 04, 2024

Abstract

Breast Cancer, one of the prime reasons for causing deaths among females across the globe. Predicting and identifying indiscretions in breast certainly supports the concerned specialists to identify this deadly disease at primary stage can save the person's life. The people affected with this dreadful disease known as breast cancer are susceptible to complex health issues and the mortal rate has increased day by day. Diagnosing breast cancer is quite laborious task and involves careful observation to screen the breast to identify irregularities in breast because of many underlying features. In this review, breast cancer prediction and classification using Machine Learning is studied to diagnose the breast cancer as ML techniques yield promising results in classification, prediction and association rule mining. Further the review focuses on the results attained by the ML techniques on different kinds of data and results are reviewed in terms of accuracy, precision, recall and f-score.

Keywords: Machine Learning (ML), Breast Cancer (BC), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), Classification and Prediction

Introduction

This kind of cancer is the most prevalent in the globe, making up 12.5% of all new cases reported each year. The most prevalent malignancy in American women's diagnoses is BC. BC accounts for almost 30% of all newly diagnosed malignancies in women each year. One in eight American women, or almost 13% of all women, may eventually get invasive breast cancer [1]. The type of cancer is determined by distinct alterations occurring in the genes, which further impact the way specific body organs function. In general, there may be cysts that develop in both breasts and that have altered hormonal processes as their source. Cancer cells first appear as a little lump that does not spread to the tissues around it. There are four phases of cancer spread: at stage 1, cancer cells grow larger and spread to other tissues. Stage 2 cancer cells grow larger and infiltrate neighboring nodes; stage 3 cancer cells enter lymph nodes; and stage 4 cancer cells spread to other organs such as the brain, lungs, and bones.

BC comes in a variety of forms, primarily classified as invasive or noninvasive. Breast cancer in situ is another name for noninvasive breast cancer. Noninvasive cancer has not spread from the initial tissue, whereas invasive cancer has spread from the breast ducts or glands to other areas of the breast [2]. The use of ML techniques to diagnose BC and provide patients with appropriate care is highly desirable. Compared to normal processes, which take longer and include more tedious activities, ML saves time when classifying the kind of cancer [3]. The body of the rest of this paper is as follows: Section 2 gives a quick review of the literature that has been carried out for breast cancer prediction. Section 3 demonstrates the methodology used and the ML techniques that has been applied and the results attained followed by section 4 which concludes the review.

Review Work

A strategy to early detection of breast cancer using heterogeneous ensemble machine learning is presented in this research. The suggested method builds the ensemble model utilizing stacking and three distinct algorithms: KNN, SVM and DT. It adheres to the CRISP-DM procedure. With the lowest log-loss of 0.56 and the highest accuracy of 78% at K = 20, the suggested ensemble model successfully rejected the null hypothesis [4]. In this study, we show how feature selection can be done using XGBoost structures. Breast Cancer dataset has been collected from the UCI machine learning repository, which has 569 instances with 31 attributes. A classification method to predict

BC was proposed to classify breast cancer data into benign and malignant classes using XGBoost. It is employed as a feature selection approach in our case, where we extract the features from the dataset that it deems significant in an intuitive manner. These characteristics are employed in several classification methods. Testing is carried out to check the accuracy of the model built by the classifier. The results attained are assessed on various parameters like Accuracy, Sensitivity, Specificity, F-Measure etc. Result analysis reveals that important features yielded by XGBoost significantly increase the model accuracy by 98% as compared to accuracy without feature selection of 97.10% compared to different classification methods [5].

Numerous studies propose the use of various machine learning classifiers, such as logistic regression, decision trees, and radial basis functions, to diagnose and separate benign tumors from malignant tumors [6]. Various classifiers, including Support Vector Machine, Naïve-Bayes, Simple Logistics, Neural Network-MLP, Random Forest, and Decision Trees, are applied to the data set. After performing cross-validation, the model was trained and tested. A few factors are used to measure the findings and determine the correctness of the classification. Every classifier's obtained results are assessed. Comparing Support Vector Machine to Multi-Classifier, the former provides higher F-score and accuracy. In a different study, Senturk et al. employed a variety of machine learning classifiers, including SVM, linear regression, decision trees, naïve bayes, and MLP, to categorize and identify breast tumors [7]. In one of the studies done to categorize breast cancer, it was discovered that logistic regression had the best classification accuracy of 97% with an extremely low error rate of 0.14, followed by KNN with 95% accuracy, decision trees with 93.14% accuracy, and REP trees with 92.44% accuracy-the lowest accuracy of all [8].

Based on Wisconsin breast cancer data taken from the UCI ML repository, an effective method for identifying and proposing a cure for breast cancer was implemented. The results show that decision trees and naïve bayes fared better, achieving 93% and 95% classification accuracy, respectively [9]. The authors used more than eight machine learning classifiers to analyze breast cancer data in a similar type of study. Cross-validation was done, which entails training and testing the model forming ten folds. When the data were examined in terms of sensitivity, specificity, and classification accuracy, it was discovered that MLP performed better than decision trees, producing the greatest accuracy of 96% [10].

This paper's main goal is to determine the best imaging modalities and deep learning strategies that can process the large dataset and produce accurate predictions. Based on the review's findings, the main methods used to classify breast cancer are mammography and histopathologic imaging. Moreover, around half of the studies made use of public datasets, and the remaining half made use of private data sources. Future researchers that hope to enhance BC diagnosis models utilizing a variety of imaging modalities can benefit from the 13 major problems that the authors addressed throughout the review [11]. The below Table 1 shows the review carried out and the different ML techniques applied by the authors to classify the BC data and results were analyzed mainly in terms of accuracy [Figure 1].

Table 1: Showing Accuracy of ML Techniques applied on BC Data.

Classifier	Accuracy
Simple Logistics	97.18%
XGBoost	97.10%
SVM	96.92%
Random Forest	96.50%
MLP	95.25%
Decision Trees	93.99%
Multi-Classifier	93.14%
Naive Bayes	74.12%
PART	75.17%
KNN	65.26%

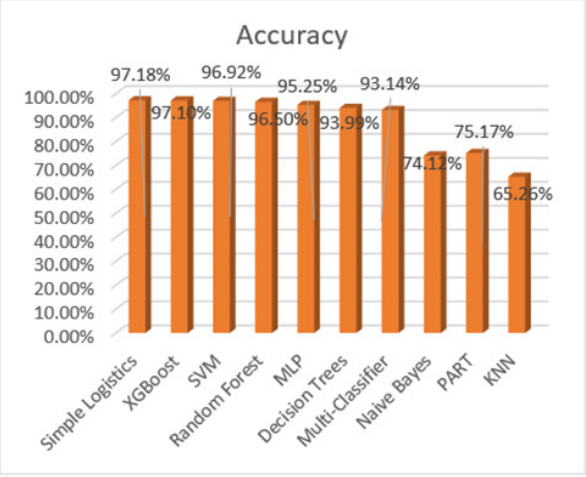

Figure 1: Showing Accuracy of ML Techniques applied on BC Data.

Conclusion

Globally, breast cancer is the most prevalent type of cancer. Genetic and inherited predispositions are among the several risk factors linked to the development of breast cancer. Breast cancers vary widely in type. Lung cancer has been surpassed by female breast cancer as the most prevalent cancer diagnosed globally. The death rate from breast cancer has declined recently in the majority of western nations, primarily as a result of improved early detection and treatment methods. In this paper, review is carried out on classifying BC data using ML techniques. ML techniques yield promising results in terms of classification, prediction and associating rules and extracting knowledge. Among all the ML techniques, MLP-a deep neural network and XG boost attained maximum classification accuracies of 99% and 97% approximately.

The prime cause for the spread of this breast cancer are observed with sedentary lifestyle, obesity, lack of physical exertion and excess intake of narcotics, hormonal abruptions, infertility, low breastfeed and genetic predisposition. There are two primary categories for women: both elevated and average risk. According to a screening research, females with thick breast are at higher risk to diagnose with this deadly disease and was observed that an additional 4.3 malignancies per 1000 instances when ultrasonography was added to mammography. Surgery, medication, and lifestyle changes are some of the choices available to women who are prone to this deadly disease and thereby can reduce that risk. The best method for lowering a woman's risk of breast cancer is a bilateral mastectomy, which is advised for all women.

References

1. https://www.breastcancer.org/facts-statistics.
2. (2020) Everything You Need to Know About Breast Cancer. Healthline.
3. Cruz JA, Wishart DS (2006) Applications of Machine Learning in Cancer Prediction and Prognosis. Departments of Biological Science and Computing Science. University of Alberta Edmonton AB 2: 2-21.
4. Nanglia S, Ahmad M, Khan FA, Jhanjhi NZ (2022) An enhanced Predictive heterogeneous ensemble model for breast cancer prediction. Biomedical Signal Processing and Control 72.
5. Sadaf K, Sultana J, Ahmad N (2023) An XGBoost-Based Classification Method to Classify Breast Cancer. In Applications of Machine Learning and Deep Learning on Biological Data pp: 75-86.
6. Chaurasia V, Pal S (2014) Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability. International Journal of Computer Science and Mobile Computing 3:10-22.
7. Senturk ZK, Kara R (2014) Breast Cancer Diagnosis via Data mining: Performance Analysis of Seven Different Algorithms. Computer Science & Engineering: An International Journal (CSEIJ) 4: 35-46.
8. Sultana J, Jilani (2018) AK Predicting Breast Cancer using Logistic Regression and Multi-Class Classifiers. International Journal of Engineering and Technology 7(4 20): 22-26.
9. Sultana J (2016) An Efficient Method to Diagnose the treatment of Breast Cancer using Multi-Classifiers. International Journal of Computer Science and Information Security 14.
10. Sadaf K, Sultana J, Jilani AK, Alabdan R (2019) Diagnosing Breast Cancer using Support Vector Machine and Multi-Classifiers. International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) 449-451.
11. Abhisheka B, Biswas SK, Purkayastha B (2023) A Comprehensive Review on Breast Cancer Detection, Classification and Segmentation Using Deep Learning. Arch Computat Methods Eng 5023-5052.