Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

Prediction of Breast Cancer Using Hybrid ML Techniques

Sheryl Catherine S. [1], Umadevi G. [2], Balaji Madhavan [3]

[1] *Student, Department of Computer Science and Engineering, Agni College of Technology, Chennai, Tamil Nadu India*

[2] *Assistant Professor, Department of Computer Science and Engineering, Agni College of Technology, Chennai, Tamil Nadu India*

[3] *Head of Department, Department of Computer Science and Engineering, Agni College of Technology, Chennai, Tamil Nadu India*

Abstract:- Breast Cancer is one of the most common diseases that occurs in Indian residents. It ranks fourth in the top ten cancers in America. Every four minutes, a woman in India is affected by this deadly disease, according to the statistics. The women living in the suburbs are more likely to get breast cancer than in the past. One in twenty-eight women in India are diagnosed with breast cancer. Urban women (1 in 22) suffer more from it than rural women (1 in 60). Projections for 2023 indicate that the United States alone will witness 1,958,310 new cancer cases and 609,820 cancer-related death. These deaths can be avoided if these tumours can be detected early. This project gives a strategy for detecting breast cancer using hybrid ML techniques. The main goal is to predict breast cancer from an existing input dataset, which has all the necessary information about Benign and Malignant cancer. This system uses two algorithms Decision Tree and Naïve Bayes.

Keywords*:* Breast Cancer, Benign, Malignant, Diagnosis, Machine Learning, Decision Tree, Naïve Bayes.

1. Introduction

Breast cancer is the most prevalent cancer. In 2020, around 2,300,000 new cases were identified worldwide, leading to approximately 688,000 fatalities. Projections for 2023 indicate that the United States alone will witness 1,958,310 new cancer cases and 609,820 cancer-related death. The occurrence of cancer varies depending on the country, region, ethnicity and lifestyle. Traditional risk prediction models use statistical approaches combined with the details of the patient to predict the risk, recurrence, and survivability of breast cancer (BC). While modest in performance, these techniques often suffer from racial bias. Machine Learning techniques have been used to predict the possibility of breast cancer. The Decision Tree and Naïve Bayes algorithms have been used to predict the possibility of a tumor present in this project. With these algorithms the model is going to predict the possibility of a tumor being present whether benign one or a malignant one. The algorithms used can be used to predict and the best one maybe defined by the accuracy of the prediction. In this project, the algorithms are made into a hybrid model. In this model, both decision tree and naïve bayes algorithms can be combined to form a hybrid model. This hybrid model increases the accuracy of the prediction as it a combination of two algorithms.

2. Literature Survey

All prior research on machine learning algorithms that have been employed for breast cancer prediction has been mentioned by the authors in [1].

It gives all information regarding the most effective ML Algorithms like K-Nearest Neighbours (K-NN), Support Vector Machines (SVM), Naive Bayes, and Random Forest. According to the results of the studies, SVM has a maximum accuracy of 97.9%. This pronouncement will help researchers choose the best machine-learning classification strategy for predicting breast cancer. Also obtained are the Wisconsin Breast Cancer

Dataset (WBCD) and Logical Regression and Decision Tree. The authors explained the hybrid strategy to increase the projected semi-structured sequential data performance [2].

The complete implementation of Logical Regression and the mathematical equations is found in [3].

Depending on the dataset and parameter selection, each method performs differently. The Logistic Regression's accuracy, sensitivity, and precision are 95.71 percent, 95.74 percent, and 97.82 percent, respectively. The data collection consists of 699 patient records from [4].

Two hundred forty-one of them, or 34.5 percent, have breast cancer, while the remaining 458, or 65.5 percent, do not have cancer. The Decision Tree (J48) delivers an accuracy of 95.59 percent in training data and 92 percent in testing data. The prediction findings of the comparison of the six standard data mining methodologies are validated using a 10-fold crossover test. The k-fold crossover validation is widely used to reduce the error caused by random sampling when comparing the accuracies of several prediction models. The accuracy of classification techniques is assessed using the attributes of mammogram images that have been chosen. For the analysis, three popular data mining methods, J48, AD Tree, and CART, were used. J48 has an accuracy of 98.1 percent, ADTree has a value of 97.7 percent, and CART has the highest accuracy value of 98.5 percent. According to the results, the CART algorithm performs well for classifying mammogram images. The ADTree algorithm performs poorly, and J48 serves as an intermediary [5].

3. Opportunities

Early Intervention and Support: When Breast Cancer is identified early on, afflicted people and their families can get prompt intervention and services.

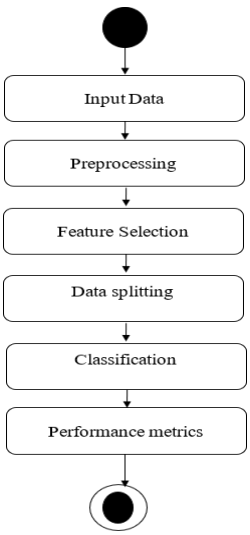
Customised Treatment Plans: ML algorithms are able to examine a variety of datasets in order to pinpoint the type of breast cancer. This allows for the creation of customised treatment programmes that are suited to each patient's needs.

Increased Diagnostic Accuracy: By combining data from several sources and spotting behavioural indicators that conventional tests might miss, machine learning approaches might improve the precision and efficiency of diagnosis.

Clinical Decision Support Systems: ML models can act as evidence-based decision support systems for Breast cancer diagnosis, treatment planning, and monitoring, supporting medical practitioners in their decision-making.

Public Health Initiatives: In order to address the rising prevalence of Breast Cancer and lessen gaps in access to diagnostic services, public health initiatives can benefit from the use of ML-driven Breast Cancer detection programmes.

4. Methodologies



The project follows the flow of the flowchart above. A few classes are then used to test and train the model in order to get accurate results. Decision Tree and Naïve Bayes, the well-known supervised learning algorithms, are used to train the model.

STEP1: Data Selection

Data selection is where the appropriate data type and source, as well as the suitable instruments to collect data are determined.

STEP2: Data Preprocessing

Data preprocessing is where raw data is transformed into an understandable form. It is one of the most important steps in data mining as raw data is not something that we can work with.

STEP3: Feature Selection

Feature selection in machine learning is the process of finding the best set of features that allows one to build useful models of studied phenomena.

STEP4: Data Splitting

Data splitting is the process of dividing data into two or more subsets. Usually, with a two-part split, one of the part is used to evaluate or test the data and the other is used to train the model.

STEP5: Classification

It is necessary to combine two different machine learning approaches, such as logistic regression and decision tree, to complete this operation.

STEP6: Performance Analysis

The ultimate outcome will be based on the overall classification and prediction.

The dataset is obtained from Kaggle which is used in training the model.

5. Conclusion

In conclusion, Breast cancer is a condition that affects many women throughout the world. Malignant and benign tumors are the two forms of breast cancer. Classifying various tumor types into relevant classifications is critical in assisting physicians in prescribing the best treatments. This paper uses Decision Tree and Naïve Bayes to provide a machine learning strategy for breast cancer prediction. Our main goal is to identify a suitable method for more efficiently predicting the occurrence of breast cancer using hybrid approaches. In this work, the performance of the suggested technique was evaluated using the Wisconsin Breast Cancer dataset. In this project, both the decision tree and naïve bayes algorithms are used in the hybrid technique to get better accuracy in predicting the occurrence of breast cancer. The obtained accuracy of the proposed method was 93.01%.

6. Acknowledgement

I appreciate the support for this study from Dr.Balaji Madhavan, Head of the CSE Department and Mrs. Umadevi.G ., Assistant Professor, CSE Department, Agni College of Technology, Chennai.

Reference

[1]. A. A. Bataineh, ‘‘A comparative analysis of nonlinear machine learning algorithms for breast cancer detection,’’ Int. J. Mach. Learn. Comput., vol. 9, no. 3, pp. 248–254, Jun. 2019.
[2]. Aalaei S, Shahraki H, Rowhanimanesh A, Eslami S (2016). Feature selection using genetic algorithm for breast cancer diagnosis: experiment on three different datasets. Iran J Basic Med Sci, 19, 476-82.
[3]. Anusha Bharat, Pooja N, R Anishka Reddy “Using Machine Learning algorithms for breast cancer risk prediction and diagnosis”, 2018 8, IEEE Third International Conference on Circuits, Control, Communication and Computing

[4]. A. R. Chaudhury, R. Iyer, K. K. Iychettira, and A. Sreedevi, ‘‘Diagnosis of invasive ductal carcinoma using image processing techniques,’’ in Proc. Int. Conf. Image Inf. Process., Nov. 2011, pp. 1–6.
[5]. B. Padmapriya and T. Velmurugan, “Classification algorithm-based analysis of breast cancer data," Int. J. Data Mining Techn. Appl., vol. 5, no. 1, pp. 43–49, Jun. 2016.