iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52012

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Breast Cancer Using Machine Learning

S. Rajayogha
Department of Computer Science and Information Technology, Kalasalingam Academy Of Research And Education.

Abstract: ***Women die from breast cancer, which is an abstract concept. Breast cancer is the most important problem. The most frequent cancer in women diagnosed globally has now surpassed lung cancer in prevalence. early detection that aids in cancer prevention. If breast cancer is to have a very high survival rate, it must be found in its earliest stages. The efficient machine learning method is utilized to categorize the data. Methods are employed in the medical field to aid in diagnosis and decision-making. This study used the Wilcoxon breast cancer dataset to do data visualization and compare various machine learning methods, including the Support Vector Machine (SVM), Decision Trees, Naive Bayes (NB), K Nearest Neighbours (K-NN), Adaboost, Xgboost, and Random Forest. The primary goal is to assess the data's accuracy in terms of each algorithm's efficiency and effectiveness in terms of accuracy, precision, sensitivity, and specificity. Our goal is to use machine learning to detect things quickly, effectively, and precisely. The experimental findings had the lowest error rate and the best accuracy (98.24%).***
Keywords: Wilcoxon, algorithms, machine learning, and detection of breast cancer.

I. INTRODUCTION:

The World Health Organization (WHO). In 2021, there will be about 963,300 deaths of women. It may increase to 2.9 million, according to the organization. Males can potentially develop breast cancer, in addition to females. Every four minutes, an Indian woman is given a breast cancer diagnosis. Breast cancer is a frequent and severe disease that can affect both men and women. As soon as the signs are recognized, it quickly progresses through the initial stage. The cells that make up this malignancy are genetically altered and aberrant cells enter these cells. is fatal after diagnosis and treatment since it spreads throughout the body. Breast cancer comes in two flavors: benign and malignant. The first is categorized as damaging and malignant, with the potential to spread to other organs. Benign is categorized as non-cancerous. Breast cancer affects women's chests, specifically the glands and milk ducts; it frequently spreads to other organs and may do so via circulation. Breast cancer is detected using a variety of methods, including biopsies, computerized thermography, and ultrasound sonography (Histological images). Patients with modest and undetectable malignancy indicators can have diagnostic mammography performed to evaluate aberrant breast cancer tissue. This method cannot be utilized to evaluate places where cancer may be suspected because of the sheer volume of photos. In examinations of women with particularly dense breast tissue, about 50% of breast tumors were not found, according to a report. Nonetheless, within two years of screening, roughly 25% of breast cancer patients receive a negative diagnosis. Thus, it is essential to make an early and prompt diagnosis of breast cancer. Many mammography-based breast cancer screenings are done regularly for all women, typically once a year or every two years.

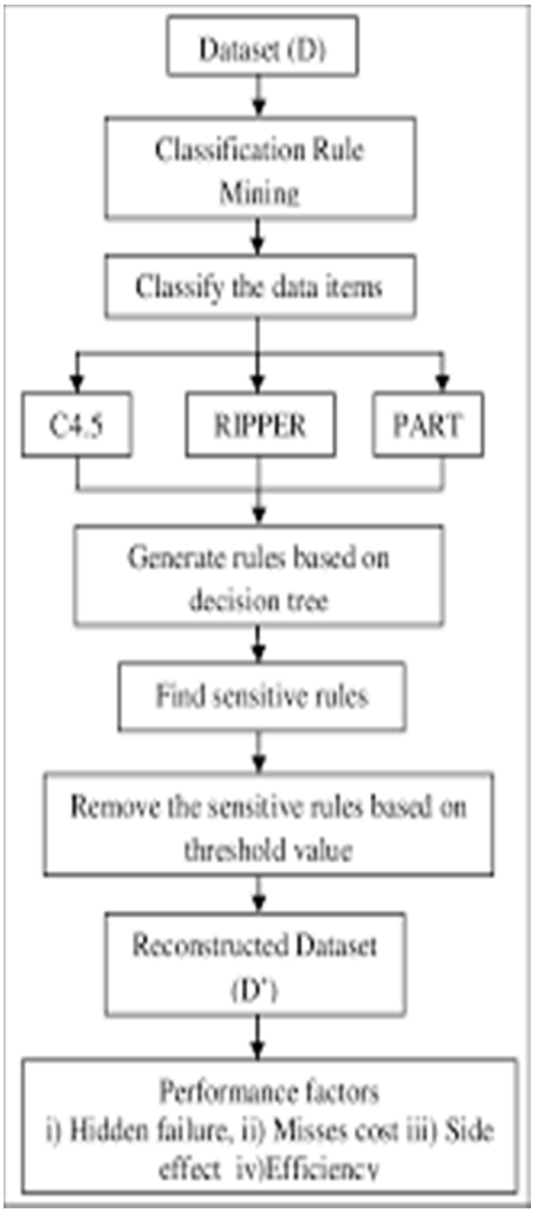
II. LITERATURE SURVEY:

1) Turgut Machine Learning process evaluated in comparison to SVM, KNN, DT, Logistic Regression, Random Forest, and ADA Boost. According to this analysis of numerous methods, the random forest has the highest efficiency at 89%.
2) Narasingarao.M. provides an overview of the research done to identify breast cancer using several algorithms and draws conclusions on the effectiveness of the algorithms.
3) Using Adaptive Reasoning Theory and the Wisconsin data set, which has 569 rows of data and 32 attributes, Junaid Ahmed was able to obtain an accuracy of 84.21%.
4) For the various datasets, Nithya used the three categorization techniques known as Decision Tree, k-Nearest Neighbor, and Naive Bayes. The authors additionally look at the error rate evaluation measures. The implementation concentrated on a certain dataset attribute type.
5) Python was used to develop the technique, which was evaluated using a dataset and yielded an accuracy of 94.74 percent while also speeding up the process. Shilpa M. and C. Nandini.
6) Hafizah compared SVM and ANN using four different breast cancer-related datasets. The study's findings demonstrated that SVM outperformed ANN in terms of performance and output. Among the features, G. S. Gc tried to extract was variance, range, and compactness. SVM classification was performed to analyze the performance.

©IJRASET: All Rights are Reserved

Their research revealed the highest variance (95%) and compactness (86%) of any study. SVM can be regarded as a suitable strategy for breast cancer prediction considering their findings.

III. ARCHITECTURE



IV. METHODOLOGY

A. Dataset Description

We got the Breast Cancer Wisconsin (Diagnostic) Dataset from Kaggle. Here, 570 patient records were employed for the analysis, and each instance had 42 attributes along with a diagnosis and features.

Every instance contains a parameter of cancerous and non-cancerous cells, and we can forecast cancer simply by inputting attributes. The values for the features are shown in numerical format. The term "Target" refers to a patient who is suffering from either

benign or malignant cancer. Benign indicates that the patient has no cancer, and by the input of features. The values of features is in Numeric Format. The ‘Target’ means the patient Who is having Whether ‘Benign’ or ‘Malignant’ Cancer state. Benign means the patient is not having Cancer and Malignant means the patient is having Cancer.

B. Data Visualization

We are going to Visualize our Numeric data with Respect to Two categories 1) Benign 2) Malignant

C. *Section Headings*

We used Google Collab as a Coding platform and get a prediction output from the Flask in Local Server. Our Methods Includes Supervised Learning Algorithms and Classification Techniques like Support Vector Classifier (SVM), Random Forest, Naïve Bayes, Decision Tree, and KNN. Dataset contains features which highly vary in units and magnitudes. So, it is required to bring all features to the same level of magnitudes. We did that by using Standard Scaling in SKLearn. Model selection is the most important step in Machine Learning. Machine Learning algorithms can be classified as: supervised learning and unsupervised learning. For Our project, we only need supervised learning. We used all Methodologies to Predict the result and Noted their Accuracy

D. Confusion Matrix and Accuracy

A confusion Matrix is used for evaluating the performance of a classification model. The Matrix compares the actual target values with predicted values by machine learning. It shows the ways in which your classification model gets confused When it makes predictions.

Accuracy is given by:

Accuracy=(TP+TN)/(TP+TN+FP+FN)=(46+66)/(46+66+0+2)*100=98.24

Where TP=True Positive,TN=True Negative,FP=False positive,FN=False Negative

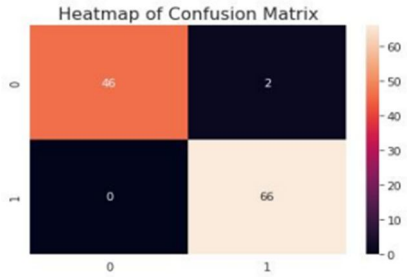
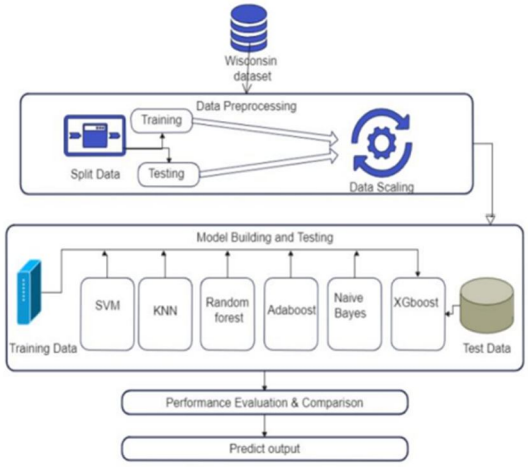


Fig. 2 Confusion Matrix

V. PROPOSED SYSTEM ARCHITECTURE :

As Shown in diagram, we first Uploaded dataset From Wisconsin Breast Cancer Dataset. After that We did Preprocessing to the data and applied Machine Learning Models, which is used in this project to predict Breast cancer.



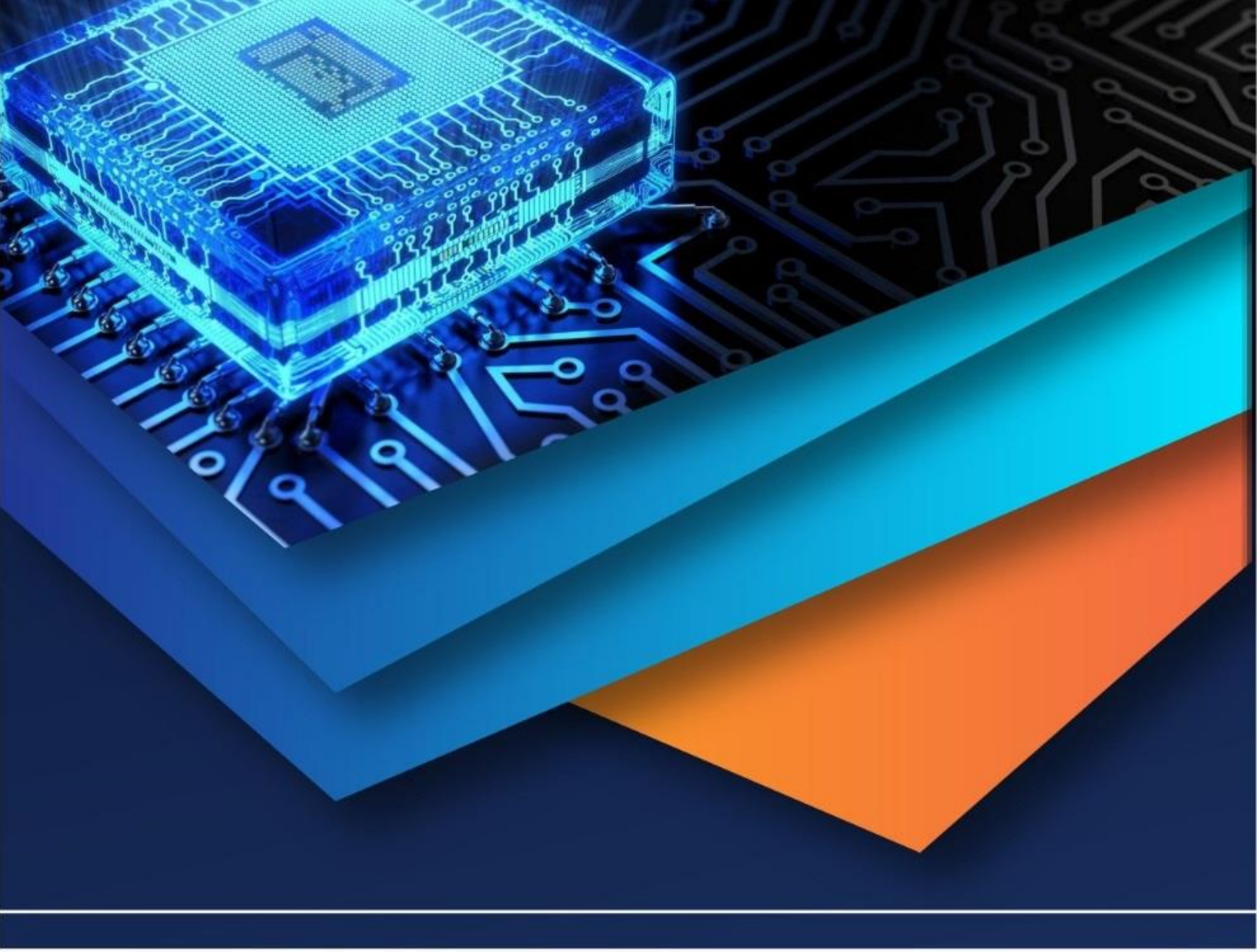
VI. CONCLUSION & FUTURE WORK

This paper examined different machine learning techniques for detection of breast cancer.

REFERENCES

[1] S. Gc, R. Kasaudhan, T. K. Heo, and H.D. Choi, "Variability Measurement for Breast Cancer Classification Mammographic adaptive and convergent systems (RACS), Prague, Czech Republic, 2015, pp. 177–182.
[2] S. Hafizah, S. Ahmad, R. Sallehuddin, and N. Azizah, "Cancer Detection Using Artificial Neural Network and Support Vector Machine: A Comparative Study," J. Teknol, vol. 65, pp. 73–81, 2013.
[3] A. T. Azar, and S. A. El-Said, "Performance analysis of support vector Neural Compute. Appl., vol. 24, no. 5, pp. 1163–1177, 2014.
[4] machines classifiers in breast cancer mammography recognition," Neural Comput. Appl., vol. 24, no. 5, pp. 1163–1177, 2014.
[5] C. Deng, and M. Perkowski, "A Novel Weighted Hierarchical Adaptive Voting Ensemble Machine Learning Method for Breast Cancer 2015.
[6] Z. Jiang, and W. Xu, "Classification of benign and malignant breast cancer based on DWI texture features," ICBCI 2017 Proceedings of the Iinternational Conference on Bioinformatics and Computational Intelligence 2017.
[7] R. Jegadeeshwaran and V. Sugumaran (2013) Comparative study of decision tree classifier and best first tree classifier for fault diagnosis of automobile hydraulic brake system using statistical features, Measurement, vol.46, pp.3247–3260.
[8] Ajith Abraham (2005), Artificial neural networks, Nature & scope of AI techniques, vol.2, pp.901-908.
[9] Jennifer Listgarten, Sambasivarao Damaraju, Brett Poulin, Lillian Cook, Jennifer DuFour, Adrian Driga, John Mackey, David Wishart, Russ Greiner and BrentZanke (2004), Predictive Models for Breast Cancer Susceptibility from Multiple Single Nucleotide Polymorphisms, Clinical Cancer Research, vol.10, pp.2725- 2737.
[10] Jaree Thongkam, Guandong Xu and Yanchun Sang (2008), Breast cancer survivability via AdaBoost algorithms, Health data and knowledge management, vol.80.
[11] V. Sugumaran, V. Muralidharan and K.I. Ramachandran (2007), Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing, Mechanical Systems and Signal Processing, vol.21

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)