

Identification and Diagnosis of Breast Cancer At Different Stages By Different Machine Learning Algorithms On The Coimbra Dataset

Manish Tiwari¹, Nagendra Singh^{*2}, Harsh Pratap Singh³, Lokendra Singh Songare⁴, Pinky Rane⁵, Shubha Soni⁶, Rakesh Pandit⁷

¹Department of computer science engineering,
Career point university,
Kota, Rajasthan
immanishtiwari@gmail.com

²Department of Electrical Engineering,
Trinity College of Engineering and Technology,
Karimnagar, Telangana, India
nsingh007@rediffmail.com

³Department of Computer Science and Engineering,
Shri Vaishnav Vidyapeeth Vishwavidyalaya,
Indore MP, India,
drharshprataps@gmail.com

⁴CSE Department,
Medi-Caps University, Indore,
lokendra.songare@gmail.com

⁵Department Computer Science and Engineering,
Medi-Caps University,
Indore, MP, India
pink.shinde@gmail.com

⁶Department Computer Science and Engineering,
Institute of Engineering and Technology Sagar, M.P.,
shubhasoni01@gmail.com

⁷Department Computer Science and Engineering,
Medi-Caps University,
Indore, MP, India
rakesh.pandit@medicapas.ac.in

*Corresponding Authors : Nagendra Singh, Email : nsingh007@rediffmail.com

Abstract— Cancer is the most deadly disease in the world. Breast cancer is the second-most common disease in women worldwide. It is the most common cancer globally among women. Annually, 12.5% of all new cancer cases worldwide Globally, 2.26 million breast cancers were discovered, and 685,000 women died from this disease. Early diagnosis of breast cancer is more difficult in developing countries than in developed countries. Using technology, if it is possible to detect cancer early and treat it on time, then many women can be cured and their lives can be saved. Early detection also leads to an increased survival rate for patients who receive clinical therapy before reaching later stages. It includes a number of risk factors, such as modifiable and non-modifiable ones. A recent survey discovered that for women above 50 years of age, the chance of getting breast cancer is about 80%. Machine learning algorithms are playing a major role in diagnosing liver cancer in its early stages and helping doctors make prompt decisions. A number of machine learning models have been executed in which the model gave better performance in terms of accuracy, and other parameters such as precision, recall, etc. are used to predict early. In this research work, the latest dataset, Coimbra, belongs to UCI machinery. It has nine features (age, BMI, glucose, insulin, HOMA, leptin, adiponectin, Resistin, MCP.1) and one classification attribute, which has values 1 and 2. 1 belongs to benign, and 2 belongs to malignant. Based on that, the supervised machine learning algorithm was applied. The WEKA tool is used to analyze the dataset. A number of algorithms are applied, such as Bayes net, multilayer perceptron, IBK, random committee, random tree, etc. More of them gave better results, and that model was chosen as the key model for breast cancer analysis.

Keywords- Cancer, breast cancer, deep learning, WEKA tool and detection of stage of cancer

I. INTRODUCTION

In the human body, when cell growth occurs uncontrollably and slowly spreads to other parts of the body, cancer disease arises. Cancer may happen anywhere in the human body. But in the case of women Breast cancer is a very common problem. In the last decade, breast cancer has been the most responsible for the deaths of women compared to other types of cancer. As per the World Health Organization (WHO), many millions of women were diagnosed with breast cancer, and out of them, many died due to a lack of medical support. The main problem is the identification of cancer in its early stages. In most cases, women don't know anything about breast cancer [1].

Breast cancer starts in the breast tissue due to abnormal cell growth. A tumor develops in the form of a lump because of the growth of abnormal cells. Tumors may first resemble other normal cells, and damaged cells may not exhibit any aberrant signs. When tumors develop due to the uncontrolled and continuous growth of cells, they spread very fast in the body. The affected cells' progress starts damaging the tissues and organs surrounding them, thereby causing breast cancer. And the development of such a tumor can be fatal [2].

Invasive and noninvasive are the main two types of breast cancer. Invasive types of cancer can start in the glands and slowly spread to whole parts of the breast. Noninvasive types of cancer cannot spread due to the original tissue. Doctors can identify the various stages of cancer (Stage 1 to Stage 4) according to the growth and size of the tumor as well as the spread rate in the near part of the breast. The early stage, or primary stage, is defined as stage 1, and at that stage, the size of the tumor is around 2 cm. The last stage is stage 4, in which tumors, can have any size, the spread rate is high, and they damage the nearby cells and tissues of the breast [3].

If cancer is detected in its primary stage, it is possible to save the lives of infected people. But if it is identified at the last stage (stage 4), there is the least chance of saving the lives of the people. Regular checkups and screenings help identify the early stages of cancer. With regular screening, clinical experts can diagnose the early stages of cancer. Biomarkers and imaging tests were favoured by clinicians for determining the stages of cancer. Scientists investigated some more useful breast imaging technologies that are effective in finding breast cancer in its early stages, like positron emission mammography, scintimammography (molecular breast imaging), Elastography, and Electrical impedance imaging (EIT)[4].

Breast cancer detection is a challenging procedure, and since there is a lack of awareness, many affected women do not receive a primary stage diagnosis, making it challenging to save their lives. If it is identified in the early stages, it is possible to prevent its progress and hence reduce the risk of

death. So identification of cancer in its early stages and the use of clinical treatments can improve the survival rate [5].

For survival for a long time, it is very essential that the detection of cancer be in the very primary stage. Cancer is curable, and doctors can save the lives of women, but there is difficulty in detection in the early stages. Once it is identified with proper treatment and supervision, patients will survive for a long time, and even many patients will cure properly [6]. Breast cancer is more dangerous for women compared to other types of cancer. Many studies show that most women's death cases increase due to breast cancer [7].

For the proper treatment of breast cancer, an accurate examination is very important. It is easier for medical experts to treat patients if the stage (1-4 stages) of cancer is correctly identified. If it is identified after the first or second stage, most of the cells and tissue are damaged, and it is very difficult to control the infection and save the life. Machine learning techniques have many novel features, and they are used for the early identification of breast cancer. Machine learning has the ability to find the correct results from complex data sets [8].

1.1 Types of breast cancer

It is mainly categorized into the following types

Ductal Carcinoma In Situ (DCIS): this is the initial stage of cancer. In this case of cancer, it has not spread to other parts of the breast.

Invasive Ductal Carcinoma (IDC): IDC is the most dangerous and common breast cancer. Out of 10 women, 8 are infected by this cancer. This type of cancer starts in the milk duct, but after some time it breaks the duct wall and starts to spread to another part of the breast.

Invasive Lobular Cancer (ILC): ILC is also dangerous because identification of this cancer is also very difficult by physical examination or imaging (mammograms), and it infects both breasts of the woman. One woman out of five suffers from this cancer.

Triple Negative Breast Cancer: As the name suggests, this cancer shows three times negative results for estrogen or progesterone receptors and also does not formulate high protein (HER2). This mostly arises in women who are older than 40 years of age. This cancer accounts for around 10–15% of all breast cancer. It grows and spreads very quickly. The treatment facilities for triple-negative breast cancer are very limited.

Inflammatory Breast Cancer (IBC): This cancer arises due to cancer cells blocking lymph vessels in the skin. The initial symptoms of this type of cancer are swelling and redness in the breast. The breast size may be inflamed. This kind type of breast cancer is seen as very rare. Approximately 1-5% of women suffered from this cancer.

This cancer has generally seen the younger women aged less than 40 years. It is more dangerous for black women.

II. LITERATURE REVIEW

The researcher employed a variety of techniques to identify breast cancer in the literature. The recommended methods for identification of breast cancer and their results are shown in Table 1.

Table 1: Methods used for breast cancer identification and outcomes of different studies

Ref. no.	Methodology	Outcomes
1	Machine learning (multilayer perceptron)	Early stage breast cancer identification(Breast Cancer Coimbra Dataset)using biomarker
3	PR Method	This work uses biomarkers to accurately detect the cancer percentage.
6	Decision Trees, ANN, Vector Machine, and K-Nearest Neighbour	These methods were used to identify the cancer stages.
8	VGGNet and GoogLeNet	Breast cytology images method used for the identification of breast cancer.
9	RSQA and AQ	They used set of data for identification of cancer stages.
10	ML(machine learning)	Breast cancer at early stages is identify by ML Method
11	SVM, k-NN, Adaboost, XGboost, Decision Tree, NB, and Random Forest	For identification of cancer used Wisconsin data set.
12	Deep neural network classifiers, SVM and DT(Decision tree)	For the prediction of breast cancer used blood investigations and anthropometric information.
13	SVM, k-NN, Adaboost, XGboost, Decision Tree, NB, Logistic Regression, Ridge Classifier, DTC and RF	For different stages of breast cancer, the authors used the Wisconsin Data Set.
14	Mammography screening techniques	Breast self-examination and screening practice are done.
15	MLP, RF(Random Forest), DT(Decision Tree) and SVM	For the prediction of breast cancer stages, data mining was used.
16	DET	Wisconsin and Coimbra datasets are used for the diagnosis of breast cancer.
17	NN, SVM, DT, Random Forest and LR	Coimbra and Wisconsin datasets are used for the diagnosis of breast cancer.
18	SVM and randomized trees classifier	The proposed combined methods give 80.23% accuracy in the identification of breast cancer.
19	SVM, RF, GB, ANN, and MPs	Coimbra and Wisconsin datasets are used for the identification of breast cancer.
20	SVM	Glucose, resistance, age, and BMI samples are used for predicting initial-stage breast cancer. The proposed method gives 85–90% accuracy in identifying the cancer.

III. RESEARCH METHODOLOGY

First Coimbra dataset was chosen from the authentic repository. The preprocessing was done on the chosen dataset using Randomizable interfaces, supervised filter, and Weighted attributes handler. Weka tool is used in this two methods are followed one is the data with replacement and the other one is the data without replacement at the implementation level. Classifiers are chosen and then divide the data as training and testing dataset applied to a number of algorithms on this dataset. Find out the best performance such as Accuracy, precession, recall, kappa statistics, etc. the algorithms provide the best result among all algorithms chosen as the final model for analysis..

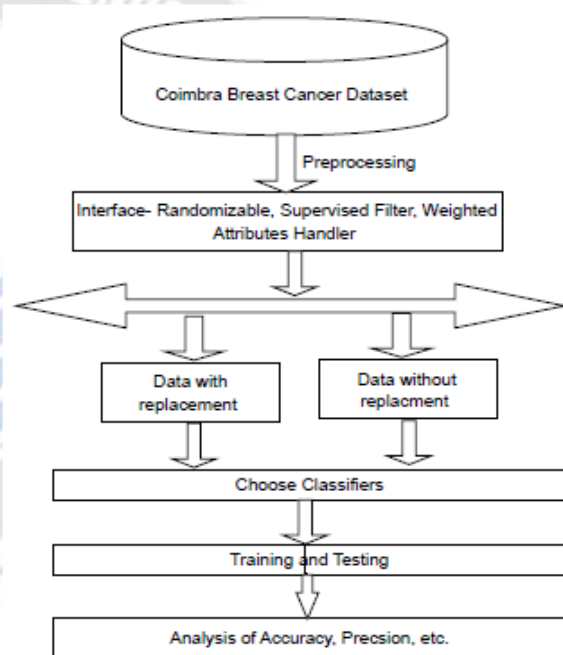


Figure1:- Methodology for analysis of dataset

3.1 Dataset Description

Numerous datasets are available on breast cancer. Many research works have been done on these datasets. Less research has been done on the Coimbra dataset. This dataset is chosen from the UCI machinery library. It has a multivariate dataset with 116 instances and a total of 10 features. Among 116 examinations 64 belong to breast cancer patients and the remaining 52 belong to healthy controls. Among these 10 features nine are independent variables such as age, BMI, Glucose, Insulin, HOMA, Leptin, Adiponectin, Resistin, and MCP-1 and 11th one is classification 1 for the absence of disease and 2 represent disease[21]. Metabolic Syndrome, particularly insulin resistance and fat women on the abdominal side is detected after 12 months of menopause chances of breast cancer. MCP.1 is one more way to say macrophage-related migration can lead to breast cancer evolution [23]. Leptin has

endocrine, paracrine, and autocrine effects on breast cancer cells in addition to these other mechanisms.

3.2 Tool to be used

The WEKA tool is used for data mining and analysis for clustering and classification purposes. Many different algorithms are available for the analysis of the different datasets. It is open-source and free software. The workbench of the WRKA includes machine learning algorithms for data mining tasks. It has the capability to perform clustering, classification, and association rules efficiently, from the preparation of data to the analysis and visualization of data. It supports different types of file formats like .arff, .csv, etc. It has the capability to perform the preprocessing easily.

Performance Parameter: A validation model is used to gain acceptance. To get realistic and reliable results, random sampling techniques are used, and repetition is done to get better results. The result in terms of accuracy, precision, and recall is calculated from confusion metrics. For the analysis, the ROC curve is used to test the estimation. The ROC curve is the true positive rate (sensitivity) plotted against the false-positive rate (1-specificity) at different threshold settings.

$$\text{Accuracy} = \frac{(\text{True Positive} + \text{True Negative})}{(\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative})} \quad (1)$$

$$\text{Precision} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Negative})} \quad (3)$$

3.2.1. Random Tree: The node of a random tree can be used as a distributed environment. It is an ensemble of multiple decision trees. Decision tree nodes are used to build classification and regression tree methods for prediction. It uses recursive partitioning to train records into segments with similar output field values. The data is split into subgroups based on the impurity index. If there is no pruning, then the tree grows to the largest extent possible. To get the score, it combined the individual scores using majority voting for classification, and the average was taken for regression.

3.2.2. Lazy k-Nearest Neighbour (IBK)

IBK stands for lazy learning. It is used for local estimation to generate the objective function. This process of local estimation can be suitable for multiple problems. The problem condition has changed, but there is still a probability of the system working more [24, 25]. It is also suitable for quick training and validation for a long time. Large space is required for stored training data for the non-parametric model in the IBK classifier, which is good background for classification and regression [26]. The nearest neighbor can be identified easily using an object editor or by applying

cross-validation to a certain upper bound. IBK is used for different search algorithms to find the nearest neighbor. It is also used in different distance formulas such as Euclidean, Minkowski, Manhattan, and Chebyshev distances [27]. It can change the old training data into new possible data by determining the window size [28].

3.2.3 Random Committee:- The random committee is based on the group of base classifiers and their average predictions. Each one uses the same data and different random numbers of seeds. It generates the final prediction based on the average of individual base classifiers. It classified the data at the lowest error in its group of algorithms, such as random forest, etc.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this paper, the dataset is selected from the UCI machinery repository. A number of machine learning algorithms are used to develop models without feature selection. In the table 3 given below, different classifiers are compared using performance parameters such as accuracy, precision, recall, kappa statistic, mean absolute error, root mean square error, etc. Using the Weka tool, many models are applied, like Bayes (Bayes Net), functions (logistic, multilayer perceptron), lazy (IBK, Kstar), meta (random committee, classification via regression, random subspace, randomizable filtered classifier, bagging), rules (PART), and trees (random forest,j48). The 10-fold cross-validation is considered for results. If changes are made in the validation, then the result may vary. There are two ways to analyze datasets: one with replacement and another without replacement.

4.1. Dataset analysis with replacement

Table 2 shows the results of the different algorithms with replacement.

Table 2: The Data analysis with replacement

Algorithms	Accuracy	Precision	Recall	Kappa Statistics
Bayes Net	94.82	95.3	95.3	89.54
Multilayer Perceptron	95.68	100	92.2	91.36
IBK	97.41	100	95.3	94.8
Kstar	95.68	100	92.2	91.36
ClassificationViaRegression	93.1	96.7	90.6	86.16
Random committee	97.41	98.4	96.9	94.7
Randomizablefilteredclassifier	99.13	100	98.4	98.2
Randomspace	94.82	95.3	95.3	89.54
PART	94.82	98.3	92.2	89.62
tree.J48	93.96	95.2	93.8	87.82
RandomForest	97.41	98.4	96.9	94.78
RandomTree	97.41	98.4	96.9	94.78

As shown above in Table 3 and in the graph analysis of different algorithms shown in Figure 2, in terms of accuracy, the randomizable filtered classifier algorithm shows well compared to other algorithms. Accuracy (99.13%); on the other hand, if the analysis has low performance in terms of accuracy, the tree J48 and classification via regression are given the lowest values of 93.96% and 93.1%, respectively. Secondly, the Random Forest, Random Tree, Random Committee, and IBK show the same accuracy, which is 97.41%. If analysis of the precision of the multilayer perceptron, IBK, or KStar performs better than other algorithms, it provides about 100% results. The minimum performance given in terms of precision is by the tree. J48 (Precision:93.96%). In terms of recall, the best performance given by the random-filtered classifier was 98.4%. classification via regression is 86.16%.

The least result among all recall algorithms like Multilayer Perceptron and Kstar, which provide equal recall at 92.2%. It is seen that kappa statistics for best performance and least performance, respectively, in IBK are 94.8% and 4.2. Dataset analysis without replacement Now same data set is analyzed

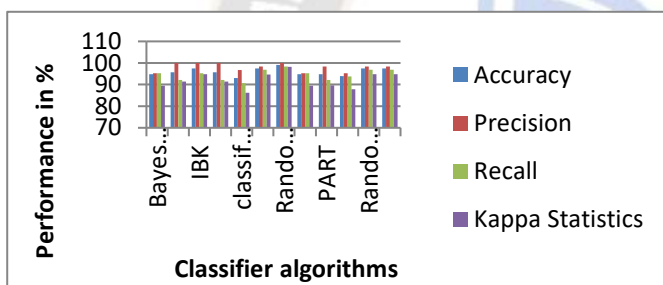


Figure 2. Graphs shows the classifiers and their performances

without replacements, a result without replacement is shown in table 3.

Table 3: Dataset analysis without replacement

Algorithms	Accuracy	Precision	Recall	Kappa Statistics
Bayes Net	64.65	69.5	64.1	29.18
Multilayer Perceptron	68.16	72.1	68.8	35.86
IBK	68.9	75	65.6	38.15
K-star	63.79	70.4	59.4	28.1
Classification Via Regression	85.34	86.2	87.5	70.32
Random committee	74.13	77.4	75	47.9
Randomizable filtered classifier	56.03	60.3	59.4	11.2
Random subspace	68.9	71.9	71.9	37.2
PART	75.86	77.3	79.7	51.03
tree.J48	73.27	72.6	82.8	45.08
Random Forest	71.55	73.1	76.6	42.18
Random Tree	72.41	74.2	76.6	44.03

Now that analysis has been done on the dataset, the analysis done without replacement is shown in table 4 and figure 3, in which the accuracy of classification via regression is best among 85.34% and the least accuracy of Bayes net is 64.65%. Different algorithms give different precision classifications, via regression is given a good 86.2%, and the minimum precision value of a randomizable filtered classifier is 60.3%. Recall and Kappa statistics show the maximum values of the tree. J48 is 82.8%, and classification via regression is 70.32%. The least value of recall and kappa statistics of algorithms Kstar and Randomizable Classifier is 59.4%, and Randomizable Filtered Classifier is 1.2.

In Figure 3, the graph shows the relationship between classifiers and their performance in percentages like Accuracy, Precision, Re-calls, Kappa Statistics. Many classifiers are used for analysis purposes, as shown in figure 3. In this graph, the algorithm without replacement is taken, and it can be seen that classification via Regression gave the better accuracy, which is more than 85.5%, and the precision of the algorithm is 86.2%. The 87.5% and 70.32% are shown for recall and Kappa statistics. Secondly, the Random committee performed better and gave the results for accuracy, precision, recall, and Kappa statistics as 74.13%, 77.4%, and 47.9%. So, without replacement, the result is not so good as compared to the result with replacement shown in figures 2 and 3.

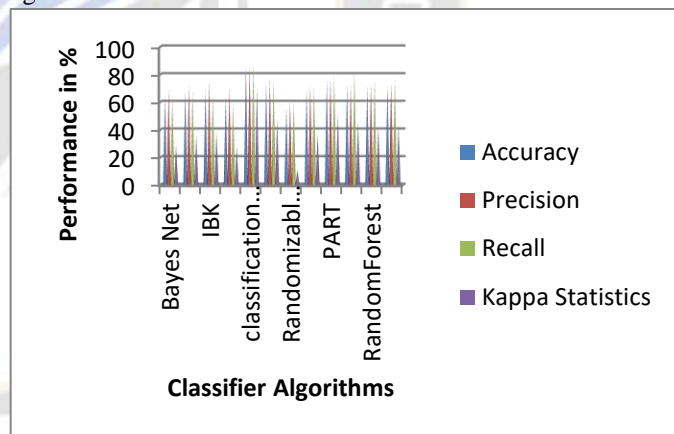


Figure 3. Graphs shows the classifier and their performances

4.3. ROC Curve with replacement with good result

The ROC curve (receiver operating characteristic) shows the performance of the classification model at different threshold levels. It plotted between the true positive rate and the false positive rate. Below are the different algorithms' ROC

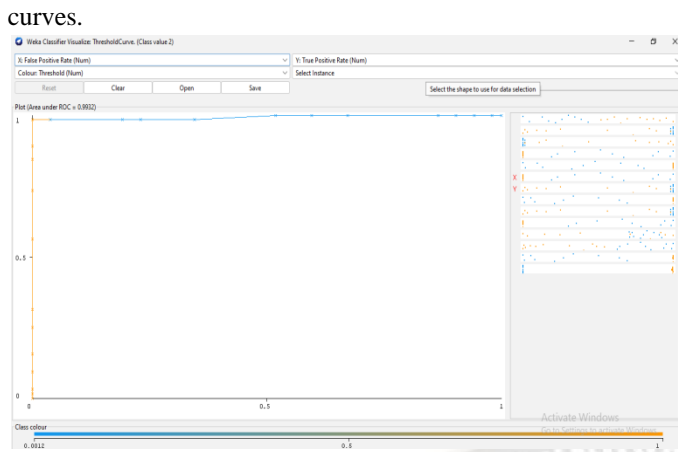


Figure 4: Shows The Area under ROC value of Randomizable filtered classifier is 0.9932

These ROC curves show the best-performing algorithms with replacement. ROC curves show the fine trade-off between clinical sensitivity and specificity for every combination of tests. It shows the usefulness of the test. The area under the ROC curve (AUROC) of a test can be used as a criterion to measure the test's discriminative ability, i.e., how good the test is in a given clinical situation.

Above figure 4, the area under the ROC curve belongs to the classifier with a replacement randomizable filtered classifier ROC value of 0.9932. This Area under ROC shows the good discriminative ability of this algorithm, which is comparatively very good among all the classifiers shown in the paper. Among them are the graphs of the true positive rate at the Y-axis and the false positive rate at the X-axis.

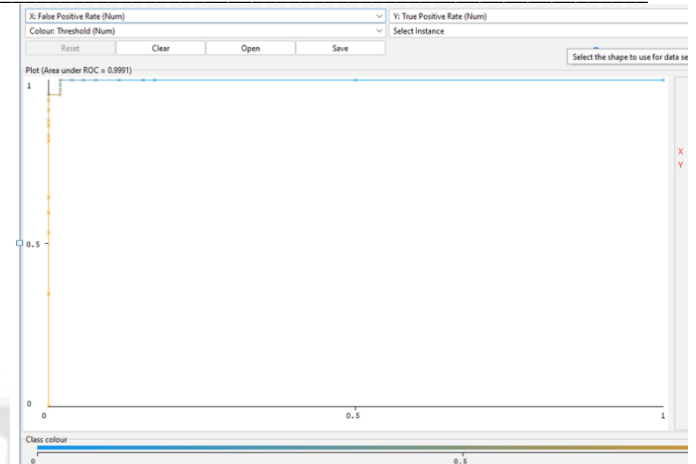


Figure 6: Shows Tree. Random Forest

As Shown above in figure 6, the Random Forest ROC curve value is 0.9991. It shows the result of the ROC at approximately 99%. This means that this classifier is good discriminative in terms of sensitivity and specificity. This classifier belongs to the Tree classifier family. The classifier's property is that it generates a number of trees, and among those, which is the best choice for the analysis purpose.

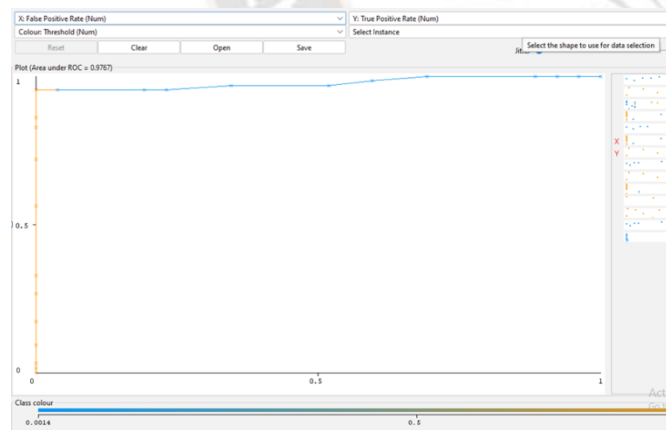


Figure 5 : Shows IBK ROC curve value is 0.9767

As Shown above in Figure 5, the IBK ROC curve value is 0.9767. It can be seen that it also gives a result of more than 97%. This classifier belongs to the Lazy classifier and shows the trade-off for clinical sensitivity and specificity for every combination of tests.

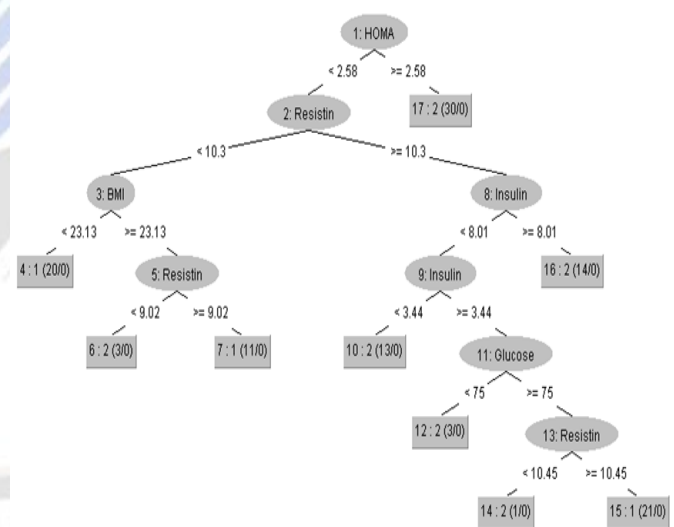


Figure 7: Tree generation by Random Tree

Figure 7 shows the tree generation by the Random tree in which, based on the entropy and information gain on the data set column, it chooses the HOMA attribute as the root node and, respectively, Resistin, BMI, etc. are chosen as the non-leaf nodes. All columns are based on information gain, whose value is high, and are chosen as nodes. The above tree generation shows the use of the tree classifier. The generation of the tree is above and uses the information gain concept.

V. CONCLUSIONS

Breast cancer is a deadly disease for women. It is a tedious job to find out about breast cancer in its early stages.

Recently, growing technology has helped them find cancer in its early stages. In this paper, the machine learning model helps find breast cancer in its early stages. The machine learning approach is applied to the dataset and based on different measures to determine the performance of the machine learning. Two approaches are applied: first, preprocessing is done, and second, analysis is done with and without replacement. As a part of the implementation, we can illustrate that the with-replacement algorithm gives better accuracy than the without-replacement algorithm. The random filtered classifier with replacement gives a better accuracy of 99.13% among all algorithms, whereas the maximum accuracy without replacement reached by the algorithm Classification Via Regression gave an accuracy of 85.34%. Finds the precision, recall, and kappa statistics to be respectively 100%, 98.4%, and 98.2. An experiment shows that with replacement, every algorithm performs better than without replacement.

REFERENCES

- [1] W. A. Khan, P.D. Arunima and Md. S. H.Onim, "A Combined PCA-MLP Network for Early Breast Cancer Detection", IEEE 978-1-7281-8402-9/21, 2021, pp. 1-6.
- [2] V. K. Singh, and H. A. Rashwan, "Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network," Expert Systems with Applications, vol. 139, p. 112855, 2020
- [3] V. Mahesh, and Mohan Kumar, "An ensemble classification based approach for breast cancer prediction", IOP Conf.: Materials Science and Engineering, pp. 1-10, 2021.
- [4] A. B.Nassif, and M. A. Talib, "Breast cancer detection using artificial intelligence techniques: A systematic literature review", Artificial Intelligence in Medicine, Elsevier, Vol.127, pp. 1-26, 2022.
- [5] Sharma D., Kumar R., and A. Jain, "A systematic review of risk factors and risk assessment models for breast cancer," Mobile Radio Communications and 5G Networks, pp. 509–519,2021.
- [6] SarthakVyas, and AbhinavChauhan, "Cancer Detection Using Machine Learning Techniques", IJRASET, Vol.10, pp. 3232-3237, 2022.
- [7] F. Bray, J. Ferlay, and I. Soerjomataram, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, Vol. 68, pp. 394–424, 2018.
- [8] S. Khan, and N. Islam, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," PRL, Vol. 124, Pp. 1–8, 2019.
- [9] Naiyer, Vaseem, Jitendra Sheetlani, and Harsh Pratap Singh. "Software Quality Prediction Using Machine Learning Application." Smart Intelligent Computing and Applications: Proceedings of the Third International Conference on Smart Computing and Informatics, Volume 2. Springer Singapore, 2020.
- [10] HibaMasood, "Breast Cancer Detection using Machine Learning Algorithm", IRJET, Vol.08, 2021.
- [11] ManavMangukiya , AnujVaghani , and Meet Savani, "Breast Cancer Detection with Machine Learning", IJTASET, Vol.10, pp.141-145, 2022.
- [12] S. Poorani, and P. Balasubramanie, "Deep Neural Network Classifier in Breast Cancer Prediction", International Journal of Engineering and Advanced Technology, Vol.9 ,pp.2106-2109,2019.
- [13] Chhaya Gupta, and Kirti Sharma, "Early Breast Cancer Detection using Various Machine Learning Techniques", IJERT, Vol. 11 Issue 06, pp.114-121, 2022.
- [14] Shatabdi Paul, and PremPrakashSolanki, "Epidemiological Study on Breast Cancer Associated Risk Factors and Screening Practices among Women in the Holy City of Varanasi, Uttar Pradesh, India", Asian Pacific Journal of Cancer Prevention, Vol. 16, pp. 8163-8171, 2015.
- [15] K.Rajendran, M. Jayabalan, and V. Sivakumar, Feasibility Study on Data Mining Techniques in Diagnosis of Breast Cancer, IJMLC, Vol. 9, No. 3, pp. 328-333, June 2019.
- [16] A.Rasool, C.Bunternghit, and L. Tiejian, "Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis", International Journal of Environment Res. Public Health, Vol.19, pp.1-19, 2022.
- [17] YixuanLi ,Zixuan Chen, "Performance Evaluation of Machine Learning Methods for Breast Cancer Prediction", Applied and Computational Mathematics, Vol. ; 7(4), pp. 212-216, 2018.
- [18] GanjarAlfian, and Muhammad Syafrudin, "Predicting Breast Cancer from Risk Factors Using SVM and Extra-Trees-Based Feature Selection Method", MDPI computers, Vol.11, pp. 1-14, 2022.
- [19] SanamAamir, and Aqsa Rahim, "Predicting Breast Cancer Leveraging Supervised Machine Learning Techniques", Computational and Mathematical Methods in Medicine, Pp.1-13, 2022.
- [20] M. Patrício, and J. Pereira, "Using Resistin, glucose, age and BMI to predict the presence of breast cancer", BMC Cancer, Vol. 18, pp.1-8, 2018.
- [21] S. Poorani, and P. Balasubramanie "Deep Neural Network Classifier in Breast Cancer Prediction", IJEAT, Vol. 9, 2019.
- [22] S. Anupam "Data Mining and Principal Component Analysis on Coimbra Breast Cancer Dataset", AIJR Proceedings, Pp.42-49, 2021.
- [23] Qian BZ, and Zhang H, "CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis", Nature. Vol.475, Pp.222–225, 2011.
- [24] K.H. Raviya and B. Gajjar, "Performance Evaluation of different data mining classification algorithm using WEKA", Indian J. Res., pp.19–21, 2013.
- [25] B. Othman, and M.F.Yau, "Comparison of Different Classification Techniques Using WEKA for breast cancer", International Conference on Biomedical Engineering, Malaysia, 2006.
- [26] N.S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression". Am. Stat. Vol.46, PP.175–185, 1992.
- [27] Nagendra Singh, "IOT enabled hybrid model with learning ability for E-health care systems", Measurement: Sensors, vol.24, pp.1-10, 2022.
- [28] M. Hossein "Machine-Learning-Based Classification Approaches toward Recognizing Slope Stability Failure", MDPI Appl. Sci., Vol.9, 2019.