ISSN No:-2456-2165

Breast Cancer Detection: A Comparative Study Using Machine Learning Models

[1]Inam, Ul Haq, [2]M. Mudasar Azeem, [3]Mubasher Hussain
[1,2,3]Faculty of Computing University of Okara, Okara 56300 Pakistan.

[4]Shazab Bashir
[4]Superior University & DPS College Model Town Lahore 54000 Pakistan

Abstract:- Breast cancer is considered one of the biggest killers in women globally. The major reason of mortality is the reason that cancer is diagnosed at later stages. Objectively, this study is conducted to compare evaluation metrics of 6 ML models such as Naïve Bayes, k-Nearest Neighborhood (K-NN's), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) and Logistic Regression (LR) on Wisconsin Breast Cancer (BC) Dataset. WEKA tool has been used to calculate the performance evaluation of these supervised ML algorithms. The literature shows that the Weka tool has been widely used in various data mining problems. The results clearly show that two models have achieved better accuracy, recall and other performance metrics in order to identify risk of breast cancer in women. These two models are K-NNs and Random Forest. In conclusion, these supervised classifiers have been trained to detect malignant and benign cells. In the future, this study may be extended for BC classification on medical images on larger dataset in order to diagnose cancer at early stages.

Keywords:- *Machine Learning Algorithms, Breast Cancer, WEKA, ML Classifiers.*

I. INTRODUCTION

There are approximately 200 types of cancers that exist in the world and the reason for most of them is not known yet. In human body, cells grow, break and regrow normally. Cancerous cells start to grow abruptly and cannot be diagnosed in the beginning due to very tiny mass. The BC in women is the most proliferating malignant cancer in China as alone in 2014 nearly 279,000 females are detected with BC, and overall, 70,000 deaths account which is approximately 13% of deaths due to breast cancer in all over the world. So, it is evident that BC is one of the critical health issues in China [1]. As cancer disease has four stages in which early diagnose can increase survival rate up to 98% while nearly 24% of total cancers in women are related to breast cancer [2], [3].

Y. Li et. al. [4] state that the BC is one of the most dangerous cancers in women that is the second main reason of death in females. So, it is obvious to attract preventive measures for research and academia. Indians have the lowest breast cancer incidents but have higher mortality rate as compared to China and USA [5].

The main and prominent kinds of cancer include carcinoma, sarcoma, lymphoma, melanoma, and leukemia. Among these, the carcinoma is the most common cancer that affects skin, lungs, breasts, pancreas, endocrine organs, and glands. Lymphoma is another type of cancer that attacks the lymphatic system or lymphocyte cells. Leukemia is a cancer of the blood cells. Sarcoma attacks bones and tissues. Melanoma or malignant melanoma is a skin cancer that attacks melanocyte cells. The other name for tumor is neoplasm. Normally tumors are diagnosed through biopsy, CT scan or MRI, Mammogram, X-Rays and others. Although cancer mortality rate is falling and survival rate is rising, yet the risk of getting cancer increases with age and higher rate is observed in developed countries. All tumors are not cancerous and nearly 90% patients die in metastasis stage. It is another fact that some benign tumors of colon and skin have tendency to be cancerous. Broadly cancer can be categorized either benign or malignant (cancerous). The unique characteristics of both benign and malignant cancers have been described in table 1.

Table 1: Benign and Malignant Cancers

Benign Cancer	Malignant Cancer
Spreads slowly	Spreads Quickly
Stays Locally (not spread) *(non-metastasize)*	Destroys Surrounding Tissues *(Invasive or metastasize)*
Small Size, Well Diagnosed	Large Size, Poorly Diagnosed
Surrounded by Fibrous Sheath	Open
Less Deaths	More Deaths
Smooth, Distinct and Regular borders	Irregular Borders
Noncancerous	Cancerous

ANN's are inspired by human intelligence of neurons to learn and respond in heuristic manners. The ANN's possesses good characteristics of parallel processing, store information on entire network, solve nonlinear complex problems, generate information from incomplete dataset, no restriction on input variables, Deep learning (DL) is comparatively new field of ANNs to achieve better classification performance. DL models uses complex hidden layer's structure that works on huge datasets to for training. The CNN is considered one of the trending DL models that are currently being used in different applications ranging from computer vision, AI, behavioral sciences and medical images [6].

Kinds of Artificial Neural Networks

- Feed-forward (FFNN)
- Recurrent (RNN)
- Modular
- Convolutional (CNN)
- De-convolutional (DCNN)

FFNN is considered one of the basic types of artificial neural networks (ANN) in which information flows from back to forward only opposite to RNN. It forwards information in one direction, normally from input to output nodes not backward as in RNN. The FFNN can have middle layers but it is not necessary. Basically, this type of neural network was designed to process large amounts of noise. FFNN is widely used in Computer Vision and image processing.

RNN forwards information from input layer to output layer. The output is now passed back to the middle nodes for weights update. Such a kind of ANN is commonly used in text-to-speech conversions.

CNN is the most common ANN that is being used these days. This model is an improved version of ANN and falls under deep learning. It works on a multiple layer perceptron through back propagation method predominantly used in image processing.

DCNN works in reverse order to the CNN network and is used to find missing features that might be less important in the beginning of CNN system's task. DCNN model is normally used in image processing.

Simple neural network contains input nodes and output nodes. No hidden or multilayers are introduced in simple neural network. In contrast to simple neural network, the MNN is multilayer neural network and it is the combination of multiple ANN that works separately from one another. The MNN is used in complex problems that need more computing resources. It is fact that many machine learning and deep learning classifiers play vital role in complex optimization problems of BC detection.

Important Kinds of Machine Learning Models [7]

- Unsupervised learning
- Semi-supervised learning
- Supervised learning
- Reinforcement learning
- Transduction
- Learning to learn
- Ensemble learning

Breast Cancer detection problems refer to challenges and issues that can arise when identifying and diagnosing breast cancer. Some of other issues in BC detection include size of dataset used, pattern recognition problems, false positive and false negative, dense breast tissue, variability in interpretation, early detection in high-risk population, need for improved imaging techniques and others. Although selecting a suitable classifier is a difficult task, yet many of the problems have been addressed through machine learning classifiers. So, a comparative analysis has been presented in the comparison results table.

II. LITERATURE REVIEW

A. ML Classifiers for BC Detection

Classification accuracy is based on assessment criteria. Machine learning (ML) is an exclusive utilization of algorithms rooted on AI techniques most specifically neural networks. The ML approach somehow gives ability to learn and improve automatically from experience without being manually programmed. Various supervised ML models have been used to detect BC tumor cells. This paper has tested six supervised ML models on BC dataset. These models include logistic regression, *K-NN's*, decision tree, random forest, *SVM* and naïve bayes.

B. K-Nearest Neighbor

The *K-NN* is regarded as non-parametric lazy supervised classifier that works based on the Euclidean distance as shown in equation (1). The value of K represents number of neighbors in the system. Let's suppose we have K=3 that means we have to take three nearest neighbor with minimum values. It provides better prediction if there are close neighbors in the feature dataset.

"It will perform classification by finding the nearest and similar data points within the corresponding dataset, and it will perform a pretrained guess depending on those classifications. Even though it seems to be very simp le to understand and develop, this technique can be used in many wide applications in various domains such as recommendation systems, semantic searching, and anomaly detection and many others. KNN can also be called as lazy learning, which means that there is no need of specific training section before classification. Instead of the repeated efforts to generalize and abstract the data classification is method makes it easier. This means that we can start classifying as soon as the data gets generalized i.e., once we've our information, there will be a square measure of some inherent issues with this sort of algorithmic program" [8].

In some cases, the K-NN provides 100% precision in BC diagnosis while MLP produces 99% precision [5]. This improved precision happens when data distribution is not well defined or complex relationships exist. In another research paper, the K-NN produces 96.22% classification accuracy in BC detection [9]. the K-NN classifier is considered excellent in pattern recognition with PCA to predict the target class with better precision and test accuracy [10]. The most widely used Euclidean Formula is given below in equation 1.

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \qquad (1)$$

Euclidean Distance

C. Decision Tree (DT)

A very popular supervised learning classifier that employs a decision tree structure of decisions similar to full binary tree that uses top-down approach. The DT uses the ID3 which calculates Entropy and Information Gain of each attribute to construct the decision tree. In general, the DT works on the process of splitting the data between true and false in general

[11]. The decision tree classifier has produced 100% precision in BC diagnosis [5]. Some cancer images are complex to diagnose as different features of interest are to be included. This can be visualized through individual decision tree. Moreover, DT handles missing values and fine tunes the algorithm.

D. Random Forests (RF)

The RF algorithm comprises of different decision trees for ensemble of categorization, regression and differentiation in training. In short RF uses decision tree structure in a randomized fashion. The RFs are basically utilized to resolve issues of overfitting in decision trees. Moreover, it is resistant to outliers and noise in data. It tackles such issues through different decision trees while the final decision is based on majority of votes as shown in Figure 1.

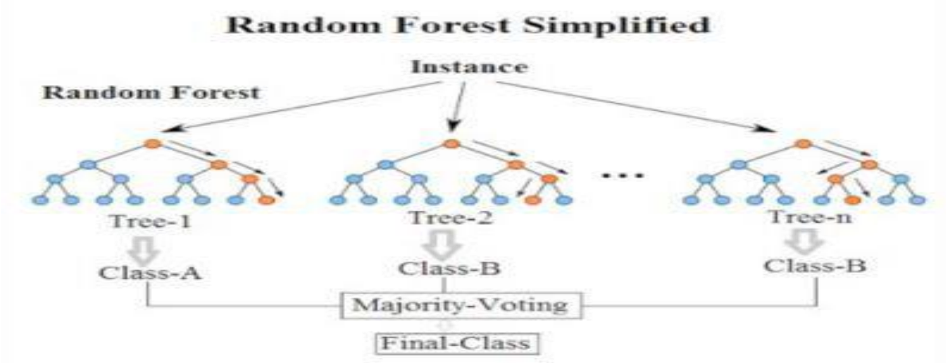


Fig. 1: Random Forest uses ID3

E. Support Vector Machine

The *SVM* is a popular and trending supervised ML algorithm to find the maximum separation distance (hyperplane) between two clusters or classes. So *SVM* algorithm is used to draw the marginal distance between two groups of objects. In fact, a hyperplane (a line) separates these objects. It works both classification and regression [11]. *SVM* works well on small and unbalanced dataset but is black-box classifier due to decision boundaries and reasoning behind predictions may be difficult to address. In another research paper, the *SVM* produces 96.42% & 96.9% classification accuracy in BC detection respectively [9], [12]. The *SVM* formula is shown in equation 2.

L(w,b,α)=12‖w‖2−∑mi=1αi[yi(w⋅x+b)−1] (2)

F. Naive Bayes

Being simple and easy to implement, the *Naive Bayes* classifier works on the basis of Bayes Theorem that works through calculating probabilities. So, it provides faster results as compared to more complex problems and robust to irrelevant features. In Bayes theorem, we determine the probability of an event A, given that B has occurred. The naïve bayes theorem formula is shown in equation 3 below [11].

$$P(A|B) = P(B|A)*P(A) / P(B) \quad (3)$$

However, the Naïve Bayes classifier may not work well on missing values in dataset.

G. Logistic Regression

This is statistical neural network that works on binary classification. It works on categorical data or binary data that is opposite to linear regression which takes continuous data values. The LR uses a logistic or sigmoid function to predict the likelihood of a binary event occurring as shown in equation 3 below.

$$S(x) = \frac{1}{1+e^{-x}} \quad (4)$$

Where ***e*** is Euler's constant. Further LR uses binary classification to classify data into two classes to take decision. Talking about other classifiers on BC Wisconsin data set, a statistical classifier *AdaBoost* achieves an accuracy of 98.77. *Random Forest* and *SVM* have achieved nearly 96.5% accuracy. However, ensemble learning of both these models enhanced accuracy by 4.3%. Another ensemble model *Bagoost* has been used to detect BCs in Indian Malva region that has achieved an accuracy of 98.21%. A comparison study of different deep learning methods is used to predict BCs and accuracy is achieved by 96.99%. The convolutional neural network achieved 97.66% accuracy [13]. In some cases, the logistic regression has shown the highest classification accuracy of 98.1% on all features of BC dataset, test accuracy ranges 95.6%-98.1%, highly correlated test accuracy 93%-95.6% [14]. Most of these ML classifiers have acquired more than 90% test accuracy [15].

In short, our comparison table have shown better accuracy and sensitivity in the Result section using WEKA tool.

III. METHOD AND PROCEDURE

This experimental study is aimed to evaluate the comparison of performance of six supervised ML algorithms for BC detection. The methodology is quantitative based on experimentation using deep learning classifiers. Weka data mining tool has been used to conduct the tests. The predictive analysis of BC dataset downloaded from Wisconsin database have been presented through steps in table 2.

Table 2: Procedure of ML Models in WEKA

Loading BC Dataset with Classes (C0, C1)
No Filters in Preprocessing
Apply Classifiers on Training Set (66% & 34%)
No Missing Values in Instances
Measure Performance Metrics
Analyze Confusion Matrix
Visualize Results

A. Algorithm

- Perform data preprocessing for missing values, noise reduction, data reduction, normalization if/when required.
- Read BC dataset.
- Divide data into features and class.
- Check features of interest for class prediction.
- Split the dataset in two training and testing set in 66:34 respectively.
- Apply ML classifiers.
- Measure accuracy and other evaluation metrics.
- Save/print performance reports of different models with class.
- Plot confusion matrices of all classifiers comparing actual with predicted class.
- Visualize Results.

B. Breast Cancer Dataset

Normally datasets provide attributes of data in different formats. In our dataset, all attributes are relevant and have nominal values of true and false, and there is no missing value in the data. Hence no preprocessing is required. Decision list contains 21 rules. These attributes contain features that may be used in various detection and prediction strategies. In this paper**,** we have used free publicly available dataset Wisconsin of UCI repository. We have compared accuracies, precision, recall, F-measure, MCC and MOC using six ML algorithms; *SVM*, *Logistic Regression*, *Random Forest*, *Naive Bayes*, *Decision Tree* and *K-NNs* and *Decision Tree* classifiers. We have shown that *SVM* and *DT* have the highest accuracy while f-test yields better results for the smaller datasets while sequential forward selection for the larger datasets. There is no duplicate feature in the data neither in values nor in the index [9]. Multiple researchers have used or referred the Wisconsin datasets [16], [12], [3], [15], [14], [9], [14], [17], [18] and many others.

C. Dataset Information

The dataset plays a crucial role in machine learning as it serves as the foundation for training, evaluating, and improving machine learning models. The quality, size, and composition of the dataset directly impact the performance and generalization ability of the resulting model. Some of the important information of dataset used is presented below.

Total Rows: 286
Total Attributes: 10

Age, Menopause, tumor-size, inv-nodes, node-caps, degree malignant, breast, breast-quad, irradiat, Class. This breast cancer dataset was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia.

Some of the important features of Wisconsin dataset have been presented in table 3. The dataset attribute “node-caps” contains 8 missing rows (3%) and only one missing row in “breast-quad”. The missing values are filtered through “*ReplaceMissingValues*” in *WEKA* in preprocessing the dataset. After preprocessing, there is not significant change in the results. Feature selection procedure is not needed and the data is already in .CSV format.

Table 3: Important Features in Dataset Contribution

Feature No.	Feature Name	Contribution to Dataset (%)
3	tumor-size=30-34	21%
1	age=55-59	33%
2	Menopause (Premeno)	45%
7	Breast (Right)	53%
10	Recurrence Events	30%

IV. EXPERIMENT AND RESULTS

In many cases, the correctly classified instances do not exceed more than 95% [19]. In predictive analysis and the reliability of the machine learning algorithms can be measured using success metrics. Positive classification means identifying correct malignant case. True Positive (TP), True Negative, False Positive (FP) and False Negative (FN) classifications are used to calculate different metrics to evaluate the results [15], [12].

- *TP = True Positive means model identifies correctly as a BC patient.*
- *TN = True Negative means model identifies correctly as an individual with no BC.*
- *FP = False Positive means model incorrectly identifies as a non-BC patient as having BC.*
- *FN = False Negative means model fails to identify a patient having BC.*

All six machine learning classifiers are tested for breast cancer detection using *WEKA* tool. The important evaluation metrics for a classifier are F-score, accuracy, precision, P4, Kappa, MCC, AUC and ROC. All these evaluation measures are calculated and evaluated. To handle the variables, the six classifiers applied in this dataset all use 66% of the data for training and 34% of the data for testing. *WEKA* data mining tool is famous handy tool so the intelligent techniques are applied [18]. We are more concerned with two performance measures i.e. accuracy and recall (true positive or sensitivity). The results of all classifiers are presented in Figures 2-8 at the end of this paper.

As it is clear in figure 8 that incidence of BC increases as age increases. The maximum chance of getting BC is in age group 30 – 70. The sensitivity and accuracy of another [1] paper is 97.14 and 97.65 respectively while our proposed method produced better results as shown in the Table 4.

Breast Cancer Prediction (99% ROC ML)

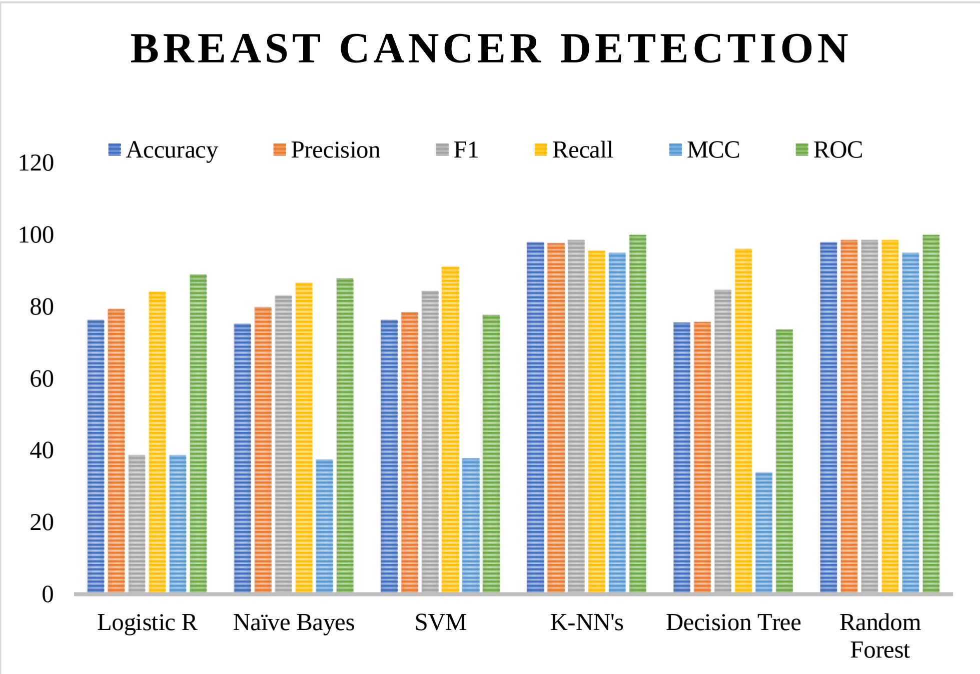
Table 4: Results Achieved BY Classifiers

Classifier Type	Performance		Existing Results	
	Recall	Accuracy	Recall	Accuracy
Logistic Regression	89.6%	76.0%	66.10	57.80 [3]
Naïve Bayes	86.0%	75.0%	N/A	74.00 [2]
K-NNs	99.5%	97.9%	97.65	97.14 [1]
SVM	91.0%	75.0%	83.33	83.33 [13]
Random Forest	98.5%	97.9%	95.65	95.71 [1]
Decision Tree	96.0%	75.8%	N/A	83.00 [2]

A. Principal Findings

In this experiment, we have applied six ML classifiers to develop BC detection model to identify the disease. Their performance is calculated and compared with each other, and the results are indicated in table 4 above. It is clear that Random Forest (*RT*) and *K-NN's* produces good results as both produce 97.9% accuracy metric. While *KNN's* outperforms RT and produces 99.5% recall as compared to 98.5% by RT. Moreover, the ROC curve of decision tree and random forest are 99.9%. Moreover, *KNN's* is slightly better in terms of specificity (TN). Talking about confusion matrix of BC dataset, the *KNN's* misses two cancer patients in classification process. While random forest misses two non-patients which is tolerable. All six machine learning algorithms are presented in Table 5.

Table 5: Six Supervised Learning Algorithms for Breast Cancer Detection



Currently, we have compared six supervised learning models for BC detection as shown in the table 5 above. The performance of *KNN's* and *RF* is optimal in BC detection. Recall is the number of positive results divided by total positives. So, recall is another good performance metric and *SVM* produces 96%. After external validation of results from the academia and industry, we will scale this experiment on different DL algorithms using ensemble learning to draw more accurate and valid results.

Limitation of this study is that if the features have different scales, then *K-NN* may not produce good results. Secondly Random Forest needs tuning of hyperparameters due to number and depth of trees.

V. CONCLUSION

The classification validation hugely depends on assessment pattern. This experimental paper has presented six supervised ML classifiers such as *k-Nearest Neighborhood*, *Logistic Regression*, *Decision Tree*, *Random Forest*, *Support Vector Machine* and *Naïve Bayes* on Wisconsin Breast Cancer Dataset. Our study has shown the performance evaluation of these supervised ML classifiers and it is clear that two models have achieved better accuracy, recall and other performance metrics on identifying women at high risk of breast cancer as risk factors. These two models are K-NNs and Random Forest.

ACKNOWLEDGMENT

The authors acknowledge the IT services, cloud servers, Internet facilities etc. provided by the University of Okara.

```
Correctly Classified Instances      218         76.2238 %
Incorrectly Classified Instances     68         23.7762 %
=== Detailed Accuracy By Class ===
           TP Rate FP Rate Precision Recall  F-Measure MCC     ROC Area PRC Area Class
           0.896   0.553   0.793     0.896   0.841     0.387   0.784    0.890    no-recurrence-events
           0.447   0.104   0.644     0.447   0.528     0.387   0.784    0.645    recurrence-events
Weighted Avg.  0.762   0.420   0.749     0.762   0.748     0.387   0.784    0.817
=== Confusion Matrix ===
  a   b   <-- classified as
 180 21 |   a = no-recurrence-events
  47 38 |   b = recurrence-events
```

Fig. 2: Logistic Regression Results

```
Correctly Classified Instances      215         75.1748 %
Incorrectly Classified Instances     71         24.8252 %
=== Detailed Accuracy By Class ===
           TP Rate FP Rate Precision Recall  F-Measure MCC     ROC Area PRC Area Class
           0.866   0.518   0.798     0.866   0.831     0.374   0.760    0.879    no-recurrence-events
           0.482   0.134   0.603     0.482   0.536     0.374   0.760    0.610    recurrence-events
Weighted Avg.  0.752   0.404   0.740     0.752   0.743     0.374   0.760    0.799
=== Confusion Matrix ===
  a   b   <-- classified as
 174 27 |   a = no-recurrence-events
  44 41 |   b = recurrence-events
```

Fig. 3: NaïveBayes Results

```
Correctly Classified Instances      218         76.2238 %
Incorrectly Classified Instances     68         23.7762 %
=== Detailed Accuracy by Class ===
           TP Rate FP Rate Precision Recall  F-Measure MCC     ROC Area PRC Area Class
           0.910   0.588   0.785     0.910   0.843     0.379   0.661    0.778    no-recurrence-events
           0.412   0.090   0.660     0.412   0.507     0.379   0.661    0.447    recurrence-events
Weighted Avg.  0.762   0.440   0.748     0.762   0.743     0.379   0.661    0.680
=== Confusion Matrix ===
  a   b   <-- classified as
 183  18 |   a = no-recurrence-events
  50 35 |   b = recurrence-events
```

Fig. 4: SVM Results

```
Correctly Classified Instances       280          97.9021 %
Incorrectly Classified Instances       6           2.0979 %
=== Detailed Accuracy By Class ===
            TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
            0.995    0.059    0.976      0.995   0.985      0.950  0.999     0.999     no-recurrence-events
            0.941    0.005    0.988      0.941   0.964      0.950  0.999     0.996     recurrence-events
Weighted Avg.  0.979  0.043   0.979      0.979   0.979      0.950  0.999     0.998
=== Confusion Matrix ===
  a   b   <-- classified as
 200  1 |  a = no-recurrence-events
  5  80 |  b = recurrence-events
```

Fig. 5. K-NN's Results

```
Correctly Classified Instances       216          75.5245 %
Incorrectly Classified Instances      70          24.4755 %
=== Detailed Accuracy By Class ===
            TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
            0.960    0.729    0.757      0.960   0.846      0.339  0.584     0.736     no-recurrence-events
            0.271    0.040    0.742      0.271   0.397      0.339  0.584     0.436     recurrence-events
Weighted Avg.  0.755  0.524   0.752      0.755   0.713      0.339  0.584     0.647
=== Confusion Matrix ===
  a   b   <-- classified as
 193  8 |  a = no-recurrence-events
  62 23 |  b = recurrence-events
```

Fig. 6: Decision Tree Results

```
Correctly Classified Instances       280          97.9021 %
Incorrectly Classified Instances       6           2.0979 %
=== Detailed Accuracy By Class ===
            TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
            0.985    0.035    0.985      0.985   0.985      0.950  0.998     0.999     no-recurrence-events
            0.965    0.015    0.965      0.965   0.965      0.950  0.998     0.995     recurrence-events
Weighted Avg.  0.979  0.029   0.979      0.979   0.979      0.950  0.998     0.998
=== Confusion Matrix ===
  a   b   <-- classified as
 198  3 |  a = no-recurrence-events
  3  82 |  b = recurrence-events
```

Fig. 7: Random Forest Results

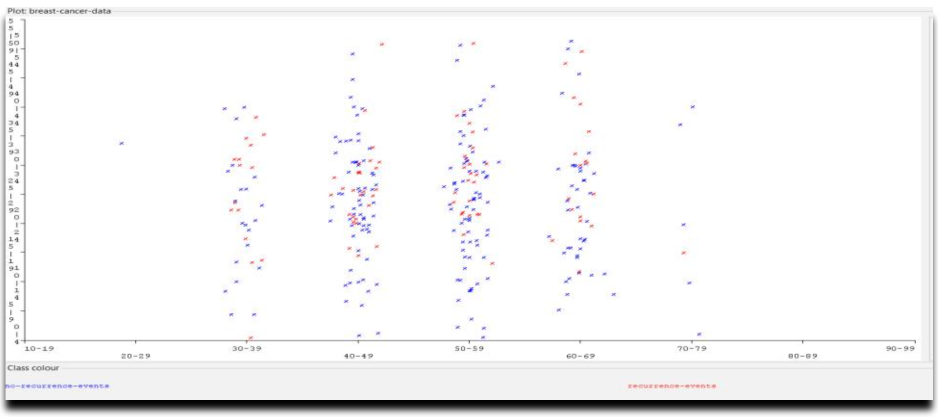


Fig. 8: BC Prediction Data Age Wise

REFERENCES

[1.] C. Hou et al., "Predicting Breast Cancer in Chinese Women Using Machine Learning Techniques: Algorithm Development," JMIR Med. Inform., vol. 8, no. 6, p. e17364, Jun. 2020, doi: 10.2196/17364.

[2.] M. Tahmooresi, A. Afshar, B. B. Rad, K. B. Nowshath, and M. A. Bamiah, "Early Detection of Breast Cancer Using Machine Learning Techniques," vol. 10, no. 3, p. 7.

[3.] S. A. Mohammed, S. Darrab, S. A. Noaman, and G. Saake, "Analysis of Breast Cancer Detection Using Different Machine Learning Techniques," in Data Mining and Big Data, Singapore, 2020, pp. 108–117. doi: 10.1007/978-981-15-7205-0_10.

[4.] Y. Li, "Performance Evaluation of Machine Learning Methods for Breast Cancer Prediction," Appl. Comput. Math., vol. 7, no. 4, p. 212, 2018, doi: 10.11648/j.acm.20180704.15.

[5.] M. Gupta and B. Gupta, "A Comparative Study of Breast Cancer Diagnosis Using Supervised Machine Learning Techniques," in 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), Feb. 2018, pp. 997–1002. doi: 10.1109/ICCMC.2018.8487537.

[6.] Mohammed Abdul Hay Abu Bakr, Haitham Maher Al-Attar, Nader Kamal Mahra, Samy S. Abu-Naser, "Breast Cancer Prediction using JNN," Int. J. Acad. Inf. Syst. Res. IJAISR, vol. 4, no. 10, p. 8, Oct. 2020.

[7.] P. Gupta and S. Garg, "Breast Cancer Prediction using varying Parameters of Machine Learning Models," Procedia Comput. Sci., vol. 171, pp. 593–601, Jan. 2020, doi: 10.1016/j.procs.2020.04.064.

[8.] M. S. Yarabarla, L. K. Ravi, and A. Sivasangari, "Breast Cancer Prediction via Machine Learning," in 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Apr. 2019, pp. 121–124. doi: 10.1109/ICOEI.2019.8862533.

[9.] R. Dhanya, I. R. Paul, S. Sindhu Akula, M. Sivakumar, and J. J. Nair, "A Comparative Study for Breast Cancer Prediction using Machine Learning and Feature Selection," in 2019 International Conference on Intelligent Computing and Control Systems (ICCS), May 2019, pp. 1049–1055. doi: 10.1109/ICCS45141.2019.9065563.

[10.] H. Rajaguru and S. C. S R, "Analysis of Decision Tree and K-Nearest Neighbor Algorithm in the Classification of Breast Cancer," Asian Pac. J. Cancer Prev. APJCP, vol. 20, no. 12, pp. 3777–3781, 2019, doi: 10.31557/APJCP.2019.20.12.3777.

[11.] Jean Sunny, Nikita Rane, Rucha Kanade, Sulochana Devi, and xavier institute of engineering, "Breast Cancer Classification and Prediction using Machine Learning," Int. J. Eng. Res., vol. V9, no. 02, p. IJERTV9IS020280, Mar. 2020, doi: 10.17577/IJERTV9IS020280.

[12.] E. A. Bayrak, P. Kırcı, and T. Ensari, "Comparison of Machine Learning Methods for Breast Cancer Diagnosis," in 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Apr. 2019, pp. 1–3. doi: 10.1109/EBBT.2019.8741990.

[13.] S. Sharma, A. Aggarwal, and T. Choudhury, "Breast Cancer Detection Using Machine Learning Algorithms," in 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Dec. 2018, pp. 114–118. doi: 10.1109/CTEMS.2018.8769187.

[14.] A. Al Bataineh, "A Comparative Analysis of Nonlinear Machine Learning Algorithms for Breast Cancer Detection," Int. J. Mach. Learn. Comput., vol. 9, pp. 248–254, Jun. 2019, doi: 10.18178/ijmlc.2019.9.3.794.

[15.] P. Ghosh, S. Azam, K. Md. Hasib, A. Karim, M. Jonkman, and A. Anwar, "A Performance Based Study on Deep Learning Algorithms in the Effective Prediction of Breast Cancer," in 2021 International Joint Conference on Neural Networks (IJCNN), Jul. 2021, pp. 1–8. doi: 10.1109/IJCNN52387.2021.9534293.

[16.] B. Sahu, S. Mohanty, and S. Rout, "A Hybrid Approach for Breast Cancer Classification and Diagnosis," EAI Endorsed Trans. Scalable Inf. Syst., vol. 6, no. 20, Jan. 2019, Accessed: Nov. 11, 2022. [Online]. Available: https://eudl.eu/doi/10.4108/eai.19-12-2018.156086

[17.] Md. M. Islam, Md. R. Haque, H. Iqbal, Md. M. Hasan, M. Hasan, and M. N. Kabir, "Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques," SN Comput. Sci., vol. 1, no. 5, p. 290, Sep. 2020, doi: 10.1007/s42979-020-00305-w.

[18.] M. K. Keleş, "Breast Cancer Prediction and Detection Using Data Mining Classification Algorithms: A Comparative Study," Teh. Vjesn., vol. 26, no. 1, pp. 149–155, Feb. 2019, doi: 10.17559/TV-20180417102943.

[19.] S. Bharati, M. A. Rahman, and P. Podder, "Breast Cancer Prediction Applying Different Classification Algorithm with Comparative Analysis using WEKA," in 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT), Sep. 2018, pp. 581–584. doi: 10.1109/CEEICT.2018.8628084.