Performance Evaluation of BCW Diagnostic Data Setusing Machine Learning Algorithms

D.Priyanka [1] , V. Yamuna [2], P.AnnanNaidu[3], G.Stalin Babu[4]

Asst.Professor,[1] Dept. of CSE, Aditya Institute of Technology and Management, Tekkali, Srikakulam, IndiaAsst.Professor[2], Dept. of CSE, Aditya Institute of Technology and Management, Tekkali, Srikakulam, IndiaAsst.Professor[3], Dept. of CSE, Aditya Institute of Technology and Management, Tekkali, Srikakulam, IndiaAsst.Professor[4], Dept. of CSE, Aditya Institute of Technology and Management, Tekkali, Srikakulam, India

dabbeerupriyanka90@gmail.com,vyamuna372@gmail.com,annanpaidi@gmail.com

Abstract: A typical illness which is influencing a huge number of ladies in today's world is breast cancer (BC). It is primarily answerable for big number of women's demises in each and every year (annually). Presently a-days, disease is treatable and rectifiable in crude stages. During their life, breast cancer is decisive among ten percentages of ladies. Breast malignant growth is the second famous reason for passing after cellular breakdown in the lungs on the planet. Constant pain, changes in the size of nipple or breast, transformation of qualities, shading (redness), variations in skin surface of breasts are the fundamental indications of BC. Most incessant characterization which prompts breast cancer is twofold (harmless malignant growth/malign cancer). Now-a–days, Machine Learning(ML) techniques are widely utilized in solving various problems of breast cancer and provide maximum rate of classification accuracy and also effective diagnostic capabilities. In the beyond five years, 88 lakh of ladies alive were determined to have this disease. It is the world's most universal malignant growth disease. In this paper, we affirm various ML algorithms(Support Vector Machine, Naive_Bayes, (RF)Random Forest, simple CART, Artificial neural network) to evaluate the performance of BCW (Diagnostic) data set in terms of various statistical measures such as AUC, CA(Accuracy), F1, precision and recall and find the best algorithm among all others which yields the high accuracy rate for prognosis of breast cancer.

Keywords: SVM, Naive_Bayes(NB), Random forest(RF), simple CART(SC), Artificial neural network(ANN).

DOI Number: 10.48047/nq.2022.20.19.NQ99088 **NeuroQuantology 2022; 20(19): 964-979**

1. Introduction

From last recent years, (BC)Breast cancer disease[1] is the second prominent cancer in case of death rates. Lung cancer, breast cancer, bone cancer, blood cancer etc are of different types of cancers, but among them breast cancer is the most dangerous one because of highest death rates. In today's era, most of the women affected with Breast Cancer due to the abnormal development of unwanted cells in breast. BC's are of two different types; those are Ductal and lobular cancer. Ductal cancer begins from the tubes also called as ducts and Lobular cancer begins in glands. Glands produce milk in breast and tubes help to transfer that milk from breast to nipple. Now we will discuss how to identify the breast cancer. Major symptoms such as thick area near breast and armpits, changes occur in size and shape of breast, changes in colour of skin on nipple or breast, severe pain occur in breast area, shading or redness occur at the tip of the nipple etc are used for identification of breast cancer. Age, irregular diets, health conditions and some of the genetic factors causes breast cancer. Diagnosis of breast cancer is the next step after identification of any of those symptoms. In general, when we went to specialist for therapy, first he will get some information about our family ancestry that anyone previously impacted with this disease or not, after that he will do Magnetic resonance imaging (MRI), Mammogram, Ultrasound etc[2] tests in regards to this disease. Three different types of Breast Cancer treatments are there. They are primary treatment, adjuvant treatment and palliative treatment. In primary treatment, either removed or killed the cancer cells from the breast. Surgery is one of the best primary treatment for removing or killing of those cancer cells. If we ignore the symptoms of breast cancer earlier then it leads to sensitive radiation or chemotherapy. Even after the successfully completion of primary treatment some of the cells are left or remained then in that case the primary treatment is not sufficient and moved to the next higher level of treatment called Adjuvant treatment**.** In adjuvant treatment, any of the type of therapies such as chemotherapy, radiation therapy and hormone therapy are used to kill the remaining alive cancer cells after the successful completion of primary treatment. To reduce the side effects of above treatment or symptoms which cause the cancer, the third variety of treatment called palliative treatment is used. That palliative treatment reduces the pain in breast and also avoids the frequent occurrence of breathing problems [3].

As part of the study of this paper, we diagnosis the symptoms of breast cancer by utilizing different machine learning algorithms like SVM, random forest, naives_bayes, simple CART, Artificial neural network. In this proposed system we find accuracy of each algorithm and use the algorithm which gives the best Accuracy. To find the accuracy, we have collected the dataset from UCI Repository and implemented on those algorithms. After finding the breast cancer we will give

eISSN 1303-5150

treatment based on their condition.

This paper includes the following objectives

- ✓ Study of related existing work.
- ✓ Describes various machine learning models.
- ✓ Defining Methodology – Data Collection, Preprocessing, test data, train data.
- ✓ Applied Machine learning Models on collected data set.
- ✓ Model Performance analysis

The level of this paper is to depict the difficulties of existing work and to feature the new assessment direction. The remaining part of this paper is sorted into various sections which mostly focus only on predicting the symptoms of breast cancer. The next section, Section 2 outlines basic knowledge on various related papers of breast cancer. The particulars of the datasets description, proposed model, pre-processing and implementation of various algorithms are included in next section. Preliminary setup is given in Section 4 followed by the result analysis which shows in Section 5 and summarization of the findings of research, and conclusion in the last section.

2. RELATED WORK

In today's world, Breast Cancer is the second most dangerous cancer disease among all the other cancer diseases in view of deaths. Most of the researchers worked on breast cancer dataset by applying various classification methods including Neural_Network, Decision tree, Naive Bayes, Support Vector Machine, Logistic Regression and so on. Some of the related works of researchers are explained as follows.

Mousannif et al compared 4 different learning algorithms such as SVM, Naive-Bayes, K- NearestNeighbor and C4.5 to determine the efficiency and also the accuracy in breast cancer detection and prediction on wiscon breast cancer (original)data set. They concluded that among those algorithms, the supervised algorithm svm gives highest accuracy of 91.73% to predict cancer cells in breast cancer. The data mining tool, WEKA is used to implement those algorithms [4].

Kabila et al , two algorithms: XGBoost and random forest are utilized to find breast cancer Risk prediction in terms of accuracy. Here, with 74.73% of maximum accuracy achieved by random forest algorithm for the breast cancer dataset with 275 numbers of instances and 12 numbers of features [5].

Padhi et al used two different varieties of supervised algorithms for detection of breast cancer. Farthest first and k-means algorithms are used to find accuracy in healthy and sick patients. Finally concluded with highest accuracy algorithm is k-means in healthy and farthest first in sick patients and also used two different classifiers such as J48 and Naive bayes algorithm to find accuracy. For

implementation purpose data set was taken from UCI ML repository and all those algorithms are implemented in WEKA tool [6].

Laghmati et al used 4 different classification or regression algorithms such as Binary SVM, Ada- Boost, K-NearestNeighbor and decision tree to implement whether the patient effected by malignant tumour or benign tumour. To implement those algorithms, Wiscon breast cancer dataset is used. Finally, in terms of specificity, k-NN model concluded wih highest accuracy rate of 99.12% and binary SVM gives 98.86% of classification accuracy [7].

Khuriwal et al used ensemble supervised machine learning algorithms to diagnosis the symptoms of the breast cancer disease. The input datasets called Wisconsin diagnosis (WD) breast cancer data was collected from UCI repository. At first, logistic regression algorithm was implemented on given input and then applied neural network algorithm for each individual which is used for calculating the performance measures such as precision and recall etc [8].

The below table Table 1 depicts the similar identifications of distinct papers on breast cancer data set using classification algorithms. In Table 1, first column shows the names of authors. Algorithms are depicted in second column. Next column gives report on results. Fourth column includes data set and year of publication and last column represents the reference number

Table 1: Similar identifications of distinct papers on breast cancer data set usingclassification algorithms

Authors	Algorithms Used	Results	Data Set and Year of Publication	Reference
Vikas Chourasia et al	Logistic Regression (LR), Decision Tree (DT), Support-vector clustering (SVC), K-Nearest Neighbors (KNN), Random Forest (RF) and Naïve Bays (NB)	to find out the accuracy obtained by votin g classifier (Meta level)	Wisconsin Diagnostic Breast Cancer (WDBC), 2020	[9]

Subrata kumar Mandal et al	Naive Bayes(NB), Decision Tree, Logistic Regression (LR) ,	Logistic Regression (LR) has highes taccuracy rate with Minimum tim ecomplexity.	Breast Cancer, 2017	[10]
Meriem Amrani et al	Naive Bayes (NB),K-nearest Neighbor (KNN)	when compared with Naive Bayes, K-	Breast Cancer data set, 2018	[11]

		nearest neighbour ha smaximum accuracy		
B.M Gayathri et al	Naive Bayes Classifier, Relevance Vector Machine	More than 95% of accuracy	Wisconsin original, 2016	[12]
H Wang et al	Support Vector Machine	reduces the variance by 97.89% an dincreases accuracy by 33.34%	Wisconsin Breast Cancer, Wisconsin Diagnostic Breast Cancer, and the U.S. National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program breast cancer datasets	[13]
Hiba Asri et al	Support Vector Machine (SVM), Decision Tree (C4.5), Naive Bayes (NB) and k Nearest Neighbors (k-NN)	SVM gives the maximum accuracy rate (97.13%) with lowest error-rate	Wisconsin Breast Cancer (original), 2016	[14]
Madhu Kumar e t al	LR, SVM, KNN	KNN gives maximum accuracy	WBCD, 2018	[15]

3. Methodology

3.1 Data-Set Description

In this paper, the dataset called BreastCancerWisconsin (Diagnostic) dataset is collected as input from UCI repository [16]. The dataset depict attributes of the cell nuclei present in the outline of diagram. Total numbers of instances in Breast Cancer Wisconsin (Diagnostic) Data Set are 569 and having 32 numbers of attributes. It is clear that to diagnosis the symptoms of breast cancer disease, mean of texture, radius and perimeter are termed as influential variables. Smoothness, fractal dimension and compactness are the other types of moderate influence variables. Those influential and moderate influential variables are utilized to refine

the classification accuracy of the model. We implement five different types of machine learning algorithms on above input dataset to diagnosis the symptoms of breast cancer disease. We find the accuracy of each algorithm and finally concluded with the best algorithm which produces output as maximum accuracy rate for classification of breast cancer data. The below table, Table 2 portrays the depiction of different genuine esteemed features of Breast Cancer Wisconsin (Diagnostic) Data Set.

Table 2: Description of various real valued features of Breast Cancer Wisconsin(Diagnostic) Data Set.

Sl. No	**Real valued features**	**Description**
1	Texture	SD of greyscale
2	Radius	Arithmetic mean of distances between centre and points identified on surface
3	Area	Cell's area
	Perimeter	Cell's circumference
5	Smoothness	limited change in terms of length of the cell's radius
6	Compactness	(Perimeter * perimeter) / cell's area - 1
7	Concavity	acuteness of contour's concave-portions
8	concave points	number of contour's concave portions
9	Symmetry	Symmetry
10	Fractal dimension	(approximate value of coastline) - one

3.2 PROPOSED Model

In proposed work, different supervised classification algorithms such as SVM, Artificial neural network (ANN), Naïve Bayes, Random Forest (RF) and Simple Cart (SC) are used to predict the occurrence of breast cancer. The proposed architecture consists of four different phases such as data collection (input), data preprocessing, implementing various classifiers such as SVM, Artificial neural network (ANN), Naïve Bayes, Random Forest (RF) and Simple Cart (SC), evaluation of performance and diagnosis. Figure 1 exhibits the steady flow of the proposed architecture and also explains about the complete perspective of the system.

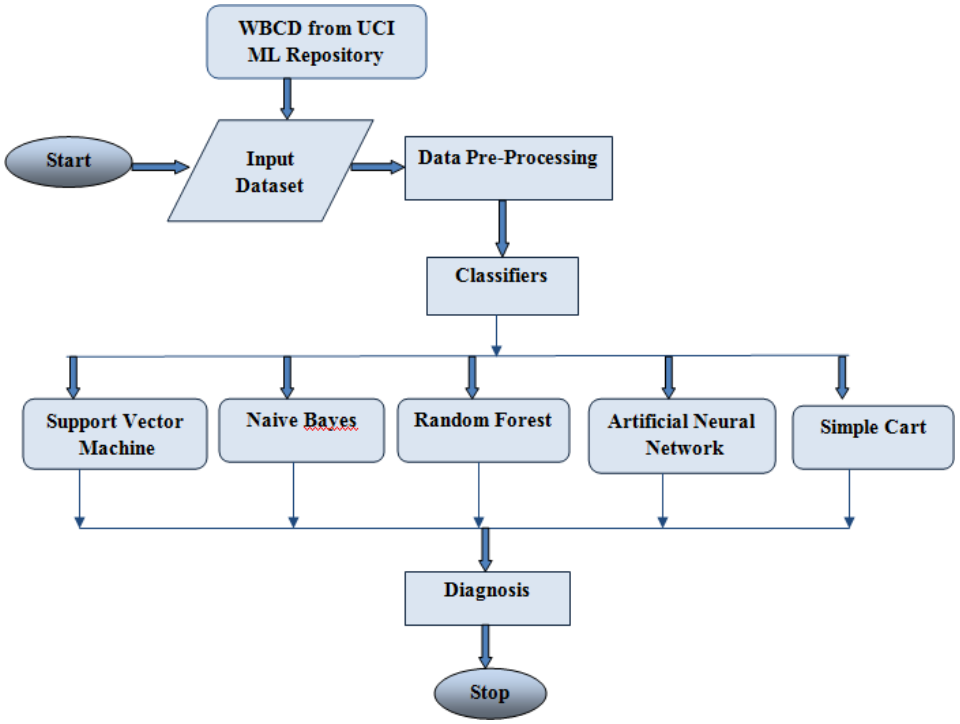

Figure 1: Proposed Architecture for the system

3.3 Pre-Processing Phase

Now a day, the input data is collected from different miscellaneous sources. Because of large size of data, databases are endangered to missing data, uncertainty or noisy data. To remove the uncertainty or filling the missing values one of the significant phase called pre-processing is used. In this paper, we apply various pre-processing techniques on breast cancer dataset to remove unnecessary data and fill the missing values. After pre- processing, various classification algorithms are applied on the data set for finding the best classification algorithm which gives the highest accuracy.

3.4 Algorithms Used

Classification is one of the mining process used to categorize the given input data into various classes called labels. It is mainly utilized for predicting the modelling problems. Classifiers can be applied on both structured type of data and unstructured data.

As part of this paper, we have compared five different types of supervised algorithms such as support vector machine, artificial neural network, Naive-Bayes, Decision Tree (simple CART), and Random forest. We have to classify and analyze dataset of breast cancer. We implemented those algorithms on breast cancer dataset by using Scikit-Learn software which is an open-source software developed in Python environment and finding out the best algorithm which gives most accurate result for diagnosing breast cancer.

3.4.1 (SVM)Support Vector Machine

One of the supervised classification algorithms such as SVM is utilized for managing regression together with classificaion methods [17]. This learning algorithm gives the better accuracy rate with minimum size of train_data. The main idea of Support_Vector_Machine is to disparate the data points by observing the separable planes in dimensional space [18].

eISSN 1303-5150

For classifying the data points, two different types of SVM's are used. One of them is Linear SVM which is used for linear classification purpose and the other type is non-linear Support Vector Machine.

3.4.2 Naive-Bayes

NB is one among the supervised learning in machine learning algorithms. This algorithm is mainly utilized for text classification algorithm by combining Bayes theorem with the assumption of "naive". For predicting the fast machine learning model, the high dimensional data (trained) is utilized for implementing Naive Bayes algorithm [19]. The below equation is used in Bayes theorm.

$$P(A/N) = \frac{P(N/A)*P(A)}{P(N)} \quad \text{(i)}$$

Where $P(A/N) = P(N1/A) * P(N2/A) * \dots\dots\dots * P(Nn/A) * P(A)$ (ii)

P(A/N) : posterior probability

P(N) : prior probability

predictorP(N/A) : likelihood

P(A) :class prior probability

3.4.3 RANDOM FOREST (RF)

To reduce the drawbacks of decision tree and maximize the precision value, one of the supervised machine learning or ensemble learning algorithm called Random Forest is used. In ensemble methods, forest represents the combination of two or more decision trees [20]. This algorithm is trained through the aggregation of bagging and bootstrapping. Bagging is an ensemble process used for improving the accuracy of various supervised learning methods.

3.4.4 ANN

Artificial neural network(ANN) is a supervised classification algorithm, which is snippet from biotic science. This algorithm mainly used in parallel processing. Learning and recall are the two processes which are involved in artificial neural networks. Process of habituate the connection weights presented at input buffer. Process which gives response to stimuli is called learning. The process which takes input from learning and giving response in artificial neural networks called recall method. Security in airways control, maintain data validation, auditing risk etc are the major applications of artificial neural networks [22].

3.4.5 SIMPLE_CART

One of the important supervised learning algorithms is CART. It expanded as Classification and regression tree. The major concepts of cart algorithm used are methodology and prediction. Hence it acts as a predictive model [23]. The model which explains how we predict the outcome variables based on their input values is called predictive model. The important algorithm which is used in simple cart algorithm is decision tree. Simple cart uses fork and split operations. Fork is mainly used to join variables and split is used to divide or split the variables.

4. Preliminary Setup

The results of those algorithms have been furnished by using one of the open source software called Scikit_learn which is refined in Python environment for libraries of machine learning on a computer system with subsequent configuration: Pentium® Dual-Core processing CPU E5300 model, @2.60GHz 2.59GHz, 2GB Random Access Memory and Microsoft Windows 8.1 Pro N OS. In this work, we affirm for productive experimentations and poor biasness on analysis of result, here 10-fold expedient has been used to isolates the data-set into 10 different folds among them 9 folds are implemented for train_data and 1 fold is used as testing data. For considering results of this cross validation, Scikit-learn (open-source software) refined in Python is used which is available freely in the internet. It is used to visualize the data at front end by data analysis and interdependent data visualization.

5 Result Analysis

This portion of the paper explains about the obtained results of the implementation and alsotheir effective analysis with respect to the algorithms used.

5.1 Measurement of CA

Confusion Matrix which is obtained for 2-class classifiers is utilized for mesuring the classification_accuracy(CA) of the proposed model. The following equation represents the formula for accuracy of the system.

$$Accuracy(A) = \frac{TPos+TNeg}{TPos+TNeg+FPos+FNeg} * 100 \quad \text{..... (iii)}$$

Where TPos(TP): gives actual values of having the symptoms of cancer TNeg(TN): gives actual values of not having the symptoms of breast cancer

FPos(FP): gives predicted values of having symptoms of breast cancer (but they don't have cancer)

FNeg(FN): gives predicted values of not having symptoms of breast cancer (but they have)

5.2 Evaluation of Performance

To dissect the order execution of various classification models, we have utilized five unique models of ML. The models were concentrated on to perceive how precisely they had the option to recognize harmless growths and dangerous cancers (carcinogenic); threateningbeing the positive class and harmless being the negative.

In addition, we have utilized different execution measure boundaries like Specificity/TN Rate, Recall/TP Rate, FN Rate, FP Rate, Precision and F1-Score, involving the Confusion matrix to give more assessments. The accompanying fragment discusses about all the outcomes acquired.

5.2.1 Performance Measure Parameters

Recall/True Positive Rate – Recall demonstrates what extent of the all out Malignant examples were accurately anticipated as Malignant.

$$Recall = TPos/(TPos + FNeg) \quad \text{.....(iv)}$$

Specificity/True Negative Rate – Specificity demonstrates what extent of the all out Benign examples were accurately anticipated as Benign.

$$Specificity = TNeg/(TNeg + FPos) \quad \text{.....(v)}$$

False Negative Rate – It demonstrates what extent of the absolute Malignant examples were mistakenly named Benign.

$$FNRate = FNeg/(TPos + FNeg) \quad \text{.....(vi)}$$

False Positive Rate – It demonstrates what extent of the absolute Benign samples were inaccurately named as Malignant.

$$FPRate = FPos/(TNeg + FPos) \quad \text{.....(vii)}$$

Precision – It shows what extent of the all predictions (all examples there were anticipated as Malignant) were really positive.

$$Prec = TPos/(Tpos + FPos) \quad \text{.....(viii)}$$

F1-Measure – The F1-Measure addresses an outline of the overall performance ofthe model on the test dataset.

$$F1 = (2 * Prec * Recall)/(Prec + Recall) \quad \text{(ix)}$$

5.3 Analysis of algorithms used

Table 3 portrays about the relative analysis of different classification algorithms such as Support vector machine(SVM), Artificial neural network(ANN), Naïve_Bayes, Random Forest and Simple Cart by considering various statistical parameters such as accuracy rate and recall value etc. When compared with Support_Vector_Machine, Artificial NN, Naïve-Bayes, and Random Forest, the accuracy rate for Simple Cart is high. So, we identified that Simple CART is better algorithm than remaining all other algorithms in terms of recall and accuracy value. By analyzing those results it is evident that the betterresults was produced by Simple Cart algorithm in diagnosis of breast cancer.

Table 3 Analysis of different classification algorithms

Algorithms	Classification Accuracy	Statistical Parameters				
		AUC	Accuracy	F1	precision	recall
SVM	97.14%	0.901	0.971	0.929	0.914	0.932
Naive Bayes	93.47%	0.929	0.9347	0.916	0.933	0.9131
ANN	97.76%	0.951	0.9776	0.939	0.942	0.9145
Random Forest	94.69%	0.943	0.946	0.931	0.926	0.906
Simple Cart	98.23%	0.908	0.982	0.945	0.955	0.956

The study of accuracy rate for Support vector machine, ANN, Naïve_Bayes, Random Forest and Simple Cart learning algorithms is depicted in Figure 2 and the conclusion implies that the maximum accuracy rate for Simple Cart (SC) algorithm.

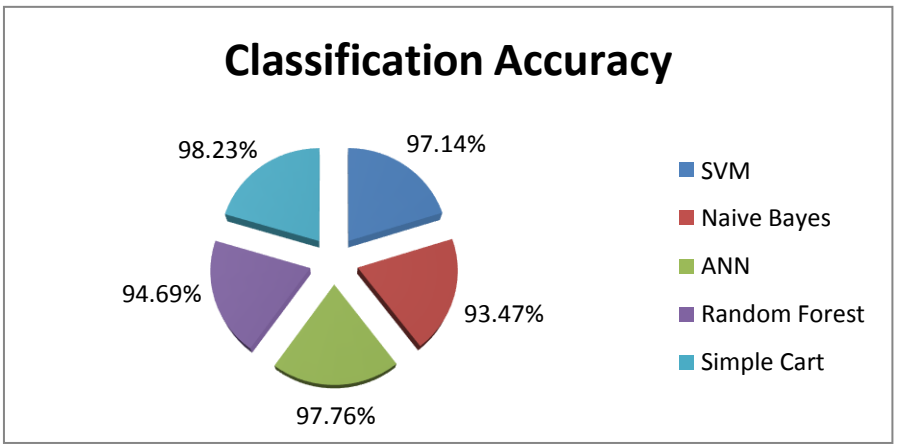

Figure 2: Accuracy rate for various machine learning algorithms

Figure 3 depicts the review of distinct statistical measures such as AUC, CA(Accuracy), F1, precision and recall values are evaluated by using different learning methods such as Support_Vector_Machine, Naïve Bayes, ANN, Random Forest and Simple Cart**.**

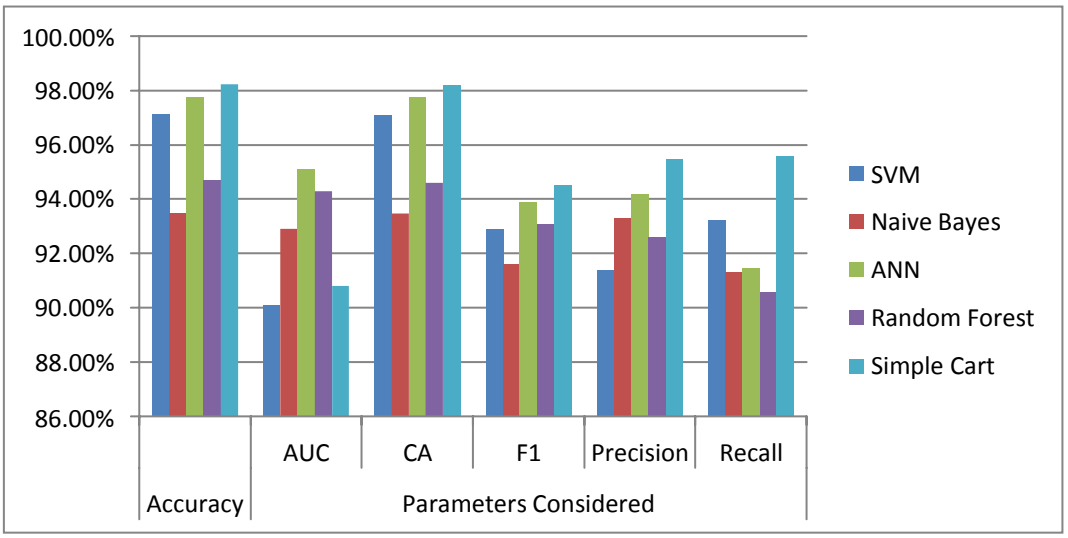

Figure 3: Analysis of Statistical Parameters

6. Conclusion:

A significant challenge in ML regions is to fabricate computationally productive and precise classifiers for biotic practices. Now a day, distinct machine learning algorithms are accessible to dissect medical data. In this paper, we utilized five among those algorithms such as SVM, Naïve Bayes, Artificial NN, Random Forest and Simple Cart on Breast Cancer Wisconsin (Diagnostic) Data Set. The effectiveness and efficiency of those algorithms are distinguished in terms of various statistical parameters such as AUC, Accuracy(CA), F1, precision and recall to find the best classification accuracy (CA). When compared with SVM, Naive Bayes, ANN and Random Forest, Simple CART scopes an maximum accuracy of 98.23%. In conclusion, Simple CART algorithm is authenticated to give better efficiency in predicting the symptoms, diagnosis of Breast Cancer and also attains the finest performance in terms of classification accuracy.

References:

[1] Ismail, N. S., & Southey, C. (2019, August). Breast cancer detection based on deep learning technique. In *2019 International UNIMAS STEM 12th Engineering Conference (Ancon)* (pp. 89-92). IEEE.

[2] Swathi, T. V., Krishna, S., & Rams, M. V. (2019, March). A survey on breast cancer diagnosis methods and modalities. In *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)* (pp. 287-292). IEEE.

[3] Amrane, M., Oukid, S., Gagaoua, I., & Ensari, T. (2018, April). Breast cancer classification using machine learning. In *2018 electric electronics, computer science, biomedical engineerings' meeting (EBBT)* (pp. 1-4). IEEE.

[4] Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, *83*, 1064- 1069.

[5] Kabiraj, S., Raihan, M., Alvi, N., Afrin, M., Akter, L., Sohagi, S. A., & Podder, E. (2020, July). Breast cancer risk prediction using XGBoost and random forest algorithm. In *2020 11th international conference on computing, communication and networking technologies (ICCCNT)* (pp. 1-4). IEEE.

[6] Padhi, T., & Kumar, P. (2019, January). Breast Cancer Analysis Using WEKA. In *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 229- 232). IEEE.

[7] Laghmati, S., Cherradi, B., Tmiri, A., Daanouni, O., & Hamida, S. (2020, September). Classification of patients with breast cancer using neighbourhood component analysis and supervised machine learning techniques. In *2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet)* (pp. 1-6). IEEE.

[8] Khuriwal, N., & Mishra, N. (2018, March). Breast cancer diagnosis using

eISSN 1303-5150

adaptive voting ensemble machine learning algorithm. In *2018 IEEMA Engineer Infinite Conference (eTechNxT)* (pp. 1-5). IEEE.

[9] Chaurasia, V., & Pal, S. (2020). Applications of machine learning techniques to predict diagnostic breast cancer. *SN Computer Science*, *1*(5), 1-11.

[10] Mandal, S. K. (2017). Performance analysis of data mining algorithms for breast cancer cell detection using Naïve Bayes, logistic regression and decision tree. International Journal Of Engineering And Computer Science, 6(2), 20388-20391.

[11] Amrane, M., Oukid, S., Gagaoua, I., & Ensari, T. (2018, April). Breast cancer classification using machine learning. In 2018 electric electronics, computer science, biomedical engineerings' meeting (EBBT) (pp. 1-4). IEEE.

[12] Gayathri, B. M., & Sumathi, C. P. (2016, December). Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer. In 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) (pp. 1-5). IEEE.

[13] Wang, H., Zheng, B., Yoon, S. W., & Ko, H. S. (2018). A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research*, *267*(2), 687-699.

[14] Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science, 83, 1064-1069.

[15] Kumari, M., & Singh, V. (2018). Breast cancer prediction system. *Procedia computer science*, *132*, 371-376.

[16] UCI Machine Learning Repository and breast cancer site:https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic).

[17] Yang, Y., Li, J., & Yang, Y. (2015, December). The research of the fast SVM classifier method. In *2015 12th international computer conference on wavelet active media technology and information processing (ICCWAMTIP)* (pp. 121-124). IEEE.

[18]. Ghosh, S., Dasgupta, A., & Swetapadma, A. (2019, February). A study on support vector machine based linear and non-linear pattern classification. In *2019 International Conference on Intelligent Sustainable Systems (ICISS)* (pp. 24-28). IEEE.

[19] Ghani, M. U., Alam, T. M., & Jaskani, F. H. (2019, November). Comparison of classification models for early prediction of breast cancer. In *2019 International Conference on Innovative Computing (ICIC)* (pp. 1-6).

eISSN 1303-5150

IEEE.

[20] Sharma, S., Aggarwal, A., & Choudhury, T. (2018, December). Breast cancer detection using machine learning algorithms. In *2018 International conference on computational techniques, electronics and mechanical systems (CTEMS)* (pp. 114-118). IEEE.

[21] Mishra, M., & Srivastava, M. (2014, August). A view of artificial neural network. In *2014 International Conference on Advances in Engineering & Technology Research (ICAETR-2014)* (pp. 1-3). IEEE.

[22] Nassif, A. B., Talib, M. A., Nasir, Q., Afadar, Y., & Elgendy, O. (2022). Breast cancer detection using artificial intelligence techniques: A systematic literature review. *Artificial Intelligence in Medicine*,102276.

[23] Brieman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and Regression Trees (Wadsworth, Belmont, CA). *Brodley CE and Utgoff PE (1995) Multivariate decision trees. Machine Learning*, *19*, 45-76.

eISSN 1303-5150