IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Comprehensive Comparison Of Different Data Mining And Machine Learning Techniques To Detect Breast Cancer

Aswini kumar mohanty
Capital engineering college, khurda

***Abstract*—**Breast cancer is the most deadly cancer and has highest mortality rate in women all over the world. One out of eight women over their lifetime will be diagnosed of breast cancer and it is recorded to be the world major cause of women's deaths. Early prediction of breast cancer can improve the survival rate of the patient. Consequently, high accuracy in cancer prediction is important to avoid any misdiagnosis. Machine learning algorithms and data mining methods are an effective way to classify data, especially in medical field, where those methods are widely used in diagnosis and analysis to make decisions. In this study, we have used five different classifiers of machine learning and data mining techniques which are Random Forest, Random Tree, Bayes net, Naïve Bayes and J48 on the breast cancer Wisconsin (Diagnostic) data set. A comprehensive comparison was made among the different classifiers. It is aimed to predict cancerous breast nodules and assess the correctness in classifying data with respect to efficiency and effectiveness of each machine learning algorithm in terms of accuracy, precision, sensitivity/recall and specificity. Experimental result indicates that Bayes net and Random Forest give the highest weighted average accuracy of 97.1% with lowest type I and II error rate. All experiments conducted in WEKA data mining tool.

1. Introduction

Breast cancer is the leading cause of death among women [1]. Breast cancer is seen in 1.5 million women yearly [1]. Breast cancers occur when one type of cell multiplies abnormally. The mass of tissues thus formed is called tumor. Tumors can be benign or cancerous. Tumors that are malignant are cancerous and can quickly spread to other organs, but benign tumors are not. Tumors are differentiated using several diagnostic methods like ultrasound, CT-scans, mammography and biopsy. However it is not very easy to differentiate the tumors even for a specialist. According to the recent research, the tumors can be classified based on statistical features. Doctors require a trustworthy diagnostic tool to classify the types of tumors. Even for specialists, however, distinguishing tumors is often challenging. In biopsy approach, the sample tissues are extracted and investigated. The breast sample are collected through a procedure called fine needle aspiration and analyzed under a microscope. Few statistical features are collected under microscope and are analyzed to predict the probability of a person having cancerous nodule. The relevant statistical features can be extracted and analyzed using various machine learning algorithm. This can significantly reduce the mortality rate of patients, by detecting the cancer at an early stage. In recent years, many machine learning approaches such as Linear Regression, Decision Tree, K-Nearest Neighbor, Random Forest, and Support Vector Machine have been utilized to categorize breast nodules. Feature selection is an important phase in classification where in the best features are selected before giving to the classification layer, which not only reduces the dimensionality of features but also improves the performance of the model.

During cell development in humans, cells develop as benign which has no negative effect on human but becomes very suspicious when this growth happens in the breast of human. Benign in the breast conditions are unusual growths or changes in the breast tissue that are not cancer.

However benign breast condition can be scary at first because the symptoms often mimic those caused by breast cancer. Although any lump formed by body cells may be referred to technically as a tumor. Not all tumors are malignant (cancerous). Most breast lumps–80% of those biopsied are benign (non-cancerous) and most breast lumps are benign tumors [2]. The diagnosis has always been a major problem in the medical field, based on various tests conducted on various patients. Tests are meant to aid the physician in making a proper and accurate diagnosis. However, miss diagnosis sometimes occurs, especially in tumor and cancerous cells since it can be difficult to make an accurate diagnosis, even for a medicinal cancer expert [3]. One of the drifting issues in the medicinal field is a diagnosis of the tumors. But early detection needs an accurate and reliable diagnosis procedure that allows doctors to differentiate benign breast tumors from malignant ones without going for surgical biopsy.

Breast cancer predictive model is investigation of the performance criterion of artificial intelligence and machine learning and several others for prediction, prognosis, detection and diagnosis of breast cancer. Data Mining (DM) is process of knowledge discovery in databases in which intelligent methods are applied in order to extract patterns. DM is set of techniques and tools applied to the non-trivial process of extracting and illustrating implicit knowledge, previously unknown, potentially useful and humanly coherent, from large data sets. A predictive model makes a prediction about values of data using known results found from different data as cited by Kharya S [4]. Data mining is the branch of computer science that used with bioinformatics for analyzing and classification biological data. The classification, clustering and other data mining methods can be exploited in combination with bioinformatics to leverage understanding of biological processes [5].

There are many data mining techniques for classification and forecast of BC outcome as Benign or Malignant [6]. In evaluation these algorithms different datasets are used by researchers to evaluate their performance.

Performance of a predictive model depends on the dataset and environment. The aim of this study is to evaluate the performance of Random Forest, Random Tree, Bayes Net, Naïve Bayes and J48 on Wisconsin Breast Cancer Diagnosis (WBCD) Dataset from UCI Machine Learning Repository created by Dr. William H. Wolberg at University of Wisconsin-Madison, US around the year 1989 and 1991 in terms of: Sensitivity, Specificity, Accuracy and Precision with a simulating environment Waikato Environment for Knowledge Analysis (WEKA). The WEKA is a collection of state-of-the-art machine learning algorithms and data pre-processing tool.

1.1 Breast cancer pervasiveness

Breast Cancer (BC) is reported to be second major cause of death in women today, [7-19]. Breast Cancer is the most commonly diagnosed disease among woman. According to the reported work by Mihaylov I stated that, one out of eight women over their lifetime will be diagnosed of Brest Cancer [20]. The report by Kumar V showed that, people used to say everyone knows someone living with breast cancer. This statement clearly indicates that Breast Cancer is common among female gender as well as a few in male genders also. In this paper the author, reported that Breast Cancer is one of the most harmful and wide spread disease among all of the diseases in medical science [21]. The authors also demonstated 18 cases of breast cancer for a very small community in Ghana in their work on a knowledge base of prevalent diseases in the original dataset report. It has been recorded by Kharya S and Soni S, that every 19 seconds somewhere around the globe a case of breast cancer is diagnosed among women and every 74 seconds a female pass on from breast cancer globally.

2. Related work

Several studies have tried with various improvisations on breast cancer classification utilizing appropriate machine learning algorithms, with superior outcomes in terms of classification system accuracy and sensitivity. Feature selection algorithms are crucial for extracting relevant features that are necessary for increasing a classification system's performance. Dhanya et al. proposed a method which uses feature selection algorithm like sequential forward feature selection, recursive feature elimination, f-test and correlation and classified the breast

nodules into benign and malignant using logistic regression, naive bayes and random forest algorithm [22]. Milon et al. used support vector machine and k-nearest neighbour for classification of breast nodules [23]. Ram et al. used principal component analysis to minimize the dimensionality of the data and then categorised them using logistic regression, k-nearest neighbours, and ensemble learning [24]. Dana et al. used support vector machine, random forest, and bayesian networks to classify the breast nodules and found that support vector machine provided the best accuracy [25]. Qaung et al. employed feature scaling and principal component analysis to reduce the number of features and then categorised them using an ensemble - voting classifier, logistic regression, SVM, and the adaboost algorithm [26]. Ahmed et al. employed logistic regression for feature reduction and classification of breast nodules [27]. Hasan et al. employed principal component analysis to reduce the number of features in the nodules and then used an artificial neural network to classify them. [28].Smita et al. employed principal component analysis for feature reduction and classified the nodules using decision tree algorithm such as CART and C4.5 [29]. Sujan et al. employed random forest algorithm after reducing the feature dimension using principal component analysis to classify the nodules and found that random forest provided better accuracy than decision tree [30]. Ahmet et al. classified the nodule using support vector machine with quadratic kernel after reducing the features using independent component analysis [31].Phonethepet.al used principal component analysis to reduce features and classified the data using J48 decision tree [32]. Liu selected highly relevant features using data correlation and independence test and classified the data using decision tree model [33]. Yang reduced the dimensionality of features using isomap and classified it using support vector machine [34]. Jain used genetic algorithm for feature selection and classified the feature using k-nearest neighbour classifier [35]. Emina used genetic algorithm to extract relevant features and classified them using Rotation Forest model [36]. Ed- daoudy used reduced feature set using association rules and classified the features using support vector machine [37]. Rahman extracted relevant features using genetic algorithm and classified using random forest [38]. Kamelused gray wolf optimisation for feature selection and classified the features using support vector machine [39]. Sharma extracted relevant features using correlation-based selection, information gain-based selection, and sequential feature selection, and then classified them using a max voting classifier [40]. The use of appropriate feature selection algorithm has significantly improved the accuracy of classification by selecting the relevant features for classification.

3. Materials and Methods

Systematically, this study will undergo this workflow in Figure 1 scientifically to present the results and conclusions (Table 2).Wisconsin Breast Cancer Diagnosis (WBCD) dataset.

The Wisconsin breast cancer (original) datasets link to the data (UCI) from the UCI machine learning Repository is used in this study. WBCD has 699 instances 2 classes 34.5% Malignant (M) and 65.5% Benign (B), and 11 attributes. The data was created by Dr. William H. Wolberg at University of Wisconsin-Madison, US around the year 1989 and 1991. Sample of the data is tabled with 10 variable in Table 2 and attributes of the data are presented in Table 1.

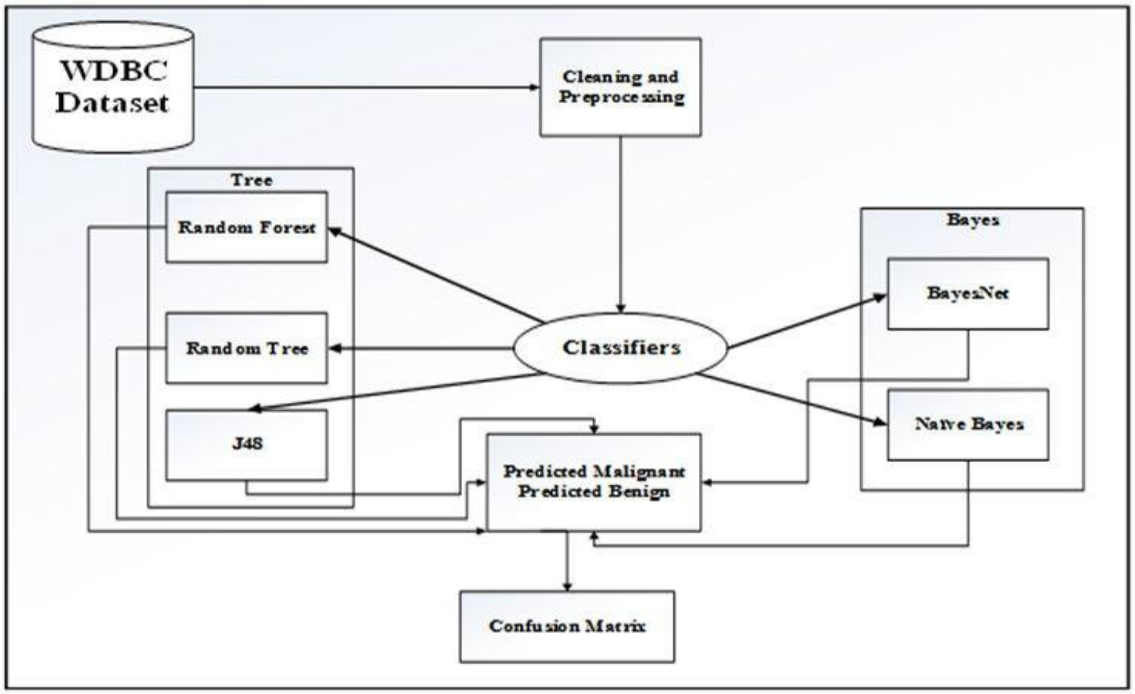

Figure 1. Conceptual model.

Table 1. Attributes of the dataset.

Clump Thickness (CT)	Uniformity of Cell size (UC)	Uniformity of cell Shape (US)
Marginal Adhesion (MA)	Single Epithelial Cell Size (SE)	Bare Nuclei (BN)
Bland Chromatin (CB)	Normal Nucleoli (NN)	Mitoses (MI)

Table 2. WBDC dataset sample.

CT	UC	US	MA	SE	BN	CB	NN	MI	CLASS
5	1	1	1	2	1	3	1	1	B
4	1	1	3	2	1	3	1	1	B
8	7	5	10	7	9	5	5	4	M
1	1	1	1	2	1	1	1	1	B
10	10	7	8	7	1	10	10	3	M
5	7	4	1	6	1	7	10	3	M
1	1	1	1	2	1	1	1	1	B
4	8	8	5	4	5	10	4	1	M

These data records were created in Excel sheet, saved in CSV format and it was then converted to ARFF format for it to be readable by WEKA. Value for each attribute is integer ranges of 1-10 inclusive.

4. Experiment

In order to compare the performance of Random Forest, Random Tree, J48, Baye Net and Naïve Baye, an experiment is conducted on bench mark dataset. All experiments on the classifiers described will be conducted using libraries from Weka machine learning environment.
WEKA contains a collection of machine learning algorithms for data pre-processing, classification, regression, clustering and association rules. Machine Learning techniques implemented in WEKA are applied to a variety of real world problems. The program offers a well-defined framework for experimenters and developers to build and evaluate their models.

5. Classifiers

Classifier as component of Figure 1 consists of the 5 techniques: Random Forest (Rand. F), Random Tree (Rand. T), J48 are Tree classifiers and Bayes Net (BN) and Naïve Bayes (NB) are: bayes classifier. In classifier the WBCD dataset was used in each of the classifier and 10-folds cross validation was applied on the dataset. Using 10-folds cross validation means, the dataset was broken down into ten sets. Each set represented 10% from the original dataset to allow every slice of the dataset to take a turn as a testing data. For each round, the experiment used nine sets for training process and the reminder ones for the testing process.

In executing each of the classifier in the WEKA the running time to build the classifier in generating of confusion matrix is recorded in Table 3 and the confusion matrix for each of the classifier is presented in Table 4. Table 5 also illustrates the performance of each classifier of terms of class Table 7.

Table 6 presented the weighted average performance for each of the classifier.

Table 3. Execution time of the classifiers.

Classifier	Time to build the model/sec.
Ran. T	0.08
Ran.F	0.66
J48	0.19
BayesNet	0.2
Naïve Bayes	0.11

Table 4. Confusion matrix of the classifier.

Classifier	B	M	Class
Ran. T	442	16	B
	14	227	M
Ran.F	444	14	B
	6	235	M
J48	438	20	B
	18	223	M
BayesNet	442	16	B
	4	237	M
Naïve Bayes	436	22	B
	6	235	M

Table 5. Deductions from confusion matrix with respect each class.

Classifier	TPR	FPR	Acc.	MCR	Precision	F-Measure	MCC	ROC Area	PRC Area	Class
Ran. T	0.965	0.058	0.964	0.036	0.969	0.967	0.905	0.956	0.96	B
	0.942	0.035	0.942	0.058	0.934	0.938	0.905	0.955	0.907	M
Ran. F	0.969	0.025	0.969	0.031	0.987	0.978	0.937	0.989	0.994	B
	0.975	0.031	0.975	0.025	0.944	0.959	0.937	0.989	0.972	M
J48	0.956	0.075	0.956	0.044	0.961	0.958	0.88	0.955	0.955	B
	0.925	0.044	0.967	0.033	0.918	0.921	0.88	0.955	0.902	M
Bayes Net	0.965	0.017	0.965	0.035	0.991	0.978	0.938	0.992	0.996	B
	0.983	0.035	0.983	0.017	0.937	0.96	0.938	0.992	0.982	M
Naïve	0.952	0.025	0.952	0.048	0.986	0.969	0.914	0.988	0.995	B
Bayes	0.975	0.048	0.975	0.025	0.914	0.944	0.914	0.983	0.942	M

Table 6. Weighted average (W) for classifier.

Classifier	TPR	FPR	Acc.	MCR	Precision	F-Measure	MCC	ROC Area	PRC Area
Ran. T	0.957	0.05	0.957	0.043	0.957	0.957	0.905	0.956	0.942
Ran. F	0.971	0.027	0.971	0.029	0.974	0.971	0.937	0.989	0.972
J48	0.946	0.064	0.946	0.054	0.946	0.946	0.88	0.955	0.937
Bayes Net	0.971	0.023	0.971	0.029	0.972	0.972	0.938	0.992	0.991
Naïve Bayes	0.96	0.033	0.96	0.04	0.962	0.96	0.914	0.986	0.976

Table 7. Performance metrics.

S/N	Name	Description	Formula
Eqn (1)	Precision (Prec)	Cases the classifier predicts Malignant, how often is it correct	$Prec = \frac{TP}{TP + FP}$
Eqn (2)	Recall	Cases where it's actually Malignant, how often does the classifier predicts Malignant, however it sometime called sensitivity or True Positive Rate (TPR)	$TPR = \frac{TP}{TP + FN}$
Eqn (3)	F-Measure	F-measure or score is Harmonic Mean of Precision and Recall.	$F - measure = \frac{2 * TPR * Prec}{TPR * Prec}$
Eqn (4)	MCR	Wrong made by the classifier or error rate	$MCR = \frac{FP + FN}{TP + FP + FN + TN}$
Eqn (5)	Accuracy (Acc)	Correct predictions made by the classifier	$Acc = \frac{TP + TN}{TP + FP + FN + TN}$

6. Results and Discussion

Figure 2 presents TPR (sensitivity) for the positive class and TNR (Specificity) for Negative class i.e. Cases where it is actually malignant/benign, how often does the classifier predicts malignant/benign respectively. Bayes Net had the highest sensitivity of 98.3%and random forest highest specificity of rate 96.9%

Figure 3 presents the accuracy of the classifier which is the correct predictions made by each classifier. Naïve Net had the highest accuracy for the positive class of 98.3%. Random forest had highest accuracy rate for the Negative class of 96.9%. Overall accuracy of the classifier was 97.1% which was attained by random forest and Naïve Bayes.

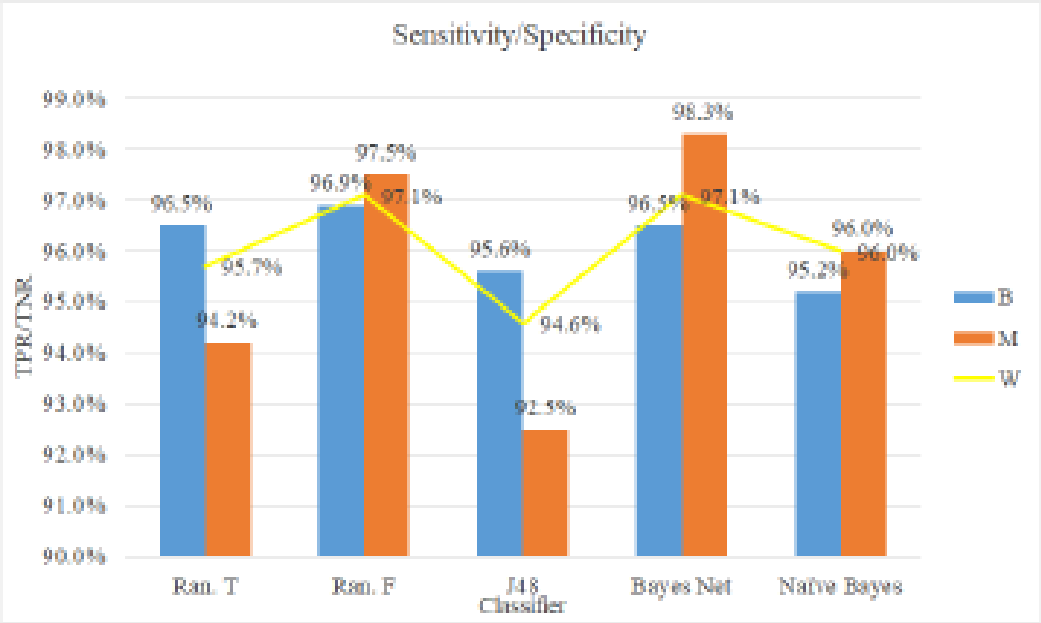

Figure 2. True positive/negative rate for each classifier.

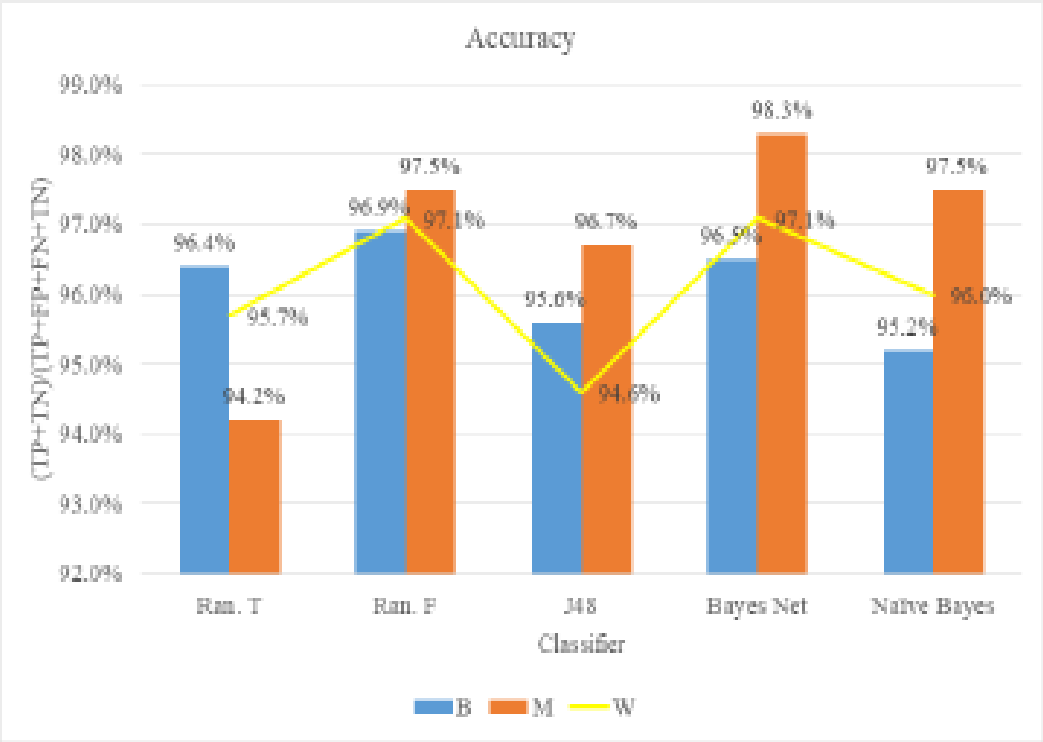

Figure 3. Accuracy of each of the classifier.

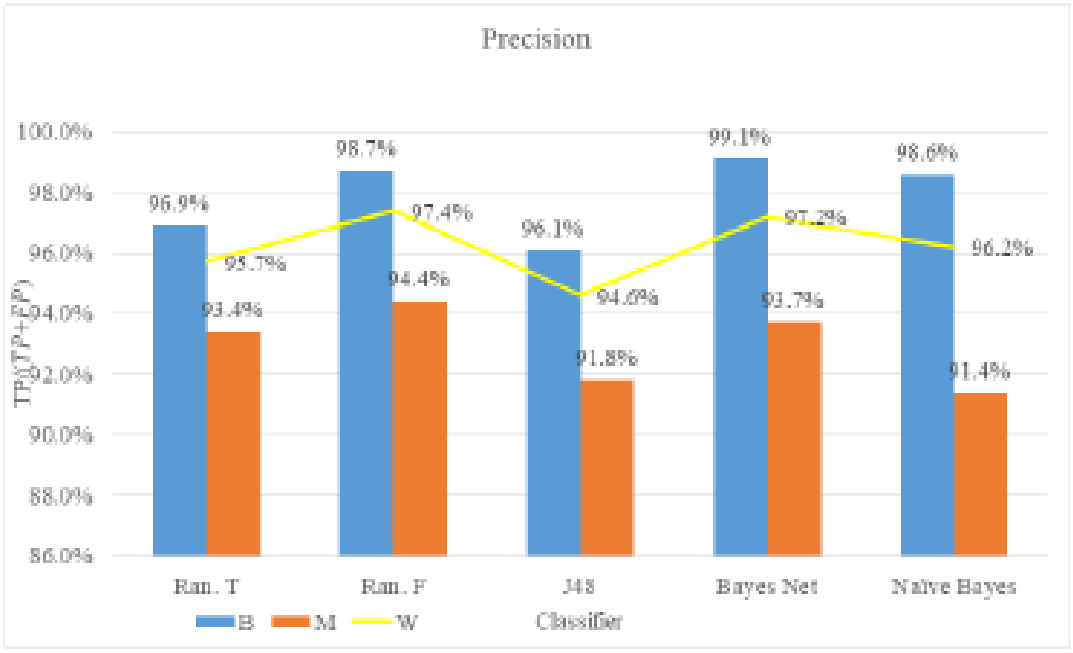

Figure 4. Precision rate of each classifier.

Figure 4 reports on precision i.e. cases the classifier predicts malignant, how often is it correct for the positive class. Random forest had the highest of rate 94.4%. The weighted average for a classifier random forest of rate 97.4%. For the negative class precision i.e. case the classifier predicts benign how often is it correct Bayes Net had the highest rate of 99.1%. With W of rate 93.7%. Overall random forest had the highest precision of 97.4%.

7. Conclusion

Breast cancer is a dread disease and its quick detection is most urgent, Many researchers as well as academicians are rigorously trying to detect the disease early by several methods. In many cases the detection is not producing the better accuracy which is a matter of concern. In this study presented, aimed to assess the performance of different classifications of data mining and machine learning algorithms in term of accuracy, sensitivity, specificity and precision. The outcome of the study concluded that random forest has the highest performance in terms of specificity (96.9%), accuracy (97.1%) and precision (97.4%). Bayes Net and Random Forest had the highest accuracy of 97.1%. Bayes Net had the highest sensitivity of 98.3%. Bayes Net and random forest took 0.2sec and 0.66sec to build the classifier respectively. Random forest took a lot of time in the learning process.

8. References

[1] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram,Rebecca L. Siegel , Lindsey A. Torre, Ahmedin Jemal,"Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", CA: cancer journal for clinicians ,Volume 68, Issue 6;68: pp 394–424, 2018.

[2] Baldwin CM. “Different Kinds of Breast Lumps,” Stony Brook Cancer Center, Breast Care Center, New york. 2013.

[3] Hamza A. “An Enhanced Breast Cancer Diagnosis Scheme based on Two-Step-SVM Technique,” Int J Adv Comput Sci Appl 8 (2017): 158-165.

[4] Kharya S, Agrawal S and Soni S. “Naive Bayes Classifiers: A Probabilistic Detection Model for Breast Cancer,” Int J Comput Appl 92 (2014): 26-31.

[5] Aavula R. “A Survey on Latest Academic Thinking of Breast Cancer Prognosis”. Int J Appl Eng Res 13 (2018): 5207–5215.

[6] Asri H, Mousannif H, Al H and Noel T. “Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis.” Procedia Comput Sci 83 (2016): 1064–1069.

[7] Kharya S and Soni S. “Weighted Naive Bayes Classifier: A Predictive Model for Breast Cancer Detection.” Int J Comput Appl 133 (2016): 32–37.

[8] Mandal S. “Performance Analysis of Data Mining Algorithms for Breast Cancer Cell Detection Using Naïve Bayes, Logistic Regression and Decision Tree.” Int J Eng Comput Sci 6 (2017): 20388–20391.

[9] Asri H, Mousannif H, Moatassime H, and Noel T, et al. “Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis.” Procedia Comput Sci 83 (2016): 1064–1069.

[10] Sankareswari MSA and Phil M. “A Proportional Learning of Classifiers Using Breast Cancer Datasets.” Int J Comput Sci Appl 3 (2014): 223–232.

[11] Kumari M and Singh V. “Breast Cancer Prediction system.” Procedia Comput Sci 132 (2018): 371–376.

[12] Edriss E, Ali E, and Feng WZ. “Breast Cancer Classification using Support Vector Machine and Neural Network.” Int J Sci Res 5 (2016): 1–6.

[13] Ayele F. “Constructing a Predictive Model for Detection of Breast Cancer.” Int J Comput Sci Eng 8 (2018): 17529–17532.

[14] Valluri R and Sowjanya M. “Prediction of Breast Cancer Using Stacking Ensemble Approach.” Int J Manag Technol Eng 9 (2019): 1857–1867.

[15.Kumar V, Mishra BK, Mazzara M and Thanh DNH, et al. “Prediction of Malignant & Benign Breast Cancer: A Data Mining Approach in Healthcare Applications,” Adv Data Sci Manag 37 (2019): 1–8.

[16] Islam MM, Iqbal H, Haque MR and Hasan MK, et al. “Prediction of breast cancer using support vector machine and K-Nearest neighbors,” 5th IEEE, Dhaka, Bangladesh. 2018. 226–229.

[17] Chaurasia V, Pal S and Tiwari BB. “Prediction of benign and malignant breast cancer using data mining techniques.” J Algorithms Comput Technol 12 (2018): 119-126.

[18] Shajahaan SS, Shanthi S and Manochitra V. “Application of Data Mining Techniques to Model Breast Cancer Data.” Int J Emerg Technol Adv Eng 3 (2013): 1-8.

[19] Talukdar J and Kalita SK. “Detection of Breast Cancer using Data Mining Tool (WEKA).” Int J Sci Eng 6 (2015): 1124-1128.

[20] Mihaylov I, Nisheva M and Vassilev D. “Application of machine learning models for survival prognosis in

breast cancer studies." Inf 10 (2019): 1-13.

[21] Appiah S, Adekoya AF, Bapuuroh C and Akowua-kwakye C. "Health and Medical Informatics A Knowledge-Base of Prevalent Diseases in Sunyani Municipality, Ghana Using Ontological Engineering," J Heal Med Informatics 10 (2019): 8.

[22] 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," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019, pp. 1049-1055.

[23] M. M. Islam, H. Iqbal, M. R. Haque and M. K. Hasan, "Prediction of breast cancer using support vector machine and K-Nearest neighbors," *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, 2017, pp. 226-229

[24] R. MurtiRawat, S. Panchal, V. K. Singh and Y. Panchal, "Breast Cancer Detection Using K-Nearest Neighbors, Logistic Regression and Ensemble Learning," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 534-540, doi: 10.1109/ICESC48915.2020.9155783.

[25] D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), 2016, pp. 1-4

[26] Q. H. Nguyen *et al.*, "Breast Cancer Prediction using Feature Selection and Ensemble Voting," *2019 International Conference on System Science and Engineering (ICSSE)*, 2019, pp. 250-254.

[27] A. F. Seddik and D. M. Shawky, "Logistic regression model for breast cancer automatic diagnosis," *2015 SAI Intelligent Systems* Conference (IntelliSys), 2015, pp. 150-154.

[28] H. Hasan and N. M. Tahir, "Feature selection of breast cancer based on Principal Component Analysis," *2010 6th International* Colloquium on Signal Processing & its Applications, 2010, pp. 1-4

[29] S. Jhajharia, S. Verma and R. Kumar, "A cross- platform evaluation of various decision tree algorithms for prognostic analysis of breast cancer data," 2016 International Conference on Inventive Computation Technologies (ICICT), 2016, pp. 1-7.

[30] S. Ray, A. AlGhamdi, A. AlGhamdi, K. Alshouiliy and D. P. Agrawal, "Selecting Features for Breast Cancer Analysis and Prediction," 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), 2020, pp. 1-6.

[31] A. Mert, N. Kilic and A. Akan, "Breast cancer classification by using support vector machines with reduced dimension," *Proceedings ELMAR-2011*, 2011, pp. 37-40.

[32] P. Douangnoulack and V. Boonjing, "Building Minimal Classification Rules for Breast Cancer Diagnosis," 2018 10th International Conference on Knowledge and Smart Technology (KST), 2018, pp. 278-281.

[33] L. Yi and W. Yi, "Decision Tree Model in the Diagnosis of Breast Cancer," *2017 International Conference on Computer Technology, Electronics and Communication (ICCTEC)*, 2017, pp. 176-179.

[34] X. Yang, H. Peng and M. Shi, "SVM with multiple kernels based on manifold learning for Breast Cancer diagnosis," *2013 IEEE International Conference on Information and Automation (ICIA)*, 2013, pp. 396-399.

[35] Jain, R., Mazumdar, J. A genetic algorithm based nearest neighbor classification to breast cancer diagnosis. Australas Phys Eng Sci Med 26, 6 (2003)

[36] Aličković, E., Subasi, A. Breast cancer diagnosis using GA feature selection and Rotation Forest. Neural Comput & Applic 28, 753–763.

[37] Ed-daoudy, A., Maalmi, K. Breast cancer classification with reduced feature set using association rules and support vector machine. Netw Model Anal Health Inform Bioinformatics 9, 34 (2020).

[38] El_Rahman, S.A. Predicting breast cancer survivability based on machine learning and features selection algorithms: a comparative study. J Ambient Intell Human Comput 12, 8585–8623 (2021).

[39] Kamel, S.R., YaghoubZadeh, R. & Kheirabadi,M. Improving the performance of support-vector machine by selecting the best features by Gray Wolf algorithm to increase the accuracy of diagnosis of breast cancer. J Big Data 6, 90 (2019).

[40] Sharma, A., Mishra, P.K. Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis. Int.j. inf. tecnol. (2021).