Research/Review Article

FEDPOLEL JOURNAL OF APPLIED SCIENCES
https://fedpoleljournalofappliedsciences.com/

Application of Machine Learning to Breast Cancer Diagnosis

Bukola Badeji-Ajisafe[1*], Adewole L.B.[2.], and Adetunmbi A.O.[3]

University of Medical Sciences, Ondo[1]
Federal University Oye – Ekiti[2]
Federal University of Technology, Akure[3]

*bolubadeji@unimed.edu.ng

Article Info

Received: 28/04/2021
Accepted: 17/09/2021

Keywords

Breast Cancer, Machine Learning, Bagging Ensemble, Decision Tree, Support Vector Machine

Abstract

Breast Cancer is a major health crisis globally, and it is the most frequent cancer in women which is associated with fibrous tissue proliferation. The tissue surrounding growth contract clinically, and presents as damping of the skin and in drawing of the nipple. The need for urgent and accurate diagnosis necessitates the application of machine learning techniques to breast cancer diagnosis. Machine learning technique has been used for extensive research in breast cancer diagnosis with emphasis on diagnosis accuracy through the simplification of the disease either by performing feature selection, in addition to other pre-processing steps, or not before classification. In this paper, two ensemble approach namely: Bagging and random forest methods on three base-learners – Support Vector Machine (SVM), Decision Tree (DT), and Naïve Bayes (NB) were used and implemented to show that few necessary attributes may be required for BC diagnosis. Prior to classification, data pre-processing was carried out to handle missing values and data scaling to normalize the range of independent variables. UCI dataset, with 9 features, was used for the study. The results show that all the techniques show roughly similar results on four different metrics (Accuracy, Precision, Recall, and False Alarm Rate)..

1. INTRODUCTION

Breast cancer (BC) is one of the most frequently diagnosed cancers in women worldwide with millions of new cases diagnosed yearly and recorded by the public health expert. It is the second commonest cancer with increasing incidence rate in women between the ages of 45-55. [1,2] rated BC as the second most common cause of death after lung cancer in the West and one of the most dangerous diseases among women in Nigeria.

The mortality and morbidity rate ranges from 36.5 to 50.2 per 100,000 casualties, due to lack of timely and accurate diagnosis most especially in developing countries of the world where there are gross inadequate specialist and required facilities to aid its diagnosis and treatment most especially in rural communities. With limited health facilities, effective detection and diagnosis of patients is difficult. These are the established diagnostic tools for various diseases and disorders, and play a major role in cancer-diagnosis. Supplementing this technique with automated classification and segmentation tools is gaining importance to reduce errors and time needed to make a conclusive diagnosis. The need to have efficient and effective clinical detection and diagnosis necessitates the need for a predictive model that will assist the medical expert in the accurate diagnosis of the affected patients. A number of machine learning approaches have applied to breast cancer diagnosis such as AdaBoostMI, Real AdaBoost [4], Artificial Neural Network [5], Support Vector Machine [4,5], Decision Tree [4,5,6], Naïve Bayes [4,5,6,], among others.

2. REVIEW OF RELATED LITERATURE

This section presents the existing literature consulted on Breast Cancer disease diagnosis using data mining approach.

[4] proposed breast cancer diagnosis models using ensembles approach on these base learners - Random Forest, Radial basis function and neural network algorithms, while the ensemble techniques are AdaBoostMI, Real AdaBoost, and MultiBoost, to improve prediction accuracy. Predictions of the three ensemble techniques exceed or equals to 97% while ANN returns the worst prediction accuracy of 88% and requires more time training time. [5] presented a systematic review of ensemble approaches used in breast cancer classification. 193 articles were reviewed in which homogeneous, single classification, and ensembles approaches were used for the analysis, it was reported that the homogeneous approach was the most widely used in solving classification problem and the ensemble approach shows that there are unresolved issues in the field of breast cancer that need the researcher's attention. The overall result shows that ensemble approaches outperformed a single classifier in terms of accuracy.

Also, [6] measured how automated ensemble learning approach was used to improve diagnosis and treatment of the disease at the early stage. Six popular ensemble methods and fourteen base learners were used for automatic detection of breast cancer. The empirical result shows that ensemble learning can improve predictive performance of the base learners on a medical domain and from the comparative experiments, random subspace ensemble outperformed others. However, the study did not measure the performance of other classifiers like the neural network with random subspace for breast cancer diagnosis.

[7] classified breast cancer dataset using ensemble approach with three base learner such as neural fuzzy (NF), K-nearest neighbour (KNN), and quatric classifier. The result shows that the ensemble approach has the highest accuracy performance when compared to that of the individual single model. While[8] research focus on different stages in cancer, in which diagnosis is done at the early stage through automation using different machine learning technique for the development of predictive models. The results show an effective and accurate decision making. In [9] their research was based on genetic programming and machine learning algorithms with the aim of constructing and optimizing a learning algorithm system that will accurately differentiate between benign and malignant breast tumours. The genetic programming technique was used to select the best features and perfect parameter values of the machine learning classifiers and the performance was measured based on sensitivity, specificity, precision, accuracy, and the ROC curves. The result shows that the genetic programming can automatically find the best model by combining feature pre-processing methods and classifier algorithms. [3] reported that the application of imaging technologies in the detection of BC is basically through X-ray mammography, Computer tomography (CT), and Magnetic resonance imaging (MRI).

[10, 16] used machine learning tool in classification of breast cancer, the research was implemented using feed forward back propagation network (FFBPN) for classification of breast cancer cases to malignant or benign. The aim was to design an Artificial Neural Network (ANN) with high and acceptable level of accuracy by selecting the number of hidden layers, number of neurons in the hidden layer and the type of activation functions in hidden layers. Validation of the research were obtained from Wisconsin Breast Cancer Database (WBCD)using three transfer functions such as LOGSIG, TANSIG, and PURELINE in neural network architectures. The result generated show that ANN

performance of different number of neurons in hidden layer 20, 21, 22, 23, 24 neurons show that the best network design is that one with three hidden layers, 21 neurons in the hidden layer, and TANSIG as activation function.
[11] employed six machine learning techniques - GRU-SVM[1], Linear Regression, Multilayer Perceptron (MLP), Nearest Neighbor (NN) search, Softmax Regression, and Support Vector Machine (SVM) classifiers to determine the best performing algorithm on breast cancer diagnosis dataset in terms of prediction accuracy. The performance of all the classifiers are adjudged very good but revealed that multiplayer perception 99% outperformed others for breast cancer detection. Ensemble approach not exploited.
All the research presented in the review specified the need for further research on breast cancer diagnosis which would include an ensemble approach for better classification. Hence, the need for this study arises.

3. METHODOLOGY

Analysis of the various techniques of breast cancer diagnosis was carried out using the UCI data repository [11], which has a total of 700 instances with nine conditional attributes and one decision attribute. All the attributes were discretized with variable field length as shown in Table 1. Each instance in the dataset was assigned either Yes or No depending on the patient's condition. The combination of the available conditional attributes (symptoms) gives room for classification (diagnosis) of each record. This decision attribute denotes the result of the diagnosis carried out. Each of the conditional attributes was assigned a class, there are **two** classes of breast cancer in this case, cancerous, non-cancerous (breast cancer never diagnosed normal, Seventy Percent (70%) partition from this data was used as a training set and 30% as a testing set. The number of records in training is 490 and 210 for testing.

The features of the UCI Wisconsin Breast Cancer Data set is given in Table 1.

Table 1: Attributes of Breast Cancer

Data ID	Abbreviation	Attribute	Attribute type	Field Length
1	CT	Clump thickness	Discrete	Variable
2	BN	Bare Nuclei	Discrete	Variable
3	MM	Mitoses	Discrete	Variable
4	CH	Cell shape	Discrete	Variable
5	CO	Chromatin	Discrete	Variable
6	NN	Normal Nucleoli	Discrete	Variable
7	EP	Epithelial	Discrete	Variable
8	CS	Cell size	Discrete	Variable
9	AD	Adhesion	Discrete	Variable

3.1 Features Selection Analysis

Information gain and Gain Ratio was used in ranking and selection of best subset attributes from the dataset which will improve classification performance and computational time.

3.1.1 Information Gain

It studied the information content, let node N represents tuples of partition D. the attributes with the highest information gain is chosen as the splitting attribute for node N.
The amount of information needed after the partitioning to arrive at an exact classification is measured by

$$Info_A(D) = -\sum_{j=i}^{v} \frac{|Dj|}{|D|} * info(D_j) \quad (1)$$

The term $\frac{|Dj|}{|D|}$ act as the weight of the *jth* partition. $Info_A(D)$is the expected information required to classify a tuple D based on partitioning by *A*. The smaller the expected information required, the greater the purity of the partitions. Information gain is defined as the difference between the original information requirement (that is, based on just the proportion of classes) and the new requirement (that is, obtained after partitioning on A). That is,

$$Gain(A) = Info(D) - Info_A(D) \quad (2)$$

Gain(A) tells how much would be gained by branching on A. It is expected reduction in the information requirement caused by knowing the value of A. The attribute with the highest information gain, (Gain(A)), is chosen as the splitting attribute at node N.

3.1.2 Gain Ratio

Gain ratio is an improvement on information gain, which is biased towards test with many outcomes. To overcome this bias there is need for GR which is used in selecting relevant features to improve classification performance and reduce computational time, since it applies normalization to information gain using a split information value defined analogously with *info (D).*

$$Splitinfo(A) = -\sum_{j=i}^{v} \frac{|Dj|}{|D|} * \log_2\left(\frac{|Dj|}{|D|}\right) \quad (3)$$

This values represents the potential information generated by splitting the training data set, D, into v partitions, which correspond to the outcomes of a test on attribute A for each of outcome, it considers the number of instances having that outcome with respect to the total number of instances in D. the attribute with the maximum gain ratio is selected as the splitting attributes. The gain ratio is defined as

$$Gain\ Ratio = \frac{Gain(A)}{Splitinfo(A)} \quad (4)$$

3.1.3 Illustrating the concept of classification using information Gain and Gain Ratio

Table 2 shows breast cancer cases of 13 patients with four conditional attributes clump thickness (CT), Cell Shape (CH), Cell Size (CS), Adhesion (AD) and one decision attributes.

Table 2: Sampled Breast Cancer information

Data ID	CT	CS	CH	AD	EP	MM	BN	CO	NN	Class ID
1	No	Yes	Yes	No	No	No	Yes	No	Yes	Non-cancerous
2	Yes	No	No	Yes	Yes	Yes	No	Yes	No	Non-cancerous
3	No	Yes	No	No	No	No	Yes	No	Yes	Non-cancerous
4	Yes	No	Yes	No	No	Yes	No	No	No	Non-cancerous
5	Yes	Yes	No	No	No	Yes	Yes	No	Yes	Non-cancerous
6	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Non-cancerous
7	Yes	No	No	No	No	Yes	No	No	No	cancerous
8	No	No	No	No	No	No	No	No	Yes	cancerous
9	No	No	No	Yes	Yes	No	No	Yes	No	cancerous
10	No	No	No	No	No	No	No	No	Yes	cancerous
11	Yes	No	No	No	Yes	Yes	No	No	No	cancerous
12	No	No	Yes	No	No	No	No	No	No	cancerous
13	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Non-cancerous

$$\sum_{i=1}^{m} P_i \log_2(P_i) \quad (5)$$

3.1.4 Gain Ratio as the Splitting Criteria

Computation of gain ratio for the attribute CT based on this equation

$$Splitinfo(A) = -\sum_{j=i}^{v} \frac{|Dj|}{|D|} * \log_2\left(\frac{|Dj|}{|D|}\right) \quad (6)$$

$$Gain\ Ratio = \frac{Gain(A)}{Splitinfo(A)} \quad (7)$$

The splitting will start with Cell Size and Cell Size will serve as the root node in this case since it has the highest Gain Ratio, others will be based on the comparison of other Gain Ratio.

4. Model Building

The training phase which is made up of a set of *M* different classifiers *(C1,C2,…,Cm)* was trained and the test data was passed into this classifiers to make predictions. The prediction of the classifiers was combined into an ensemble (Bagging) using majority voting algorithm which was used in aggregating the performance of the classifiers. The algorithms that were used to train the classifiers are Naive Bayes, Support vector Machine and Decision tree algorithm.

4.1. Decision Tree (DT): Decision tree uses divide and conquer approach to construct a decision rules for solving classification problems using information gain ratio which avoids the bias of selecting attributes with many values [13] . In data mining classifications learning, the goal of prediction using decision tree algorithm is to learn a mapping function from input variable x to output variable c , given a labeled set of input-output pairs as: Algorithm is to learn a mapping function from input variable x to output variable c , given a labeled set of input-output pairs as:

$$D = \{(x, c)\}\frac{N}{i=1} \quad (8)$$

In Eq. (8) $\mathcal{D}$ is called the training set, and N is the number of training samples. In the simplest setting, each training input x is a D -dimensional vector of numbers, which are called features or attributes. Also the response variable ***c*** is the class output variable. Decision tree makes its classification using information gain which is a measure of the differences in entropy from before to after the current set S for which entropy is calculated is split on an attribute A . Therefore, in order to make classification, the attribute with the highest information gain is seen as the best classifier for making decision, and this is calculated as:

$$Entropy\ H(S) = -\sum_{i=1}^{n} p(C)\log_2 p(C) \quad (9)$$

H(S) = entropy of current set, C = set of class Non-cancerous and cancerous, n = number of attributes.

Accordingly, the information gain by a training dataset is defined as:

$$info\ Gain\ (S) = H(S) - \sum_{I^n}^{n} \frac{|S_i|}{|S|} H(S) \quad (10)$$

4.2. Random forest – is an ensemble classification algorithm which is induced from boot strap samples of training data, using random feature selection in the tree induction process. Prediction is made by aggregating (majority vote for classification or averaging for regression) the predictions of the ensemble. Random forest uses information content of a mode as the splitting criterion. Information content is defined as$I(N) = |S|H(S) - |S_L|H(S_L)|S_R|H(S_R)$

Where $|S|$= input sample size*;* $|S_{L,R}|$= size of left, right subclasses of *S.* $H(S)$= Shannon entropy of$S = \sum_{i=\pm 1}^{m} P_i\, log_2(P_i)$.

At each tree split, a random sample of m features is drawn from P variables and only those m features are considered for splitting.

$$m = \sqrt{p}\ or\ log\ 2^p \quad (11)$$

Where P is the number of features. The best variable /split-point is picked among the m and node splits into two daughters nodes. Boots-trapping is applied at the beginning to generate different subsets leading to different trees. The trees are then ensemble using majority vote or regression method.

4.2 Support Vector Machine (SVM)

[12]defined SVM as finding hyperplane in a space different from that of the input data x, it can also be defined a hyperplane in features space induced by a kernel K (the kernel defines a dot product in that space. SVM realizes two things: the hypothesis space used by SVM, and the loss functions used. The SVM find an optimal hyperplane as the solution to the learning problem. The simplest formulation of SVM is the linear one where the hyperplane lies on the space of the input data x. The hypothesis space is a subset of all hyperplane of the form:

$F(x) = w.x + b.$ (12)

SV classification

$min\|f\|_x^2 + c\sum_{i=1}^{i} \varsigma$ (13)

Subject to: $y_i f(x_i) \geq 1 - \xi_{ix}\ for\ all\ i.\ \ \xi_{ix} \geq 0x$

Variable ξ_i are called slack variable and the error made at point (x_i, y_i).

4.3 Naïve Bayesian Classification

It is a probabilistic model of what is happening in data, which estimates the class for new data item. [13, 14, 15] applied Naïve Bayesian in solving various problems and the result were successful. Below is the step by step application of naïve bayes:

(i). Given a training set of tuples and their associated class labels. The training data has n-attributes, which may be categorical or numeric, {A1, A2, …, A3} and each tuple is represented by an n-dimensional attribute vector X = ($x1$, $x2$,…, xn), depicting n measurements made on the tuple from n-attributes.

(ii). Suppose that there are k classes, C1, C2,…,Ck and each data point belongs to one of the k classes. The goal is to develop a Bayes classifier M(X) →Ci , where X can be any point , not necessarily a member of the training data. Given a tuple, X, the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. That is, the naïve Bayesian classifier predicts that tuple X belongs to the class Ci if and only if $P(\ c_i \mid X\ ) > P(c_j \mid \leq j \leq k, j \neq i)$.

4.4 The Bagging Ensemble Approach

The approach was used to improve the stability of the classification accuracy in [16]. The basic idea of bagging ensemble is to aggregate the predictions generated from data mining classifiers (Naïve Bayes, Decision tree and Support vector Machine) using a bootstrap sample of the training data to generate new hypothesis. Bootstrap samples is generated by sampling D^1 instances from the training set samples with replacement ($D^1 \leq D$). During the bagging Ensemble process, the base learners Decision Tree, Naive Bayes and Support vector machine used in this study are trained and built from each bootstrap sample of the original Breast Cancer dataset and their predictions are combined with bagging ensemble using majority voting method to form a consensus model for making final decision used in classification. It can be determined using Eq. (13).

$C_f = \underset{i}{\text{argmax}} \sum_{i=1}^{m} Jry_i$ (13)

Where J_r represents the decision of the r^{th} classifier given class y_i, C_f represents final prediction. Bagging aims at improving the accuracy by creating an improved composite classifier, M* by amalgamating the various outputs of learned classifier into a single prediction. Given a set, D, of d tuples, bagging works as follows. For iteration *i (i=1,2,….,k),* a training set, Di of d tuples is sampled with replacement from the original set of tuples D.

5. EVALUATİON METRİCS

The performance measures are calculated from:

True Positives (TP), the number of cancerous correctly classified as cancerous

a) True Negatives (TN), the number of non-cancerous programs correctly classified as non-cancerous

b) False positives (FP), the number of non-cancerous programs falsely classified as cancerous
c) False Negative (FN), the number of cancerous falsely classified as non-cancerous

The models were evaluated based on four criteria described below:

a) Accuracy*:* it reflects the number of correctly classified predictions from all predictions made, which is the ratio of correct predictions to the total predictions given as

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

b) Precision: it measures the proportion of the patient without breast cancer disease, which is the ratio of true positives to the overall positive predictions. It is otherwise referred to as precision or positive predictive value given as:

$$Precision = \frac{TP}{TP+FP} \quad (15)$$

c) False Alarm Rate (FAR): the proportion of actual negative breast cancer cases that were predicted as positive cases by the model. This is simply the ratio of false positives (false alarms) to the total negatives

$$FAR = \frac{FP}{TN+FP} \quad (16)$$

d) Recall*:* it measures the proportion of the patient with breast cancer disease, also is the ratio of True positives to the total positives i.e.

$$Recall = \frac{TP}{FN+TP} \quad (17)$$

6. EXPERİMENTAL SETUP AND RESULTS DİSCUSSİON

Three different predictive models were used on breast cancer dataset which consist of 700 instances, the training set consist of 70% of each instances of both class which was used to build the models and the remaining 30% instances of both class was used in determining the efficacy of the classification of individual classifiers as shown in first is the base learners- NB, DT, and SVM, while the second one was the bagging ensemble approach based on majority voting. The number of records in the training set is 490 (70%) and 210 (30%) in testing. Breast cancer dataset consist of 9 attributes which were reduces to 7 using WEKA ranking gain ratio feature selector method. İt was used to calculate and compare the statistical measure of the performance of a binary classification test. Table 3 shows the Breast cancer features using Gain Ratio feature selector.

Table 3: Breast cancer features ranking using Gain Ratio feature selector

Feature	Features	Gain Ratio
8	Nucleoli	0.399
5	Eplithelia	0.395
2	Cell Size	0.386
6	Nuclei	0.374
3	Cell Shape	0.314
7	Chromatin	0.303
9	Mitosis	0.299
4	Adhesion	0.271
1	Clump Thickness	0.21

The confusion matrices table of Naïve Bayes, SVM, decision tree, random forest, ensemble classifiers and the detailed performance of each model with all the 9 attributes is presented in Table 4, 5, and 6 respectively. Also, figure 1 show the performance evaluation chart with all the 9attributes used in determining the performances of the classifiers.

Table 4: Confusion Matrix Table of Naïve Bayes, SVM and Decision tree models performance with all the 9 attributes

Classifiers	Naïve Bayes	SVM	Decision tree

	Predicted Positive	Predicted negative	Predicted Positive	Predicted negative	Predicted Positive	Predicted negative
Active positive	TP = 128	FN = 6	TP = 127	FN = 7	TP = 128	FN = 7
Actual negative	FP = 6	TN = 70	FP = 6	TN = 69	FP = 6	TN 69

Table 5:

Classifiers	**Random Forest**		**Ensemble Bagging**	
	Predicted Positive	Predicted negative	Predicted Positive	Predicted negative
Active positive	TP = 129	FN = 7	TP = 128	FN = 5
Actual negative	FP = 5	TN = 69	FP = 5	TN = 72

Confusion Matrix Table of Random Forest, Ensemble Bagging models performance with all the 9 attributes

Table 6: Results of Individual Classifiers with all the 9 attributes

Classifier	**Naïve Bayes**	**Decision tree**	**SVM**	**Random Forest**	**Ensemble Bagging**
Accuracy	0.94	0.94	0.93	0.95	0.95
Precision	0.96	0.95	0.96	0.94	0.95
False Alarm Rate	0.08	0.08	0.08	0.09	0.04
Recall	0.96	0.95	0.94	0.98	0.96

The confusion matrixes table of Naïve Bayes, SVM, decision tree, random forest ensemble classifiers and the detailed performance of each model with all the 7 attributes is presented in Table 7, 8, and 9 respectively. Also, figure 2 shows the performance evaluation chart with 7 attributes used in determining the performances of the classifiers.

Table 7: Confusion Matrix Table of Naïve Bayes, SVM and Decision tree models performance with all the 7 attributes Gain Ratio features Selector

Classifiers	**Naïve Bayes**		**SVM**		**Decision tree**	
	Predicted Positive	**Predicted negative**	**Predicted Positive**	**Predicted negative**	**Predicted Positive**	**Predicted negative**
Active positive	TP = 127	FN = 7	TP = 128	FN = 6	TP = 127	FN = 7
Actual negative	FP = 6	TN = 70	FP = 5	TN = 71	FP = 6	TN = 70

Table 8: Confusion Matrix Table of Random Forest, Ensemble Bagging models performance with all the 7 attributes

Classifiers	**Random Forest (RF)**		**Ensemble Bagging**	
	Predicted Positive	**Predicted negative**	**Predicted Positive**	**Predicted negative**
Active positive	TP = 128	FN = 3	TP = 129	FN = 7

Actual negative	FP = 7	TN = 72	FP = 3	TN = 71

Table 9: Results of Individual Classifiers with all the 7 attributes

Classifier	Naïve Bayes	Decision tree	SVM	Random Forest	Ensemble Bagging
Accuracy	0.94	0.94	0.95	0.96	0.98
Precision	0.96	0.96	0.96	0.96	0.97
False Rate	0.08	0.07	0.07	0.07	0.04
Recall	0.96	0.95	0.96	0.95	0.96

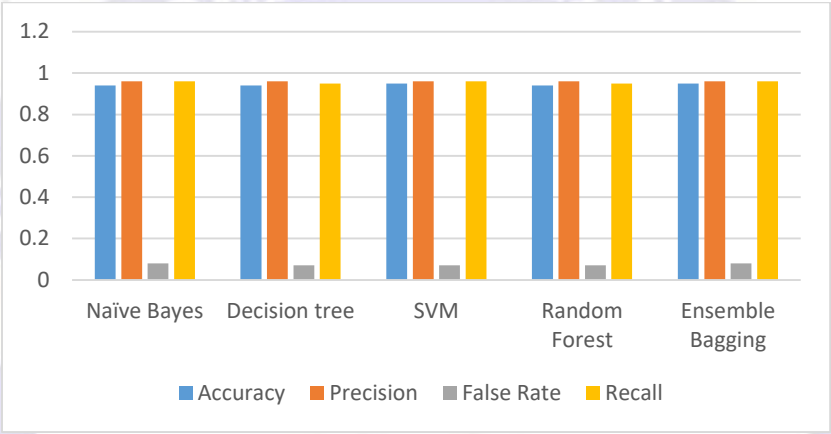

Figure 1: Performance Evaluation Chart with 9 attributes from Breast cancer dataset

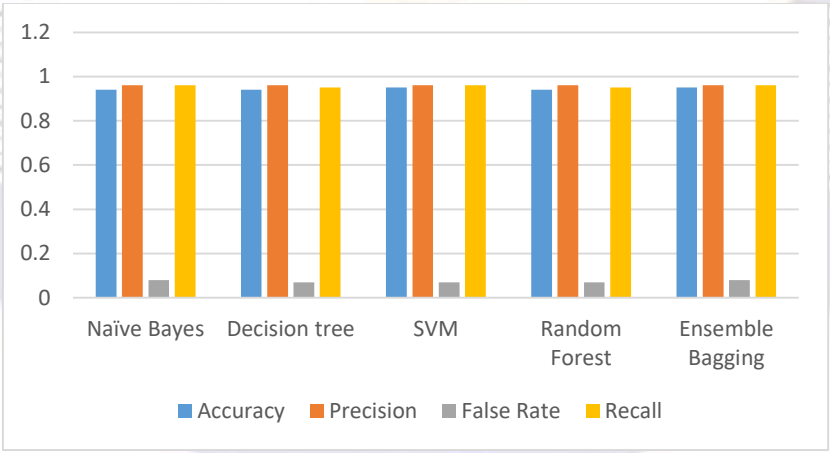

Figure 2: Performance Evaluation Chart with 7 selected attributes from the Breast cancer dataset

The performance of individual base learner classifiers, bagging and random forest for all the attributes were shown in figure 1. From table 6, the results were good, the accuracy ranges from 93 to 95%, precision from 94 to 97% and recall rate from 94 to 98%, while the false alarm rate ranges from 0.07 to 0.08. Also, figure 2 shows the performance of individual base learner classifiers, bagging and random forest for 7 selected attributes. The result reported from table 9 were good - for 7 attributes, the accuracy ranges from 94 to 95%, precision return 96%for all the classifiers, recall rate from 95 to 96%, and the false alarm rate ranges from 0.07 to 0.08. From the generated result above, comparing random forest and bagging ensemble it can be deduced

that the breast cancer disease can still be detected in patients with a few number of features. Hence, the last two attributes shown in Table 4 are of little significance to the detection of breast cancer.

Breast cancer is a deadly disease killing millions of women worldwide, this is quiet worrisome and it has been identified that breast cancer deaths are as a result of many factors which includes lack of diagnosis, poor diagnosis, wrong interpretation of MRI scan results, self-medication, shortage of medical hospitals and facilities and above all shortage of medical experts. These have led to the recent increase in the use of artificial intelligence in medicine since BC is a big challenge mostly for women in the rural areas where there are no medical experts.
This paper provide various data mining techniques used in building an effective diagnostic model for breast cancer detection as stated in the pre-existing literatures. Since human lives are directly involved, accuracy of such model must be considered. The use of ensemble learning approach and forest was used to aggregate the output of various data mining models (Naïve Bayes, SVM, decision tree) was demonstrated in this research over breast cancer dataset as one of the measure by which accuracy of these computer aided diagnostic model could be improved. The proposed method was implemented using a python programing language. The test results show that all the techniques show roughly similar results on four different metrics, it can be deduced that breast cancer disease can still be detected in patients with a few number of features.

Future Direction

Developing nations are often faced with the challenges of gross shortage of man-power, hence, the need for an automated diagnosis system. We intend to validate the developed model on locally sourced breast cancer data in order to evaluate the generality of the classification.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] World Cancer report (2014). World Health Organization

[2] Amin, S.M., Ewunonu, H.A., Oguntebi, E., and Liman I.M., (2017). Breast cancer mortality in a resource-poor country: a 10 year experience in a tertiary institution. *Sahel Medical Journal*, 20 (3): 93-97.

[3] Abdallah1, Y.M., Elgak1, S., Zain, H., Rafiq, M., Ebaid, E.A., and Elnaema, A.A., (2018). Breast cancer detection using image enhancement and segmentation algorithms. *Biomedical Research*, 29 (20): 3732-3736.

[4] Adegoke, V. Chen, D., Banissi, E., and Barikzal, S., (2017). Prediction of Breast Cancer Survivability Using Ensemble Algorithm. *International conference on smart system and technologies,* pp. 223-231. Osijek, Croatia 18 – 20 Oct, 2017.

[5] Hosni M., Ibtissam, A., Ali Idris, Juan .M. Carrillo de Gea, and Jose L. F. A. (2019). Reviewing Ensemble Classification Methods in Breast Cancer. *Computer Methods and Programs in Biomedicine, 177: 89-112.*

[6] Onan, A. (2015): On The Performance of Ensemble Learning for Automated Diagnosis Of Breast Cancer. *Artificial intelligence perspectives and applications. AISC*, 347: 119-129.

[7] Hsieh S.L., Hsieh S.H., Cheng P.H., Chen C.H., Hsu K.P., Lee I.S., Wang, Z., and Lai, F., (2012): Design Ensemble machine learning model for Breast Cancer diagnosis. *Journal of medical systems.* 36(5): 2841-2847.

[8] Joshi, M., and Joshi, A., (2017). On Comparative Study of Breast Cancer classification using Ensemble in Statistics Modelling. *International Journal of Computer Technology and Statistics. IJCTS* 8(1): 18-21.

[9] Dhahri, H., Al Maghayreh, E., Mahmood, A., Elkilani, W., and Nagi, M.F. (2019). Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms. *Journal of Healthcare Engineering,* (2019):1-11.

[12] Abdel-Ilah L., Šahinbegović H. (2017) Using machine learning tool in classification of breast cancer. In: Badnjevic A. (eds) CMBEBIH 2017. IFMBE Proceedings, vol 62. Springer, Singapore. https://doi.org/10.1007/978-981-10-4166-2_1.

[13] Abien Fred Agarap (2018): On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. : ICMLSC '18: Proceedings of the 2nd International Conference on Machine Learning and Soft Computing. Pages 5–9https://doi.org/10.1145/3184066.3184080

[14] UCI machine learning repository: Breast cancer Disease Dataset. Retrieved from http://archive.ics.uci.edu/ml/datasets/Breast_Cancer_disease

[15] Wahba G. (1990). Splines Models for Observational Data, *Series in Applied Mathematics, Vol. 59, SIAM.*

[16] Christopher, K., Darren, M., William, R., and Fredik, V. (2003). Bayesian Event classification for intrusion detection, *Proceedings of the 19th Annual Computer Security Applications Conference (ACSAC'03), 2003.*

[17] Amor, N.B., Beferhat, S. and Elouedi, Z.(2004): Naïve Bayes vs Decision Trees in Intrusion Detection Systems, *ACM Symposium on Applied Computing, pp. 420 – 424.*

[18] Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2): 123-140.

[19] Abdel-Ilah L., and Šahinbegović H. (2017) Using machine learning tool in classification of breast cancer. In: Badnjevic A. (Ed.) CMBEBIH 2017. IFMBE Proceedings, vol 62:3-8.