International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

An Evaluation of the Wisconsin Breast Cancer Dataset using Ensemble Classifiers and RFE Feature Selection Technique

Sulyman Age Abdulkareem[a]*, Zainab Olorunbukademi Abdulkareem[b]

[a]*Institute for Communication Systems, Home of 5G and 6G Innovation Centre, University of Surrey, Guildford, GU2 7XH, UK*

[b]*Computer and Information Sciences Department, University of Strathclyde, Glasgow, G1 1XQ, UK*

[a]*Email: s.abdulkareem@surrey.ac.uk*

[b]*Email: zainab.abdulkareem@strath.ac.uk*

Abstract

Breast cancer represents one of the deadliest diseases that records a high number of death rate annually. It is the most common type of cancer and the main cause of death among women worldwide. Machine learning (ML) approach is an effective way to classify data, especially in medical field. It is widely used for classification and analysis to make decisions. In this paper, a performance comparison between two ensemble ML classifiers: Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) on the Wisconsin Breast Cancer Dataset (WBCD) is conducted. The main objective of this study is to assess the correctness of the classifiers with respect to their efficiency and effectiveness in classifying the dataset. This was done by utilizing all and reduced features of the dataset that were generated with Recursive Feature Elimination (RFE) feature selection technique. Four metrics were used in the study: Accuracy, Precision, Recall and F1-Score to evaluate the classifiers. All experiments were executed within Anaconda Environment with Jupyter Notebook and conducted using Python programming language. Experimental result shows that XGBoost with 5 reduced feature using RFE feature selection technique gives the highest accuracy (99.02%) with lowest error rate.

Keywords: Breast Cancer; WBCD; XGBoost; RF; RFE; Ensemble Classifiers.

---

* Corresponding author.

1. Introduction

cancer is a large group of diseases that can start in almost any organ or tissue of the body when abnormal cells grow uncontrollably [1]. The disease recorded an estimated 9.6 million deaths in year 2018. It is a heterogeneous disease that can be divided into several distinct types. Breast cancer, one of the types of cancer, is second most recurrent cancer peculiar to women after lung cancer [2]. It is also one of the prime causes of death among them globally. Yearly, approximately 124 of 100,000 women are diagnosed with breast cancer, with an estimate of 23 of 124 of them dying from the disease [3]. The World Health Organization (WHO) also reported that 25% of the women in the USA are diagnosed with this type of cancer at some stage in their lives. Additionally, it was noted that in Nigeria, over 115,000 active cancer cases were recorded in 2018, with over 70,000 deaths from the disease. Furthermore, 22.7% of the total instances were breast cancer cases with 16.4% deaths recorded. The breast cancer trend chart was able to demonstrate that by year 2040, the recorded cases would have increased from 26,310 in 2018 to about 50,921 active cases [1]. Figure 1 depict an image of a breast cancer tumor sizes, T1 being the smallest (2cm or less), and T4 being the biggest (any size of tumor growing into the chest wall).

Table 1: List of Acronyms

Acronym	Full Form
ACC.	Accuracy
AI	Artificial Intelligence
ANN	Artificial Neural Network
BN	BayesNet
CART	Classification and Regression
DT	Decision Tree
FN	False Negative
FP	False Positive
FNA	Fine Needle Aspiration
FST	Feature Selection Technique
IBk	Instance Based Learner
ICBC	Iranian Center for Breast
ID3	Iterative Dichotomiser 3
kNN	k-Nearest Neighbour
LDA	Linear Discriminant Analysis
LP	Linear Perceptron
LR	Logistic Regression
ML	Machine Learning
MLP	Multi-Layer Perceptron
NB	Naïve Bayes
OOB	Out of Bag
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
QDA	Quadratic Discriminant
RF	Random Forest
RFE	Recursive Feature Elimination
SVM	Support Vector Machine
SMO	Sequential Minimal
TN	True Negative
TP	True Positive
USA	United States of America
XGBoost	eXtreme Gradient Boosting
WBCD	Wisconsin Breast Cancer
WHO	World Health Organization

Consequently, detecting cancer at an early stage gives a 30% chance of it being treated effectively, while late detection makes its treatment complex [4,5]. In detecting breast cancer at an early stage, some techniques are employed by medical practitioners such as: surgical biopsy which has approximately 100% correctness [4], Fine

Needle Aspiration (FNA) with visual interpretation with 65% to 98% correctness [6], and mammography with 63% to 97% correctness [7]. However, the first technique is the most reliable but comes with a costly price tag. In this paper, ensemble Machine Learning (ML) classifiers will be used to evaluate the Wisconsin Breast Cancer Diagnosis (WBCD) dataset which is based on the FNA technique.

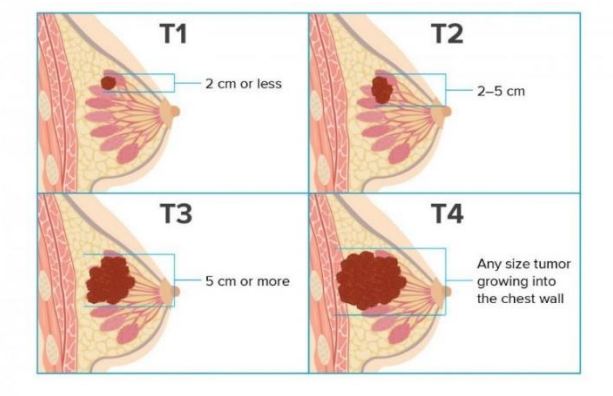

Figure 1: Breast Cancer Tumor Sizes

Machine Learning is a subset in the Artificial Intelligence (AI) research domain which allows machines learn a particular task through training from input dataset to acquire experience [8]. The use of the ML approach has been dominant in the last few decades in the development of predictive models that aids effective decision-making in various domains. One of such is the cancer research domain, where the approach can be used to identify distinct patterns in dataset, and subsequently make a prediction. Numerous research studies have been published over the last two decades that tried to achieve the best performance for the computational interpretation FNA samples [9]. In this study, two ensemble ML classifier: eXtreme Gradient Boosting (XGBoost) and Random Forest (RF) classifiers are used to test the WBCD dataset. The rest of the paper is organized as follows: Section 2 discusses some related work. Section 3 discusses the cancer dataset that will be used for the experiment. The experiment methodology, and the fundamental concept of the two ensemble ML classifiers being investigated are discussed in Section 4. Experiment results and discussions are provided in Section 5. Finally, Section 6 and 7 contains some recommendations and conclusion for this work.

2. Related Work

The research using machine learning classifiers in the medical domain has been prevalent for a long time, especially in the diagnosis of breast cancer. Classification task is one of the popular types of machine learning tasks. Several research studies have been conducted by applying ML classifiers on different medical data, one of which is the WBCD. Results from previous studies have demonstrated that the used classifiers produced good classification accuracies. A study titled "Breast Cancer Diagnosis on Three Different Datasets Using Multi-Classifiers" was conducted by Salama and his colleagues [10] using WEKA data mining tool. The study utilized three distinct breast cancer datasets with WBCD being one of them. Binary classification task was performed on the datasets using five different ML (NB, MLP, J48, SMO, and IBK) classifiers. Experiment results from the WBCD was able to demonstrate that the SMO classifier had the best classification accuracy with (96.9957%),

while the IBK classifier has (94.5637%) accuracy. Furthermore, a fusion of J48 and MLP with PCA was able to produce a higher level of accuracy with (97.568%), with the least accuracy result being (95.8512%) for J48 and NB with PCA. A study by [11] to analyze feature selection with classification on three different breast cancer datasets. The study was based on binary classification, and it utilized the CART classifier. Experimental result was able to demonstrate that the classifier had the best classification for the WBCD dataset with 94.84% accuracy without feature reduction. Different feature selection techniques were applied on the dataset and highest level of accuracy was gotten from the combination of the CART classifier + the *Exhaustive* feature selection technique with 6 features and 95.13% accuracy. Reference [12] compared the performance criterion of some ML classifiers (*RepTree*, *RBF Network*, and *Simple Logistic*). The dataset used in the study was provided by University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia. The data set has 10 attributes and total 286 rows. Experiment result was able to demonstrate that the Simple Logistic classifier had the best accuracy 74.5% for the binary classification task. Asri and his colleagues [13] used some machine learning algorithms for breast cancer risk prediction and diagnosis. In the study, a performance comparison between different ML algorithms: SVM, C4.5, NB and kNN on the WBCD is conducted. The study had the objective of assessing the correctness of the selected algorithms in classifying the cancer data. Experimental results demonstrated that SVM gave the highest accuracy (97.13%) with lowest error rate. All experiments were conducted with WEKA data mining tool. Reference [14] conducted a study to compare the performance of ML algorithms for breast cancer detection and diagnosis. In the paper, three ML algorithms were used; SVM, RF, and BN. The WBCD dataset was used in the experiment to evaluate the performance of the algorithms. The paper used the WEKA data mining tool for simulation, and it reported 97% accuracy. In addition, classification performance varies based on the method that is selected. The SVM had the highest performance in terms of accuracy, specificity, and precision. However, RFs had the highest probability of correctly classifying the breast cancer data. Ivančáková and his colleagues [15] compared different ML methods on Wisconsin Dataset. The dataset that was used had 569 distinct records and 32 attributes. In the study six ML classifiers were utilized for the experiment simulation. The dataset was divided into ratio (60:40, 70:30, and 80:20) for evaluation purpose. In addition, the performance of the classifiers was evaluated with the actual and an under-sampled version of the breast cancer dataset. The results from the experiment was able to demonstrate that the SVM classifier had the best accuracy result for both versions of the dataset ratio 70:30 and 97.66% for the actual dataset. The under-sampled dataset with ratio 80:20 and 97.91% accuracy. Shajahaan and his colleagues [16] applied data mining techniques to model breast cancer data. The experiment was conducted using some ML classifiers; RF, CART, C4.5, ID3, and NB on the WBCD dataset. In addition, TANGARA and WEKA data mining tool were used for the data simulation. Result from the experiment was able to demonstrate that the NB classifier had 97.42% accuracy, making it the best classifier in the study. The study by Amrane and his colleagues [17] was based on breast cancer classification using two (kNN and NB) machine learning classifiers. The classification performance of the classifiers was evaluated using the WBCD data. Results from the experiment simulation was able to demonstrate that the kNN was the better classifier in terms of accuracy with 97.51% against 96.19% for NB. However, NB had a shorter total execution in comparison to the kNN classifier. Bayrak and his colleagues [18] compared the performance of two (ANN and SVM) ML methods for breast cancer diagnosis in their study. The WBCD data was used in the experiment to evaluate the methods with WEKA data mining tool. Based on the performance metrics of the applied methods, SVM (Sequential Minimal Optimization Algorithm) had the

best performance with 96.9957% accuracy for the diagnosis and prediction of the dataset. Ahmad and his colleagues [19] conducted their study using three (*C4.5*, *ANN*, and *SVM*) ML techniques for predicting breast cancer recurrence. The dataset used in the study was obtained from the Iranian Center for Breast Cancer (ICBC). It had 1189 records, 22 predictor variables, and one outcome variable. The pre-processing stage reduced the records to 547, as 642 records had missing data. Experiment result demonstrated that the SVM classifier had the better accuracy with 95.7% in comparison to C4.5 and ANN with 93.6% and 94.7% respectively. All experiment simulations were conducted using WEKA data mining tool. Showrov and his colleagues [20] used some machine learning algorithms for breast cancer risk prediction and diagnosis. In the study, a performance comparison between different ML algorithms: SVM, ANN, and NB on the WBCD is conducted. The study had the objective of assessing the correctness of the selected algorithms in classifying the cancer data. Experimental results demonstrated that SVM gave the highest accuracy (96.72%) with lowest error rate. All experiments were conducted using Python programming language. A study titled "Breast classification using machine learning" was conducted by Amrane and his colleagues [21] using WBCD dataset. Binary classification task was performed on the datasets using two ML (NB, and kNN) classifiers. Experiment results was able to demonstrate that the kNN classifier had the best classification accuracy with (97.51%), while the NB classifier has (96.19%) accuracy. However, the authors noted that even though kNN had the best accuracy performance, its running time will increase if a larger dataset is adopted for the experiment. Sarkar and Nag [22] evaluated the classification performance of C4.5 classifier using the WBCD dataset. Binary classification task was performed using the selected classifier. Experiment result was able to demonstrate that the classifier was able to attain an accuracy of 96.71% for the classification task. All experiments were conducted using Java 7 programming language. The study was concluded with the authors suggesting that C4.5 decision tree based classification systems would result in better diagnosis at an early stage for patients who are potentially at risk of breast cancer; and would in-turn help save thousands of lives each year. A study titled "Breast cancer classification using machine learning techniques: a comparative study" was conducted by Houfani and his colleagues [23] using the WBCD dataset. Binary classification task was performed on the datasets using seven different ML (SVM, RF, DT, MLP, LR, NB, and kNN) classifiers. Experiment results was able to demonstrate that the MLP and LR classifiers had the best classification accuracy with (97.9%), while the NB had (95%) classification accuracy which was the least. The study was concluded with the authors stating that the best classifiers (*MLP* and *LR*) in the study are better performant in comparison to the other classifiers that were utilized. In this paper two ensemble ML classifier are used: Random Forest (RF) and eXtreme Gradient Boost. In addition (*Recursive Feature Elimination* (RFE)) feature selection technique was used to reduce the dataset features to evaluate the classifier performances. The learning process in ML techniques can be divided into two main categories, supervised and unsupervised learning. In supervised learning, which is our focus in this study, a set of data instances are used to train the machine and are labeled to give the correct result (*Classification*). The selected classifiers used in this study are under the supervised category.

3. Dataset

In this paper, we used the Wisconsin Breast Cancer Dataset (WBCD) that is acquired from UCI Machine Learning Repository. It has 699 instances that are classified as benign and malignant. In addition, it has 11 integer-valued attributes [24]. Table 2 depicts the features of the dataset:

Table 2: Summary of the Dataset

S/N	Attribute	Domain
1	Sample code number	id number
2	Clump Thickness	1–10
3	Uniformity of Cell Size	1–10
4	Uniformity of Cell Shape	1–10
5	Marginal Adhesion	1–10
6	Single Epithelial Cell Size	1–10
7	Bare Nuclei	1–10
8	Bland Chromatin	1–10
9	Normal Nucleoli	1–10
10	Mitoses	1–10
11	Class	2 for Benign & 4 for Malignant
Class Distribution		Benign: 458 (65.5%); Malignant: 241 (34.5%)
Total Number of All Instances		699
Number of Missing Values		16

4. Methodology

This section discusses the methodology and experiment setup for the classification task. The steps that we followed are given in details in subsequent subsection.

4.1. The Classifiers

- *Random Forest*

Random Forest classifier, RF for short works combines output of several decision trees to form an ensemble forest of trees. This is backed-up with an argument having a single decision tree can either produce a certain or very simple model [25]. The classifier is sometimes referred to as a *Bagging Classifier*. Adopting the use of the classifier increases the stability in comparison to using single decision tree classifier.

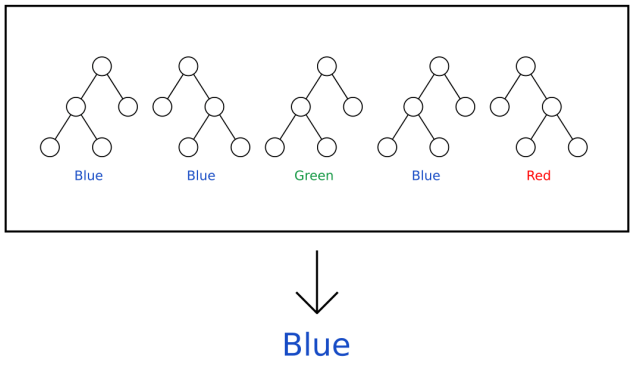

Figure 2: How RF Works

Consequently, this indicates that classifier performance is not affected by the noise of the input dataset. In addition, one of the prime reasons behind using RF in cancer detection is its ability to handle data minorities.

The RF classifier is based on a recursive method in which every iteration involves picking one random sample of size N from the dataset with replacement, and another random sample from the predictors without replacement. The data obtained is partitioned afterwards. The out-of-bag data (OOB) which is all data not chosen in the sampling process are then dropped and the above steps repeated several times depending on the number of selected trees. Lastly, a count is made over the trees used in classifying the output, and subsequently cases are then classified based on a majority vote over the decision trees [26]. Figure 2 depicts the pictorial diagram of how the RF classifiers works.

- *eXtreme Gradient Boost*

eXtreme Gradient Boost classifier, XGBoost for short works by building a succession of weak decision tree learners, with each new tree correcting try to reduce the error of the previous one. It was proposed by Chen and Guestrin [27] and was accorded as a scalable ML classifier. Recent studies by researchers has been able to highlight that some classifiers have more success rate in performing classification tasks in comparison to other classifiers. XGBoost is categorized as one of such classifiers [28]. This classifier is designed to increase the computational speed and efficiency of the machine used in performing the experiment. Figure 3 depicts the pictorial diagram of how the XGBoost classifiers works.

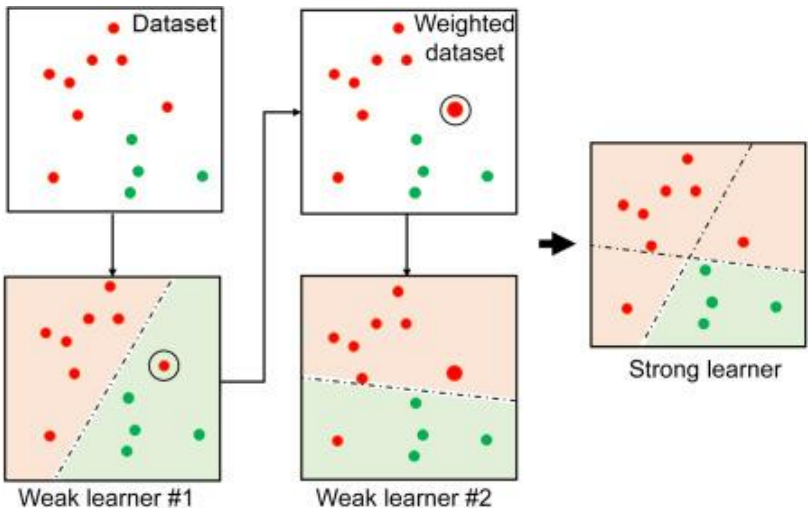

Figure 3: How XGBoost Works

4.2. The Feature Selection Technique

- *Recursive Feature Elimination*

Recursive Feature Elimination, RFE for short is the feature selection technique that we adopted in this study. It works by recursively eliminating dataset attributes and building a classifier on the remaining attributes. It uses the classifier accuracy to pinpoint which attributes (combination of attributes) contribute significantly to predicting the target class.

4.3. Data Pre-processing

The classifiers were developed using Python programming language in Jupyter Notebook IDE environment. This was adopted as python is a powerful programming language that has been used over time to conduct this type of research in some previous studies. There are many inbuilt libraries in python that we used during the pre-processing phase of the dataset. Importation of dataset into the GUI was the first step that was taken to enable us to understand and visualize the dataset. Findings was able to reveal that the 16 of the cancer records had missing data, which had to be dropped. The remaining data (683) were then utilized for the experiment. Furthermore, the (Sample code number) attribute was dropped as it contains the patient's identification number which is not relevant for classification. In addition, the experiment is in two phases, the first phase used all the dataset features. The second phased used a reduced number of features after applying the RFE on the dataset. The application of the feature selection technique pinpointed five features as the best for the classification task. The features are (*Uniformity of Cell Size*, *Uniformity of Cell Shape*, *Bland Chromatin*, *Normal Nucleoli*, and *Mitoses*). In addition, we used the min max normalization to scale the features between 0 and 1 to remove the skewness that is associated with the dataset. It can be described with the following formula [29].

$$\mathrm{y} = \frac{x - \min(x)}{\max(x) - \min(x)} \qquad (1)$$

where x is the initial value and min and max are the highest and lowest value of the attribute in each column respectively.

4.4. Implementation

The choice of the selected classifiers for this study is based on the fact that ensemble ML classifiers are known to have better performance when used for classification task in comparison to single classifiers. Specifically, we focused on the XGBoost classifier as it has recently been dominating applied machine learning and Kaggle competitions for structured or tabular data. In the experiment, we split the dataset into ratio 80/20, where 80 is the training set and 20 is the test set. Thus, the split dataset ratio is used in building and evaluating the two classifiers.

4.5. Classifiers Evaluation

Once we modelled the training data, it was evaluated with the test data to predict the outcome. The main conundrum that is to be solved in this classification task with the ensemble ML classifiers is to predict labels of unseen (*test*) data accurately. This is with the assumption that all samples are drawn from the same probability distribution and are completely independent from each other. The classifiers performances were evaluated using metrics such as Accuracy, Precision, Recall, and F1- Score. The Accuracy metric defines how correct the classifiers used in the experiment performed the classification task. Accuracy measures the proportion of true instances classified by the classifier (TP + TN) against the overall predicted classification instances. The metric is defined as follows:

$$Acc. = \frac{(TP + TN)}{(TP + FP + TN + FN)} \qquad (2)$$

where TP is the True Positive value, and it corresponds to the number of data instances that were classified correctly and were correct, TN is the True Negative value, which corresponds to the number of data instances that were classified as incorrect and were actually incorrect, FP is the False Positive value, and it corresponds to the number of data instances that were classified incorrectly but were actually correct, and FN is the False Negative value, which corresponds to the number of data instances that were classified as correct but were actually incorrect. Precision evaluates the proportion of the data instances predicted as true and were true in the experiment i.e. (the fraction of relevant instances among all retrieved instances). It is defined as follows:

$$Precision = \frac{(TP)}{(TP + FP)} \qquad (3)$$

Recall evaluates the proportion of the actual true data instances that were predicted correctly as true in the experiment i.e. (the fraction of retrieved instances among all relevant instances). It is defined as follows:

$$Recall = \frac{(TP)}{(TP + FN)} \qquad (4)$$

F1-Score evaluation metric was used in measuring the harmonic mean between precision and recall scores of the classifiers. The metric was used in finding a fair balance between the two metric values of the classifiers. It is defined as follows:

$$F1 - Score = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)} \qquad (5)$$

5. Results and Discussion

In this research, 2 ensemble ML classifiers were closely investigated and assessed with the breast cancer dataset using the aforementioned metrics. The confusion matrix for each model is formulated to evaluate the classifier. The result is given in Table 3 and 4.

Table 3: Result Summary with All Features

Metric	RF	XGBoost
Acc.	97.07%	98.53%
Precision	97%	98%
Recall	97%	99%
F1-Score	97%	98%

In the first experiment, the performance of the two classifiers is compared based on the previously listed evaluation metrics. All the dataset features were used in the experiment and with the training and test set being split into ratio 80/20. Experiment results was able to demonstrate that XGBoost performed the classification task better than RF classifier as seen in Table III. XGBoost had the highest classification accuracy with 98.53% against 97.07% for RF, making it the best classifier in the first experiment. The RF classifier had an average of 97.07%, 97%, 97%, and 97% for (*Accuracy, Precision, Recall, and F1 Score*) respectively. A further breakdown

of the results was able to reveal that the classifier had a good average accuracy score of (97.07/100%). The accuracy score was calculated using the proportion of true instances classified by the model (*TP + TN*) against the overall predicted classification instances. The precision score was able to demonstrate that 97% which is a fraction of 0.97/1.00 of relevant instances among all retrieved instances were classified correctly, while the recall score also showed that 97% which is a fraction of 0.97/1.00 of retrieved instances among all relevant instances were correctly classified. Lastly, the F1-Score was used in calculating the harmonic mean of the precision and recall values, such that the best score is (100%), and the classifier was able to record a high score (97%) as the average for the classification task. The second classifier, XGBoost had an average of 98.53%, 98%, 99%, and 98% for (*Accuracy, Precision, Recall, and F1 Score*) respectively. A further breakdown of the results was able to reveal that the classifier had a good average accuracy score of (98.53/100%). The accuracy score was calculated using the proportion of true instances classified by the model (*TP + TN*) against the overall predicted classification instances. The precision score was able to demonstrate that 98% which is a fraction of 0.98/1.00 of relevant instances among all retrieved instances were classified correctly, while the recall score also showed that 99% which is a fraction of 0.99/1.00 of retrieved instances among all relevant instances were correctly classified. Lastly, the F1-Score was used in calculating the harmonic mean of the precision and recall values, such that the best score is (100%), and the classifier was able to record a high score (98%) as the average for the classification task.

Table 4: Result Summary with 5 Features

Metric	**RF**	**XGBoost**
Acc.	98.05%	99.02%
Precision	98%	99%
Recall	98%	99%
F1-Score	98%	99%

The second experiment introduced the RFE feature selection technique to reduce the dimension of all the dataset features to the best 5 features. The experiment followed the same pattern as the first, only that feature selection was done before the min max normalization. Results was able to demonstrate again that XGBoost was the better classifier in comparison to RF as seen in Table 4. XGBoost had the highest classification accuracy with 99.02% against 98.05% for RF, making it the best classifier for the experiment. In addition, Table 5 depicts

Table 5: Confusion Matrix of XGBoost + RFE

Predicted			
Malignant	Benign		
2	129	Benign	***Actual***
74	0	Malignant	

the confusion matrix which revealed that the XGBoost classifier predicted all the malignant data correctly during the experiment. The RF classifier had an average of 98.05%, 98%, 98%, and 98% for (*Accuracy, Precision, Recall, and F1 Score*) respectively. A further breakdown of the results was able to reveal that the classifier had a good average accuracy score of (98.05/100%). The accuracy score was calculated using the proportion of true instances classified by the model (*TP + TN*) against the overall predicted classification

instances. The precision score was able to demonstrate that 98% which is a fraction of 0.98/1.00 of relevant instances among all retrieved instances were classified correctly, while the recall score also showed that 98% which is a fraction of 0.98/1.00 of retrieved instances among all relevant instances were correctly classified. Lastly, the F1-Score was used in calculating the harmonic mean of the precision and recall values, such that the best score is (100%), and the classifier was able to record a high score (98%) as the average for the classification task. The second classifier, XGBoost had an average of 99.02%, 99%, 99%, and 99% for (*Accuracy, Precision, Recall, and F1 Score*) respectively. A further breakdown of the results was able to reveal that the classifier had a good average accuracy score of (99.02/100%). The accuracy score was calculated using the proportion of true instances classified by the model (*TP + TN*) against the overall predicted classification instances. The precision score was able to demonstrate that 99% which is a fraction of 0.99/1.00 of relevant instances among all retrieved instances were classified correctly, while the recall score also showed that 99% which is a fraction of 0.99/1.00 of retrieved instances among all relevant instances were correctly classified. Lastly, the F1-Score was used in calculating the harmonic mean of the precision and recall values, such that the best score is (100%), and the classifier was able to record a very high score (99%) as the average for the classification task.

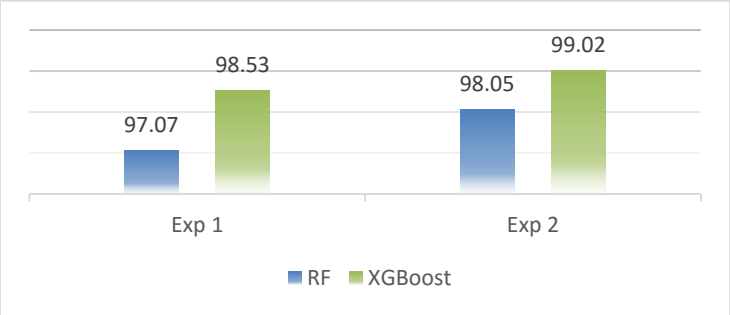

Figure 4: Classifiers Performance Based on Accuracy

Table 6: Papers using the WBCD

Ref	**Classifiers**	**Accuracy**
Lavanya and Rani [11]	CART + Exhaustive (FST)	95.13%
Chaurasia and Pal [12]	Simple Logistic	74.5%
Asri and his colleagues [13]	SVM	97.13
Bazazeh and Shubair [14]	SVM	97%
Shajahaan and his colleagues [16]	NB	97.42
Ahmad and his colleagues [19]	SVM	95.7%
Showrov and his colleagues [20]	SVM	96.72%
Amrane and his colleagues [21]	kNN	97.51%
Sarkar and Nag [22]	C4.5	96.71%
Houfani and his colleagues [23]	MLP & LR	97.9%
Sumbaly and his colleagues [30]	J48	94.36%
Bennett and Blue [31]	SVM	97.20%
Chaurasia and his colleagues [32]	NB	97.36%
Islam and his colleagues [33]	ANN	98.57%
Our Study	**XGBoost + RFE**	**99.02%**

Table VI presents some previous studies and their accuracy results in comparison to our own findings. Our result was able to demonstrate that the XGBoost + RFE performed better results from previous studies. This shows that the combination of the ML classifier and feature selection technique is more efficient and effective for performing the breast cancer classification task.

6. Recommendations

Datasets are inevitable for training supervised ML models like classification classifiers, in addition to being helpful for evaluating supervised and unsupervised ML models. Consequently, new (*more recent*) labeled breast cancer datasets should be made available publicly and evaluated with different ML classifiers to examine their classification performances. In addition, more instances and distinct features should be added to subsequent datasets to make them more robust, as the current WBCD dataset has limited instances and features. Finally, creating a predefined (*training* and *test*) split for future datasets will facilitate comparisons of different approaches evaluated of the dame dataset.

7. Conclusion

In analyzing medical data, different machine learning approaches are available. A key challenge in this research domain is developing accurate and efficient classifiers for medical applications. In this paper we investigated the use of two (XGBoost and RF) ensemble machine learning classifiers for cancer diagnosis on the Wisconsin Breast Cancer Dataset (WBCD). We compared the classifiers efficiency and effectiveness with all and reduced dataset features in terms of Accuracy, Precision, Recall and F1-Score to find the best classification accuracy. XGBoost that utilized the 5 best features of the dataset had an accuracy of 99.02% and outperformed all other classifiers. Consequently, using the RFE features selection technique was able to increase the classification performance of the classifier in comparison to when all the features were applied. This study is however limited to just two ensemble ML classifiers and one feature selection technique. It can be extended by applying other ML classifiers and more distinct feature selection techniques on the dataset. This will allow for comparison of different techniques that are adopted by other researchers in their different studies.

References

[1]. "WHO - Breast Cancer: Prevention and Control," https://www.who.int/ health-topics/cancer, 2020, Accessed December 3, 2020, from WHO - World Health Organization.

[2]. U. C. S. W. Group et al., "United states cancer statistics: 1999–2011 incidence and mortality web-based report," Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute, 2014.

[3]. "NCI. SEER: Cancer Statistics Review," 2012.

[4]. L. R. Borges, "Analysis of the wisconsin breast cancer dataset and machine learning for breast cancer detection," Group, vol. 1, no. 369, 1989.

[5]. J. G. Elmore, C. Y. Nakano, T. D. Koepsell, L. M. Desnick, C. J. D'orsi, and D. F. Ransohoff, "International variation in screening mammography interpretations in community-based programs,"

Journal of the National Cancer Institute, vol. 95, no. 18, pp. 1384–1393, 2003.

[6]. R. W. Giard and J. Hermans, “The value of aspiration cytologic examination of the breast a statistical review of the medical literature,” Cancer, vol. 69, no. 8, pp. 2104–2110, 1992.

[7]. J. G. Elmore, K. Armstrong, C. D. Lehman, and S. W. Fletcher, “Screening for breast cancer,” Jama, vol. 293, no. 10, pp. 1245–1256, 2005.

[8]. D. Michie, D. J. Spiegelhalter, C. Taylor et al., “Machine learning,” Neural and Statistical Classification, vol. 13, no. 1994, pp. 1–298, 1994.

[9]. S. Saxena and K. Burse, “A survey on neural network techniques for classification of breast cancer data,” International Journal of Engineering and Advanced Technology, vol. 2, no. 1, pp. 234–237, 2012.

[10]. G. I. Salama, M. Abdelhalim, and M. A.-e. Zeid, “Breast cancer diagnosis on three different datasets using multi-classifiers,” Breast Cancer (WDBC), vol. 32, no. 569, p. 2, 2012.

[11]. D. Lavanya and K. U. Rani, “Analysis of feature selection with classification: Breast cancer datasets,” Indian Journal of Computer Science and Engineering (IJCSE), vol. 2, no. 5, pp. 756–763, 2011.

[12]. V. Chaurasia and S. Pal, “Data mining techniques: to predict and resolve breast cancer survivability,” International Journal of Computer Science and Mobile Computing IJCSMC, vol. 3, no. 1, pp. 10–22, 2014.

[13]. H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, “Using machine learning algorithms for breast cancer risk prediction and diagnosis,” Procedia Computer Science, vol. 83, pp. 1064–1069, 2016.

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

[15]. J. Ivanc˘akov ´ a, F. Babi ´ c, and P. Butka, “Comparison of different machine ˇ learning methods on wisconsin dataset,” in 2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI). IEEE, 2018, pp. 000 173–000 178.

[16]. S. S. Shajahaan, S. Shanthi, and V. ManoChitra, “Application of data mining techniques to model breast cancer data,” International Journal of Emerging Technology and Advanced Engineering, vol. 3, no. 11, pp. 362–369, 2013.

[17]. M. Amrane, S. Oukid, I. Gagaoua, and T. Ensar˙I, “Breast cancer classification using machine learning,” in 2018 Electric Electronics, Computer Science, Biomedical Engineerings’ Meeting (EBBT). IEEE, 2018, pp. 1–4.

[18]. 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). IEEE, 2019, pp. 1–3.

[19]. L. G. Ahmad, A. Eshlaghy, A. Poorebrahimi, M. Ebrahimi, A. Razavi et al., “Using three machine learning techniques for predicting breast cancer recurrence,” J Health Med Inform, vol. 4, no. 124, p. 3, 2013.

[20]. M. I. H. Showrov, M. T. Islam, M. D. Hossain and M. S. Ahmed, "Performance Comparison of Three Classifiers for the Classification of Breast Cancer Dataset," 2019 4th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 2019, pp. 1-5,

doi: 10.1109/EICT48899.2019.9068816.

[21]. M. Amrane, S. Oukid, I. Gagaoua and T. Ensarİ, "Breast cancer classification using machine learning," 2018 Electric Electronics, Computer Science, Biomedical Engineering's' Meeting (EBBT), Istanbul, 2018, pp. 1-4, doi: 10.1109/EBBT.2018.8391453.

[22]. S. K. Sarkar and A. Nag, "Identifying patients at risk of breast cancer through decision trees," International Journal of Advanced Research in Computer Science, vol. 8, no. 8, pp. 88–91, 2017.

[23]. [23] D. Houfani, S. Slatnia, O. Kazar, N. Zerhouni, H. Saouli, and I. Remadna, "Breast cancer classification using machine learning techniques: a comparative study", Medical Technologies Journal, vol. 4, no. 2, pp. 535–544.

[24]. "UCI Breast Cancer Wisconsin (Original) Datase," https://archive.ics. uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\%28Original29., 2020, Accessed December 24, 2020.

[25]. I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," Artificial Intelligence in medicine, vol. 23, no. 1, pp. 89–109, 2001.

[26]. Y. Yasui and X. Wang, "Statistical learning from a regression perspective by berk, ra," Biometrics, vol. 65, no. 4, pp. 1309–1310, 2009.

[27]. T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794.

[28]. M. Gumus and M. S. Kiran, "Crude oil price forecasting using xgboost," in 2017 International Conference on Computer Science and Engineering (UBMK). IEEE, 2017, pp. 1100–1103.

[29]. M. Çinar, M. Engin, E. Z. Engin, and Y. Z. Ates¸c¸i, "Early prostate cancer diagnosis by using artificial neural networks and support vector machines," Expert Systems with Applications, vol. 36, no. 3, pp. 6357– 6361, 2009.

[30]. R. Sumbaly, N. Vishnusri, and S. Jeyalatha, "Diagnosis of breast cancer using decision tree data mining technique," International Journal of Computer Applications, vol. 98, no. 10, 2014.

[31]. K. P. Bennett and J. Blue, "A support vector machine approach to decision trees," in 1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98CH36227), vol. 3. IEEE, 1998, pp. 2396–2401.

[32]. V. Chaurasia, S. Pal, and B. Tiwari, "Prediction of benign and malignant breast cancer using data mining techniques," Journal of Algorithms & Computational Technology, vol. 12, no. 2, pp. 119–126, 2018.

[33]. M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, and M. N. Kabir, "Breast cancer prediction: a comparative study using machine learning techniques," SN Computer Science, vol. 1, no. 5, pp. 1–14, 2020.