Breast Cancer Detection using Machine Learning Techniques

Md. Samiul Islam
Lecturer
Department of Computer Science and Engineering
Stamford University Bangladesh

Md. Ashikuzzaman
BSc. Final year thesis student
Department of Computer Science and Engineering
Stamford University Bangladesh

Joy Mojumdar
BSc. Final year thesis student
Department of Computer Science and Engineering
Stamford University Bangladesh

ABSTRACT
According to the World Health Organization (WHO), in 2020, around 2.3 million women diagnosed with breast cancer, and 685,000 of them died globally. Though, this calculation is terrible to think, there is always a hope for the patients who are able to be diagnosed at the very early stage. Keeping this helpfulness of early diagnosis in mind, there have been proposed a lot of research works in the recent years. And, most of these researches are computer aided. This is the reason, in the recent years, machine learning techniques are getting quite noticed because of their efficiency and reliability. In this paper,6 different machine learning techniques such as Logistic Regression, Decision Tree Classifier, KNN (K-Nearest Neighbors), Random Forest Classifier, SVM (Support Vector Machine), and Gradient Boosting Classifier have been proposed to detect breast cancer.The very popular Breast Cancer Wisconsin (Original) Dataset [1] collected from UCI machine learning repository has been used to apply the proposed machine learning techniques. In this research work, 20% of data has been used for testing and the rest 80% of data has been used for training. Decision Tree Classifier outperformed the other techniques giving the highest accuracy of 96.89%. The results of other techniques were quite competitive.

General Terms
Breast Cancer Detection, Machine Learning Algorithms.

Keywords
Machine Learning, Logistic Regression, Decision Tree Classifier, KNN (K-Nearest Neighbors), Random Forest Classifier, SVM (Support Vector Machine), Gradient Boosting Classifier.

1. INTRODUCTION
Cancer is such a disease in which, some cells of the body grow in an uncontrolled way and gradually it spreads to the other parts of the body. According to the World Health Organization (WHO), in 2020, there were 2.3 million women diagnosed with breast cancer and 685,000 deaths globally. In Bangladesh, the rate is about 22.5 per 100000 females. Breast cancer has been reported as the highest prevalence rate (19.3 per 100,000) among Bangladeshi women between 15 and 44 years of age [2]. As a developing country, this statistic can be a warning for Bangladesh. If much attention is not paid to this area of disease, it will be a great threat to the women of Bangladesh. Toincrease the survival rate, the first and foremost thing is to diagnose breast cancer by classifying tumors. There are two different types of tumors such as malignant and benign tumors. Physicians need a reliable diagnostic procedure to distinguish between these tumors. But generally, it is very difficult to distinguish tumors even by the experts. If it is detected in its early stages, there is a 30% chance that cancer can be treated effectively[3]. Currently, the most used techniques to detect breast cancer in early stages are: mammography (63% to 97% correctness), FNA (Fine Needle Aspiration) with visual interpretation (65% to 98% correctness), and surgical biopsy (approximately100%correctness) [3]. But there are some downsides with all of these processes such as:

- A false-negative mammogram looks normal even though breast cancer is present. So, false-negative mammograms can give women a false sense of security, thinking that they don't have breast cancer when in fact they do.
- Although a fine needle aspiration (FNA) is a simple and quick procedure, The risks of fine needle aspiration include the possibility of cancer cells being trailed into unaffected tissue as the needle is removed. There is another risk that any abnormal cells may be missed and not detected because an FNA biopsy can only sample a small number of cells from a mass or lump.
- In the case of surgical biopsy, there are some side effects such as tenderness, pain, infection and more importantly the shape of the breast may be changed.

Keeping all of these downsides in mind, during the last decades, a tremendous amount of research has been performed related to breast cancer prediction using machine learning techniques. The main motivation behind these research works is the volume of data generating extremely fast in the field of biomedical. This data provides a rich source of information for medical research. Machine learning helpsto extract information and knowledgefrom data on the basis of pastexperiences and detect hard-to-perceive pattern from large and noisy dataset [8]. This is proved by many researchers that machine learning algorithms work better in cancer diagnosis. Getting motivated through the previous researches, the authorsof this paper also attempted to apply some popular machine learning techniques such as Logistic Regression, Decision Tree Classifier, KNN (K-Nearest Neighbors), Random Forest Classifier, SVM (Support Vector Machine), and Gradient Boosting Classifier. After analyzing their performances in terms of different performance metrics such as precision, recall, f1-score and classification accuracy, the authors have achieved some satisfactory results. Decision Tree Classifier outperformed the other techniques giving the

highest accuracy of 96.89%. The results of other techniques were quite competitive. LogisticRegression, KNN, RandomForestClassifier, SVM, GradientBoostingClassifier attained 62.79%, 75.96%, 95.34%, 74.41%, 95.34% of accuracy respectively. The rest of the paper is arranged as follows:

The section (2) will be the literature review on breast cancer detection using machine learning techniques. The section (3) will explain the fundamental concept of the 6 machine learning algorithms being experimented. The section (4) will describe the methodology that the authors have used to improve the performance of the proposed ML techniques. The section (5) will present the comparative studyof the proposed algorithms. And finally, section (6) will conclude the paper.

2. LITERATURE REVIEW

In the recent years, a lot of researches have been conducted in the field of breast cancer. And, a large quantity of these researches has used machine learning techniques for better results, efficiency and reliability.

In [4], Linear Projections and Radviz were used as visualization techniques for data exploration and feature selection. Further, Decision Tree induction algorithms were used to create models that are able to differentiate between Malignant and benign breast tumors from breast mass images. The result shows Classification and regression Trees achieved an accuracy of 96% in predicting breast cancer.

In [5], The authors have proposed six different classification techniques namely Multilayer Perceptron, Decision Tree, Random Forest, Support Vector Machine and Deep neural network for evaluation. DNN classifier had a great performance in accuracy level of 92%, indicating better results in relation to traditional models, Random Forest 50 and 100 presented the best results for the ROC curve metric and their AUC density was 94% for both.

In [6], The authors have worked with three most popular machine learning techniques namely Random Forest, KNN (K-Nearest Neighbors) and Naïve Bayes. In this paper, the very popular Wisconsin Diagnosis Breast Cancer dataset has been used as a training set to compare the performance of the three proposed techniques based on some key parameters such as accuracy and precision. The result shows KNN outperforms the two other techniques giving the highest accuracy value of (95.90%) and precision value of (98.27%).

In [7], The authors have propounded four different machine learning techniques namely ANN (Artificial Neural Network), KNN (K-Nearest Neighbors), Binary SVM (Binary Support Vector Machine) and Decision Tree. They have used Mammographic Mass dataset. The result shows ANN (Artificial Neural Network) attains the highest accuracy value of 84% which outperforms the other three techniques.

3. MACHINE LEARNING ALGORITHMS

Machine learning algorithm is such a method through which an AI system performs its task, basically predicting output from given input data. There are four types of machine learning algorithms:

- Supervised
- Semi-supervised
- Unsupervised
- Reinforcement

In this paper,the authorshave proposed 6 different popular machine learning algorithms:

3.1 Logistic Regression

Logistic Regression falls in the category of Supervised machine learning algorithm. Its main work is to predict the probability of a binary target class. It works very well on categorical data. Logistic Regression uses the sigmoid function eq.1 to convert the independent variable into the expression of probability that ranges between 0 and 1 with respect to the dependent variable.

$$y = \frac{1}{1+e^{-x}} \ldots\ldots\ldots\ldots (1)$$

Here, x => The independent variable that is neededtotransform.

e => Euler's constant (=2.71828).

y => The output.

The graph that is produced by the sigmoid function is an S-shaped curve as shown in fig1.

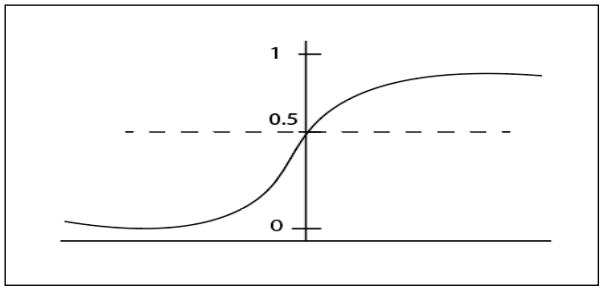

Fig 1: Sigmoid Curve

In this graph, 3 indicators: 0, 0.5, and 1 can be seen. Here, 0 means there is no possibility or no probability of a certain occurrence, 1 means there is a certain possibility, and 0.5 indicates the cut-off line. Data above the cut-off line falls into a class and data under the cut-off line falls into another class. But the data which falls along with the cut-off line (which is a very rare case) are unclassified. There are some fields where logistic regression can be used such as Fraud detection, Disease diagnosis, Emergency detection, etc.

The parameters that were used in this algorithm for this research work are:

- max_iter = 1000
- n_jobs = -1
- random_state = 0

3.2 KNN (K-Nearest Neighbors)

The KNN (K-Nearest Neighbors) is a supervised machine learning algorithm that can be used to solve both classification and regression problems. In this technique, a query data point is given. And, it is needed to calculate the distance between the query data point and all the data points of the dataset as shown in fig2. The distance is calculated usually through Euclidean distance eq.2. After calculating all the distances, it is needed to select the nearest distances based on the value of K. The value ofK is usually an odd number.

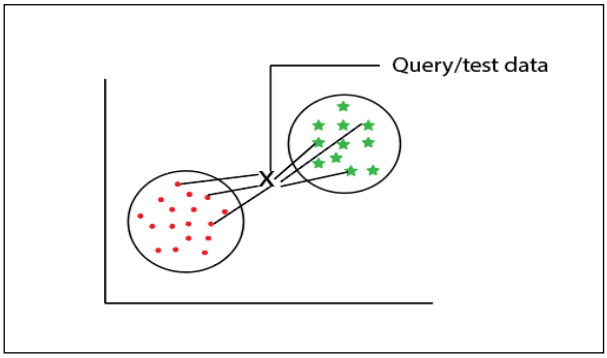

Fig 2: Calculating distance in KNN algorithm

If the value of K = 1, then the datapoint for which smallest distance is found, the query/test data point will fall into the class that this data point (for which smallest distance is found) falls into.

If the value of K = 3, then it is needed to check for the majority of target classes. The test data point will fall into the class of the majority.

In this research work, the authorsfound the K value of 15 to be the best since KNN algorithm gives the best result at this K value in terms of the used dataset.

$$d(p,q) = \sqrt{(q1-p1)^2 + (q2-p2)^2} \ldots\ldots\ldots\ldots (2)$$

The reason of not using the K value as even is the algorithm may be confused to give a decision, if the number of each class is same.

Let's take an example, there two categorical classes A and B. If the K value is 2 and one of the two nearest data points falls into class A and the other falls into B, then thereis no way to check for the majority. This is the main reason of using odd value for K.

The parameters that were used in this algorithm for this research work are:

- n_neighbors = 15
- leaf_size = 10
- n_jobs = -1

3.3 Decision Tree Classifier

Decision Tree is also a supervised machine learning technique that can be used for both classification and regression problems. But, in most of the cases, it is usually preferred for solving classification problems. When the Decision Tree algorithm is used on the training data, it generates a tree-structured model/classifier. When the model is generated, a test data point is provided to the model. Then the model tells what class the data point belongs to.A generated decision tree contains two types of nodes: (1) Decision node and (2) Leaf node as it can be seen in fig.3. Decision node is basically the node which contains some branches and Leaf node does not contain any branches but it provides the classification result. When the model processes the test data point, it gives decision based on the Decision node (Yes/No).And the test process is performed on the feature/attribute of the dataset. Depending on the result of this test process, the whole dataset gets splitted. This process is called 'Splitting Action' which is performed in Decision Tree algorithm.The first Decision node is the Root node of the Decision Tree.

There are some steps through which the Root is selected:

1. First, it is neededto calculate the Entropy value of the total system using eq.3.

   $$E = -\sum_{i=1}^{N} Pi \log_2 Pi \ldots\ldots\ldots\ldots (3)$$

2. It is needed to calculate the Gain values of all the features of the dataset using eq.4.
   $Gain(S,T) = E(S) - E(S,T) \ldots\ldots\ldots\ldots (4)$

3. After calculating the Gain values of all the features, a feature is selected as the decision/root node which has the largest gain value.
4. Steps 1 – 3 are performed iteratively as long as classification result is found.

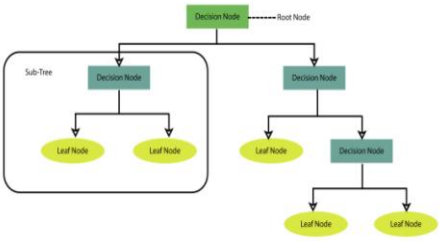

Fig 3: Generated Sample Decision Tree in Decision Tree Algorithm

The parameters that were used in this algorithmfor this research work are:

- max_depth = 500
- max_features = 5
- random_state = 50
- max_leaf_nodes = 500

3.4 Gradient Boosting Classifier

Gradient Boosting Classifiers are a group of machine learning algorithms that combine many weak learning models together to create a strong predictive model. When the Gradient Boosting is used for classification, the initial prediction for every individual is the log(odds). Just like with Logistic Regression, the easiest way to use the log(odds) for classification is to convert it to a probability. And the way to do this conversion is using Logistic function shown in eq.5.

$$logistic\ function = \frac{e * \log(odds)}{1 + e * \log(odds)} \ldots\ldots\ldots\ldots (5)$$

To correctly classify the test data-pointa threshold value is used usually taken as 0.5. The test data-point is classified based on two conditions:

- if the probability of test data > 0.5
- if the probability of test data < 0.5

The next work is to measure how good or bad the initial prediction is. This is measured by calculating Pseudo Residuals.

Residual = (Observed value - predicted value)

The residual values for all the data points in the data set are calculated and stored in a new column called residual.

Thena tree is built using all the independent features of the dataset to predict residuals.

Just like using Gradient boosting for Regression problem, the number of leaves is limited that are allowed in the tree. In practice people often set the maximum number of leaves to be between 8 and 32.

After building the tree, the output values for the leaves are calculated. When the Gradient Boosting is used for Regression problem, a leaf with single Residual has an output value equal to that Residual. But when the Gradient Boosting is used for Classification Problem, the situation is a little more complex because the predictions are in terms of the log(odds). And this leaf is derived from a probability. So, they can't just be added together to get a new log(odds) prediction without some sort of transformation. In terms of classification problem, the most common transformation is done usingeq.6.

$$\frac{\sum Residual}{\sum[Previou\ sProb * (1 - PreviousProb\ )]} \ldots\ldots\ldots\ldots (6)$$

After calculating the output values for all the leaves using the above equation, it is needed to update the predictions by combining the initial leaf with the new tree.

The new tree is scaled by a Learning Rate. The following formula is used to update the predictions:

log(odds) prediction = previous log(odds) prediction + (Learning Rate * output value of a leaf)

Another column is created for storing the new predicted probabilities.

Then, just like before it is needed to calculate the new residuals for each of the datapoints in the dataset. Now that the new Residuals have been found,a new tree can be built and thenthe output values for each leaf needs to be calculated.

The above process repeats untilthe maximum number of trees is made, or the residuals get super smaller.

The parameters that were used in this algorithm for this research work are:

- n_estimators = 20
- learning_rate = 0.075
- max_features = 2
- max_depth = 5
- random_state = 0

3.5 SVM (Support Vector Machine)

SVM is a supervised machine learning algorithm. It can be used for both classification and regression problems. In classification problem, SVM works drawing a Hyperplane between the target classes as it can be seen in fig.4.

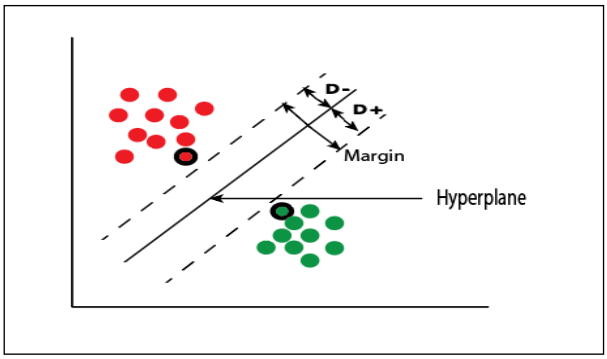

Fig 4: Classification in SVM algorithm

The hyperplane decides in which class the test data point will fall into. The dashed line above the hyperplane is drawn to indicate the nearest data point (Red) to the opponent data points (Green). The dashed line under the hyperplane is drawn to indicate the nearest data point (Green) to the opponent data points (Red). Because of these two dashed lines two distances (D- and D+) are created. D+ is the distance to the nearest positive points and D- is the distance to the nearest negative points. And ifthese two distances are joined, another distance can be found called 'Margin'. This margin has a great significance in deciding the hyperplane. Margin usually plays an important role in deciding which hyperplane exists and which does not exist. The two data points with black stroke based on which the two dashed lines were drawn are called support vectors. The support vectors influence the position and orientation of the hyperplane.

3.6 Random Forest Classifier

Random Forest Classifier is kind of an ensemble classifier which uses Decision Tree algorithm in a randomized way. In the very first step, Random Forest Classifier generates a Bootstrap dataset taking random data points from the original dataset. In the Bootstrap dataset, having duplicate data points is possible. From the bootstrap dataset a decision tree is generated taking a subset of variables at each step of the tree.Following this particular step again and again a large number of decision trees is generated.At the prediction state, when an unknown test data point is given, Random Forest Classifier takes the data point and applies it to all the generated decision trees. Each Decision tree gives a result. Based on these results, the algorithmcounts the majority votes and classifies the test data pointbased on the majority of votes. The following fig.5 explains the voting system.

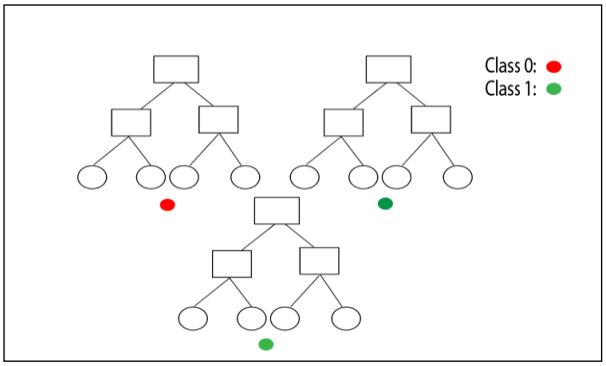

Fig 5: How voting works in Random Forest Classifier

If the voting result is like the scenario shown infig.5 for a particular test data point, then that data point will fallinto the category of Class 1. Because, the Class 1 gets majority of votes.

The parameters that were used in this algorithm for this research work are:

- max_depth = 500
- n_estimators = 100
- max_features = 5
- max_leaf_nodes = 500
- n_jobs = -1
- random_state = 50

4. METHODOLOGY

The authors of this paperconducted thewhole work based on two categories as it can be seen in fig.6 and fig.7. The two categories are:

1. Applying all the machine learning techniques without any feature engineering (applying on noisy data) and analyzing the performances.
2. Applying all the machine learning techniques with feature engineering (applying on cleaned data) and analyzing the performances.

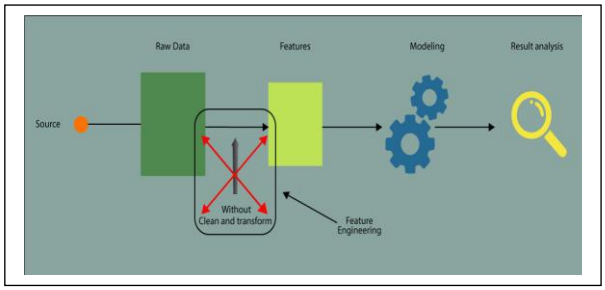

Fig6: Analysis without any feature engineering

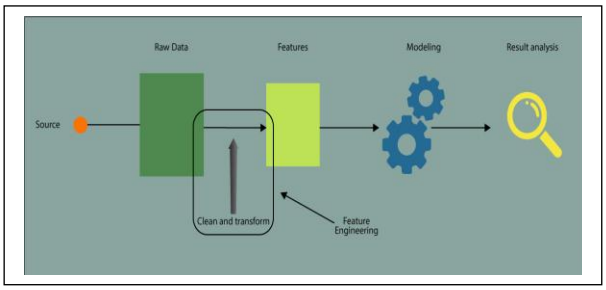

Fig 7: Analysis withfeature engineering

4.1 Dataset

The authors of this paperused the popular Breast Cancer Wisconsin (Original) Dataset collected from thevery popular UCI machine learning repository [1]. The dataset has 699 instances and 10 attributes. The dataset's attribute information is shown in table 1.

Table 1. Dataset's attribute information

Sample code number	id number
Clump Thickness	1 - 10
Uniformity of Cell Size	1 - 10
Uniformity of Cell Shape	1 - 10
Marginal Adhesion	1 - 10
Single Epithelial Cell Size	1 - 10
Bare Nuclei	1 - 10
Bland Chromatin	1 - 10
Normal Nucleoli	1 - 10
Mitoses	1 - 10
Class	(2 for benign, 4 for malignant)

The dataset has 16 missing values, all of which are under the 'Bare Nuclei' attribute. There are 458 benign and 241 malignant records. Fig.8 shows the countplot for the 'Class' attribute.

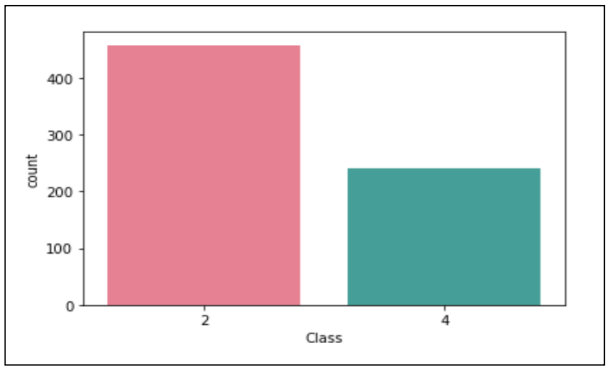

Fig8: Countplot for the Class attribute

4.2 Feature Engineering

In machine learning, feature engineering is basically the pre-processing step. Through this process,important features from raw data can be extracted. This process improves theprediction accuracy of the predictive model. In machine learning, Feature engineering contains four main processes:

1. Feature creation.
2. Transformations.
3. Feature Extraction.
4. Feature Selection.

In this research work,the whole work has been conducted in two ways:

1. Without feature engineering.
2. With feature engineering.

4.2.1 Without Feature Engineering

In this stage, the authorscollected the raw dataset from UCI machine learning repository andapplied machine learning techniques to that raw data. Table 2 shows the performance comparison of the 6 algorithms without feature engineering.

Table 2. Performance analysis of all the algorithms without feature engineering

Algorithms	Accuracy	F1-score	Precision	Recall
Logistic Regression	89.28%	88.98%	88.75%	89.31%
Decision Tree Classifier	**90.71%**	**90.45%**	**90.20%**	**90.79%**
KNN	65.71%	58.61%	66.60%	60.09%
Random Forest Classifier	**92.14%**	**91.92%**	**91.66%**	**92.27%**
SVM	59.28%	37.21%	29.64%	5.0%
Gradient Boosting Classifier	**91.42%**	**91.12%**	**91.12%**	**91.12%**

This analysis shows that Random Forest Classifier gives the highest (92.14% of accuracy) without any feature engineering. In the next step,their performances will be analyzed after feature engineering the data.

4.2.2 With Feature Engineering

In this research work, the following steps were performed as feature engineering:

1. Handling missing values.
2. Plotting boxplot to identify outliers.
3. Removing outliers with 3-standard deviation.
4. Applying PCA (Principal Component Analysis).

Now,the above steps will be described one by one.

4.2.2.1 Handling Missing Values

The real-world dataset often contains a lot of missing values. Many machine algorithms fail to work if the dataset contains missing values. The dataset used in this research work also contains some missing values (16 missing values). So, to make the algorithms classify properly, it was needed to handle these missing values. All of these 16 missing values are under the 'Bare Nuclei' attribute of the dataset. There are several ways of handling missing values. But the two most popular of them are:

- Removing the instances containing missing values.
- Replacing the missing values with mean/median/mode of the particular feature.

In this research work, the second approach has been chosen as the authors didn't want to lose any data from the dataset. The mean value (3.463519313304721) of 'Bare Nuclei' feature was calculated and replacedwith all the missing values.

4.2.2.2 Plotting Boxplot to Identify Outliers

In machine learning, outlier is a datapoint that is markedly differs from the rest of the data points. It basically shows thevariables that should not be considered when training the models ondataset. There are several ways of identifying this outlier. One of them is Boxplot. The fig.9 and fig.10 show the boxplot for each of the attributes without 'ID' and 'Class' attribute.

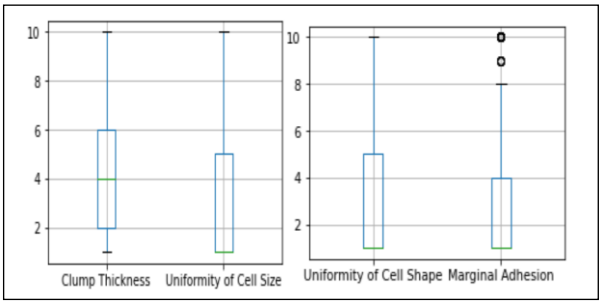

Fig9: Boxplot for identifying outliers

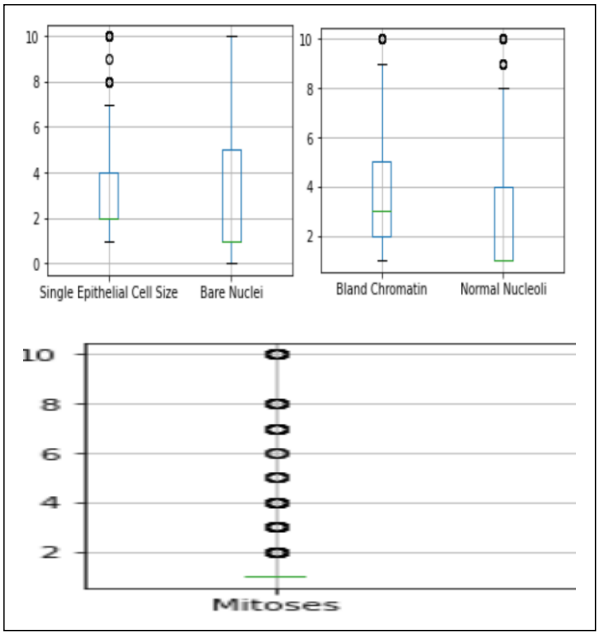

Fig 10: Boxplot for identifying outliers

In fig.9 and fig.10, it can be seen that 'Marginal Adhesion', 'Single Epithelial Cell Size', 'Bland Chromatin', 'NormalNucleoli' and 'Mitoses' attributes containoutliers. 'Mitoses' attribute contains the most outliers. After removing all of these outliers using 3-standard deviation, we were able to reduce 54 instances from the initial 699 instances. Now, the rest of the works will be performed on (699 - 54) = 645 instances out of 699.

4.2.2.3 Applying PCA (Principal Component Analysis)

In machine learning, PCA (Principal Component Analysis) is a technique used to reduce dimensions. Whenworking on real-life machine learning problems, there may be a thousand columns or features. In this circumstance, it is needed to find the most important features to correctly classify a particular problem. PCA plays an important role in this case. PCA helps to reduce the number of columns and as a result the overall complexity reduces. It reduces the number of columns that doesn't mean it removes the actual columns. It simply produces new variables that are constructed as the mixtures of the initial variables. We set the parameter 'n_components' as 5 to produce 5 new variables that are the most important. Thenall of thealgorithms were appliedon the dataset with these new important variables. And, the performances of some of them got improved drastically.

5. COMPARATIVE STUDY

Now, the performances of all the proposed machine learning techniques are going to be analyzed and compared. Table 3 shows the performance comparison after feature engineering the data.

Table 3. Performance analysis of all the algorithms after feature engineering

Algorithms	Accuracy	F1-score	Precision	Recall
Logistic Regression	62.79%	55.89%	55.95%	57.10%

Decision Tree Classifier	**96.89%**	**96.06%**	**95.18%**	**96.92%**
KNN	75.96%	56.74%	68.65%	57.00%
Random Forest Classifier	**95.34%**	**94.22%**	**92.30%**	**96.87%**
SVM	74.41%	42.66%	37.20%	5.0%
Gradient Boosting Classifier	**95.34%**	**94.11%**	**92.69%**	**95.88%**

This analysis shows that Decision Tree Classifier outperforms all the rest of the classifiers in terms of each of the performance metrics attaining the accuracy of (96.89%), the f1-score of (96.06%), the precision value of (95.18%) and the recall value of (96.92%).

6. CONCLUSION

In this paper, the authors have proposed 6 different machine learning techniques namely Logistic Regression, Decision Tree Classifier, KNN (K-Nearest Neighbors), Random Forest Classifier, SVM (Support Vector Machine), and Gradient Boosting Classifier to detect breast cancer. In this research work, the very popular Breast Cancer Wisconsin (Original) Dataset [1]collected from UCI machine learning repository has been used. The dataset contains 699 instances and 10 attributes. In the first step,all of the proposed algorithms were applied on the raw data without applying any feature engineering technique. In that case,Random Forest Classifier gave the highest accuracy of 92.14%. In the second step, the authors handled all of the 16 missing values using the mean value replacing technique, removed all the identified outliers using 3-standard deviation and then applied PCA to get the 5 most important features to achieve better performance. The analysis shows that after applying all of these techniques some of the algorithms attained much better accuracy than before. Decision Tree Classifier outperformed all the rest of the classifiers in terms of each performance metrics attaining the accuracy of (96.89%), the f1-score of (96.06%), the precision value of (95.18%) and the recall value of (96.92%). Whereas, the most competitive techniques such as Random Forest Classifier attained the accuracy of (95.34%), the f1-score of (94.22%), the precision value of (92.30%) and the recall value of (96.87%), Gradient Boosting Classifier attained the accuracy of (95.34%), the f1-score of (94.11%), the precision value of (92.69%) and the recall value of (95.88%).

In this research work, the authors have used Jupyter Notebook (version 6.4.5) as thework environment.Python programming language(version 3.10.0) has been used as the primary programming language. The PC configuration was:

- Processor: AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx 2.10 GHz.
- Ram: 8.00 GB.
- System type: 64-bit operating system, x64-based processor.
- Operating system: Windows 11 Home Single Language Edition.

7. ACKNOWLEDGMENTS

We, the authors of this paper are reallygrateful to theUCI machine learning repository team for providing such an authentic dataset on breast cancer. We are also thankful to the creator of the dataset Dr. William H. Wolberg (physician) at the University of Wisconsin Hospitals, Madison, Wisconsin, USA.We would also like to show our appreciation to the anonymous reviewers for their invaluable comments and suggestions.

8. REFERENCES

[1] Breast Cancer Wisconsin (Original)Dataset collected from UCI machine learning repositorycreated by Dr. William H. Wolberg (physician) at the University of Wisconsin Hospitals, Madison, Wisconsin, USA. https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+%28original%29

[2] Begum SA, Mahmud T, Rahman T, Zannat J, Khatun F, Nahar K, Towhida M, Joarder M, Harun A, Sharmin F. Knowledge, Attitude and Practice of Bangladeshi Women towards Breast Cancer: A Cross Sectional Study. Mymensingh Med J. 2019 Jan;28(1):96-104. PMID: 30755557.

[3] Borges, Lucas Rodrigues. "Analysis of the wisconsin breast cancer dataset and machine learning for breast cancer detection." Group 1.369 (1989): 15-19.

[4] T. Brito-Sarracino, M. Rocha dos Santos, E. Freire Antunes, I. Batista de Andrade Santos, J. Coelho Kasmanas and A. C. Ponce de Leon Ferreira de Carvalho, "Explainable Machine Learning for Breast Cancer Diagnosis," 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), 2019, pp. 681-686, doi: 10.1109/BRACIS.2019.00124.R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[5] F. Teixeira, J. L. Z. Montenegro, C. A. da Costa and R. da Rosa Righi, "An Analysis of Machine Learning Classifiers in Breast Cancer Diagnosis," 2019 XLV Latin American Computing Conference (CLEI), 2019, pp. 1-10, doi: 10.1109/CLEI47609.2019.235094.

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

[7] S. Laghmati, A. Tmiri and B. Cherradi, "Machine Learning based System for Prediction of Breast Cancer Severity," 2019 International Conference on Wireless Networks and Mobile Communications (WINCOM), 2019, pp. 1-5, doi: 10.1109/WINCOM47513.2019.8942575.

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