Predict Breast Cancer Using Several Machine Learning Algorithms

Eslam Yahia[*], Maria Sameh[**], Mira Rafat[**], Ester Elzek[**], Selvia Ashraf[**], Sandy Esam[**], Steven Waheed[**]

Faculty of Computers and Information, Assiut University
Bioinformatics Department

Abstract- Breast cancer is a type of tumor that occurs in the tissues of the breast. It is the most common type of cancer found in women around the world and it is among the leading causes of death in women. Our main focus is to analyze different Machine Learning algorithms in order to find out the most appropriate algorithm that will give the highest accuracy of prediction. The main purpose of this paper is to highlight four algorithms of machine-learning algorithms that are being used for breast cancer prediction. Machine learning algorithms such as Logistic Regression (LR), K-Nearest Neighbor (KNN), Decision Tree (DT), and Random Forest (RF) are the most popular techniques. The results showed that the decision tree algorithm has the highest accuracy of about 100%, then the logistic regression and KNN algorithm has the accuracy of about 98%, and for Random Forest algorithm has the accuracy of about 97%.

Keywords- Machine Learning, Breast Cancer Prediction, Classification, Supervised Learning, Tumor detection

I. INTRODUCTION

Breast cancer is a common and dangerous disease in women, cancer is the form of abnormal cells that come into these mutated cells. Spreads throughout the body, leading to death in diagnosis and treatment. Breast cancer affects both men and women, but it is more common in women.

There are two types of breast cancer, Malignant and Benign. The first is classified as harmful has the ability to infect other organs and is cancerous, Benign is classified as non-cancerous. This disease infects the women's chest and specifically glands and milk ducts, breast cancer spreads to other organs frequently and could be through the bloodstream.

These two types classified in four main types: normal, benign, in-situ carcinoma and invasive carcinoma.

1. A benign tumor involves a minor change in the breast structure. It is not harmful and does not classify as a harmful cancer.
2. In cases of in-situ carcinoma the cancer is only in the mammary duct lobule system and does not affect other organs. This type is not dangerous and can be treated if diagnosed early.
3. Invasive carcinoma is considered to be the most dangerous type of breast cancer, as it can spread to all other organs. Breast cancer can be detected using several methods including: X-ray mammography, ultrasound (US), Computed Tomography (CT), Portion Emission Tomography (PET), Magnetic Resonance Imaging (MRI) and breast temperature measurement.

Breast cancer can be diagnosed using one of two approaches: histopathological image analysis or genomics

- Histopathology images are microscopic images of breast tissue that are extremely useful in early treatment of the cancer.

- As for genomics, radio-genomics is an emerging research field focusing on multi-scale associations between medical imaging and gene expression data. Radio-genomics provide both radio-logical and genetic features that can enhance diagnosis. It can analyze tissues at the molecular level, helping with prediction and early detection of cancer.

The BRCA gene test is a blood test that uses DNA analysis to identify harmful changes (mutations) in one of the breast cancer susceptibility genes - BRCA1 and BRCA2.
Women who inherit mutations in these genes are at increased risk of developing breast cancer and ovarian cancer compared to the general population.

Those most likely to have a genetic mutation, based on a personal or family history of breast or ovarian cancer, are offered testing for the BRCA genes. Women at high risk of breast and ovarian cancer aren't routinely tested for BRCA genes.

Gene test results are not always clear. A positive result means that you carry a genetic mutation that increases your risk of cancer, and you can work with your doctor to manage this risk. A negative result could mean that you don't have the mutation, or you may have another genetic mutation that doctors haven't yet discovered. The test may also identify another form of the gene that doctors aren't aware of. In these situations, it's not always clear what the results mean for cancer risk.

Most women considering genetic testing undergo genetic counseling. Genetic counseling can help you understand what the test results mean for your health, help you decide if genetic testing is right for you, and recommend a specific set of genetic testing based on your family's health history.

Machine learning and Data mining techniques are straightforward and effective ways to understand and predict data.
To classify the cancer is benign or malignant we will use:

1. Logistic Regression Algorithm: is a supervised ML algorithm that is used to predict the probability of certain classes based on some dependent variables.
2. K-Nearest Neighbor Algorithm: is a supervised ML algorithm that is used to predict both classification and regression problems based on a number of the nearest points.
3. Decision Tree Algorithm: is a supervised ML algorithm that uses the divide and conquer concept, it divides the dataset into smaller numbers of subsets. These smaller sets of data can make predictions with the highest level of precision.
4. Random Forest Algorithm: is a supervised ML algorithm that is used to solve both classification and regression problems, it contains several decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.

II. PROBLEM DEFINITION AND BACKGROUND

PROBLEM DEFINITION: -

Breast cancer is a condition in which the breast's cells expand out of control. Breast cancer takes various forms. Which breast cells turn into cancer determine the type of breast cancer.

Different areas of the breast might emerge to breast cancer. There are three basic components of a breast: connective tissue, ducts, and lobules.

The glands that release milk are called lobules. Milk travels through tubes called ducts to the nipple. The connective tissue, which is made up of fatty and fibrous tissue, envelops and holds everything in place. The ducts or lobules are where most breast cancers start. Blood and lymph vessels are two ways that breast cancer can travel outside of the breast. Breast cancer is said to have metastasized when it spreads to other body regions.

Breast cancer causes
- A family history.
- inheriting mutations in the BRCA2, BRCA1 (more common with Ashkenazi Jewish heritage) and CHEK2.
- exposure female hormones (natural and administered)
- getting your first period before turning 12.
- a previous diagnosis of breast cancer.
- a past medical history containing specific non-cancerous breast diseases.

Early breast cancer identification should take into account two main clinical scenarios, similar to any other primarily chronic condition. The diagnosis is either the result of an early examination of a patient's complaint or a physical exam finding, or it is established by a screening test that detects the disease in an asymptomatic patient before it creates symptoms.

Breast cancer diagnoses at advanced stages are more common in LMIC. According to the information that is now available, 70% of breast cancer patients in HIC present with stages I–II, compared to fewer than 50% of those diagnosed in LMIC.

It can be hard to diagnose breast cancer which is inflammatory. Frequently, a screening mammography or a physical examination will not reveal any lumps. Additionally, the thick breast tissue that most women with inflammatory breast cancer have makes it more harder to identify the disease during a screening mammogram.

As early diagnosis and treatment are connected to better prognoses as compared to worse outcomes caused by significantly delayed diagnosis, the delayed diagnosis is more to blame than the disease that caused patient's mortality.

In this paper, we will use the Kaggle breast cancer dataset, and it contains the following features:
1. ID number.
2. Diagnosis (M = malignant, B = benign).

We compute ten real-valued features for each cell nucleus:
A. radius (mean of distances from the center to points on the perimeter).
B. texture (standard deviation of gray-scale values).
C. perimeter.
D. area.
E. smoothness (local variation in radius lengths).
F. compactness (perimeter^2 / area - 1.0).
G. concavity (severity of concave portions of the contour).
H. concave points (number of concave portions of the contour).
I. symmetry.
J. fractal dimension ("coastline approximation" - 1).

For each image, the mean, standard error, and "worst" or worst feature—the mean of the three largest values—were calculated, yielding a total of 30 features. For instance, field 3 is Mean Radius, field 13 is Radius SE, field 23 is Worst Radius.

All feature values are recoded with four significant digits.

BACKGROUND: -

Each year, breast cancer affects hundreds of thousands of women. Despite significant improvements in breast cancer diagnosis and treatment, some patients arrive with metastatic illness at the time of initial diagnosis, and the risk of recurrence endures for decades after the initial occurrence.

Breast cancer is a malignancy that may develop in either or both breasts. This implies that it may also represent a threat to life. This disease always begins as a tumor, much like many types of cancer do. This happens when malformed cell divisions result in tumors that spread and transform into breast cancer. This cancer can start in several locations within the breast and has a terrible effect on both men and women. Bloodstream metastasis, or the spread of breast cancer to other body organs, can occur. Tumors in the brain, bones, liver, lungs, and other parts of your body are possibilities.

Bone fractures, blocked blood vessels, and spinal cord pressure are possible dangers.

It included A few early signs may include a new lump in the breast or underarm, nipple itching or discharge, and changes in the texture of the skin above.

The most typical types of breast cancer include:
• Invasive ductal carcinoma. The cancerous cells start off in the ducts before spreading to other areas of the breast tissue. Additionally, invasive cancer cells have the ability to expand, or metastasis, to other bodily regions.
• Invasive lobular carcinoma. The lobules are where cancer cells first start, then they travel from the lobules to the nearby breast tissues. These invasive cancer cells can also infect different body regions.

There are a Few early symptoms may include new lump in the underarm or in breast, itching or discharge from the nipples, and skin texture change of the nipple or breast.
The most common kinds of breast cancer are:

- Invasive ductal carcinoma. The cancer cells start dividing in the ducts and then grow outside the ducts into other positions of the breast tissue. Invasive cancer cells can also spread, or metastasize, to other positions of the body.
- Invasive lobular carcinoma. Cancer cells start dividing in the lobules and then spread from the lobules to the breast tissues that are close by. These invasive cancer cells can also spread to other positions of the body.

There are many ways to diagnosis the breast cancer such as:

- ***Breast ultrasound***. A machine that uses sound waves to make pictures, called sonograms, of areas inside the breast.
- ***Diagnostic mammogram***. If you have a problem in your breast, such as lumps, or if an area of the breast looks abnormal on a screening mammogram, doctors may have you get a diagnostic mammogram. This is a more detailed X-ray of the breast.
- ***Breast magnetic resonance imaging (MRI)***. A kind of body scan that uses a magnet linked to a computer. The MRI scan will make detailed pictures of areas inside the breast.
- ***Biopsy***. This is a test that removes tissue or fluid from the breast to be looked at under a microscope and do more testing. There are different kinds of biopsies (for example, fine-needle aspiration, core biopsy, or open biopsy)

The algorithms that will help us to classify the cancer is benign or malignant are:

- ***Logistic Regression Machine Learning:*** is basically a supervised classification algorithm (a type of machine learning in which machines are trained using "labelled" data, and on the basis of that trained data, the output is predicted) of Machine Learning

algorithms (shown in Figure 1). This simply means it fetches its roots to the field of Statistics and the main role of Logistic Regression in Machine Learning is predicting the output of a categorical dependent variable from a set of independent variables. In simple words, categorical dependent variable means a variable that is dichotomous or binary in nature having its data coded in the form of either 1 (stands for success/yes) or 0 (stands for failure/no).

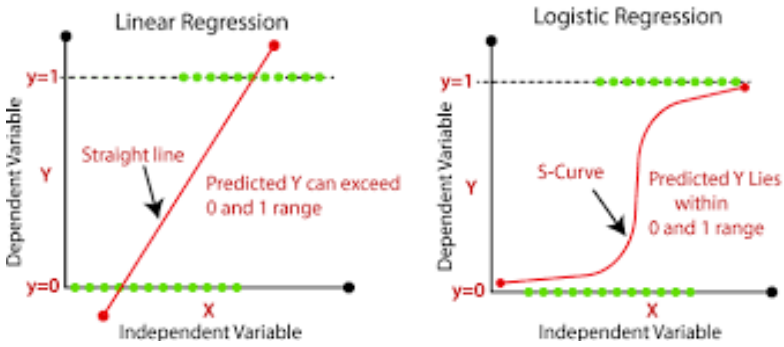

Figure 1. Logistic Regression

- ***The K-Nearest Neighbor Algorithm:*** the name refers to finding the k-nearest neighbors (shown in Figure 2) to make a prediction for unknown data that can solve both classification and regression problems:
    1. *In the classification problem*, the KNN algorithm will attempt to infer a new data points class by looking at the classes of the majority of its K-Neighbors.
    2. *In regression problems*, the KNN algorithm will predict a new data points continuous value by returning the average of the K-Neighbors values.

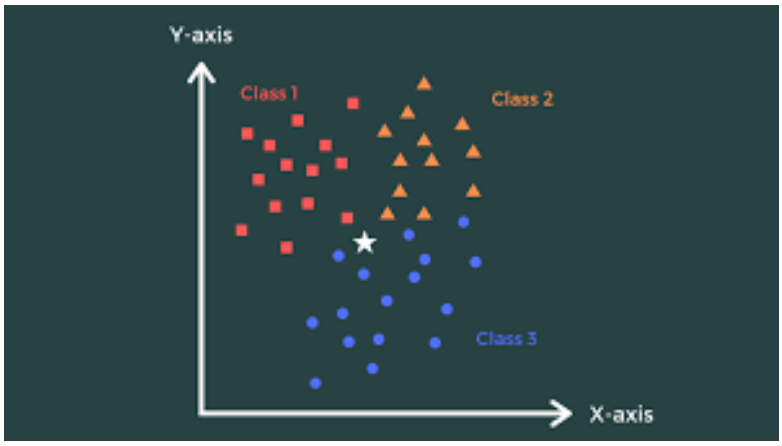

Figure 2. K-Nearest Neighbor.

- ***A Decision Tree Algorithm:*** It is a tool that has applications spanning several different areas. Decision trees (shown in Figure 3) can be used for classification as well as regression problems. The name itself suggests that it uses a flowchart like a tree structure to show the predictions that

result from a series of feature-based splits. It starts with a root node and ends with a decision made by leaves.

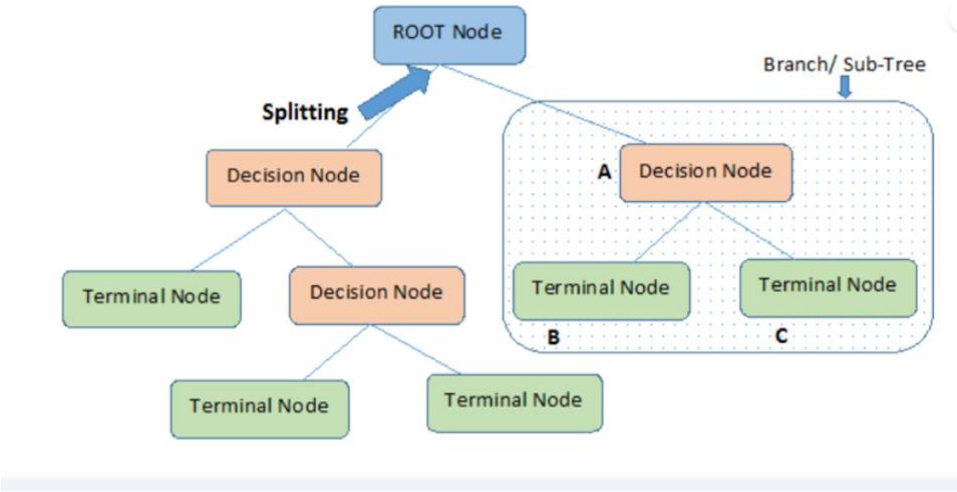

Figure 3. Decision Tree

- ***Random forest Algorithm:*** is a commonly-used machine learning algorithm, which combines the output of multiple decision trees to reach a single result. It is a commonly used classification and regression algorithm. As classification and regression are the most significant aspects of machine learning we can say that the Random Forest Algorithm (shown in Figure 4) is one of the most important algorithms in machine learning.

The working of the Random Forest Algorithm is quite intuitive. It is implemented in two phases:

1. *First*. is to combine N decision trees with building the random forest.
2. *Second*. is to make predictions for each tree created in the first phase.

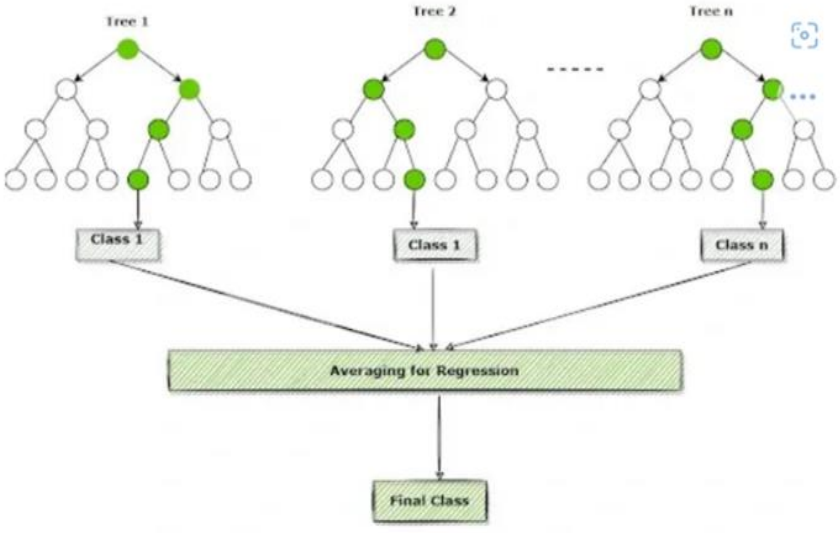

Figure 4. Random Forest

III. *RELATED WORK & METHODOLOGY*

RELATED WORK:-

There are a lot of studies in this field that use different algorithms to detect breast cancer through imaging and through genomics.

The authors in [1] use the Random Forest algorithm to analyze medical case diagnoses of breast cancer and obtain high prediction accuracy. They use the Wisconsin breast cancer dataset (WDBC) that has ten real-valued features are computed for each cell nucleus and get 98.8% accuracy and 99.3% Receiver Operating Characteristic (ROC), which shows the trade-off between sensitivity or (TPR) and specificity (1 - TPR).

The author in [2] uses a Logistic Regression algorithm from sklearn package and he also uses the same dataset as the authors in the paper [1] (WDBC). it starts from the measurement data of biopsy cells in women with breast cancer, and the classification results show that by selecting maximum texture and maximum perimeter the accuracy will be 96.5%, which is better than other methods like selecting mean radius and the mean texture the accuracy can reach 90.48%.

The authors in [3] use mammography images dataset, and talk about the Radiologists to predict if the mammography images have cancer or not, but they may miss about 15%. they use a new method with high accuracy, it consists of two main parts. First, the image processing techniques are used to prepare the mammography images for the feature extraction process, and the features are used as input for Back Propagation Neural Network (BPNN) model and Logistic Regression (LR). In this paper they extracted 209 images for 50 patients cases who have breast cancer. And they get 0.07 Mean Square Error (MSE), and the regression value of the neural network exceeded 93.7%.

The authors in [4], use genetic sequencing and histopathological images to detect and treat breast cancer with the help of deep learning and machine learning. They also provide recommendations to researchers who will work in this field. They found that according to National Breast Cancer Foundation 2020 alone, more than 276,000 new cases of invasive breast cancer and more than 48,000 non-invasive cases were diagnosed in the US. They used a lot of deep learning and machine learning algorithms that show high accuracies to detect breast cancer.

The authors in [5] focused on breast cancer in India. They found that there is a chance of fifty percent fatality in a case as one in two women diagnosed with breast cancer die in the cases of Indian women. Their paper compared three machine learning algorithms (Random forest, KNN, Naïve Bayes) which are commonly used for breast cancer detection, And they used Wisconsin breast cancer dataset (WDBC). They found that Random Forest has 94.7 accuracy, KNN has 95.9 accuracy, and Naïve Bayes has 94.47 accuracy.

METHODOLOGY:-

Our target topic is breast cancer detection using different machine learning algorithms. We ended up using more than 20 of the most recent papers related to breast cancer and machine learning.

We use the following search statements: *"Machine Learning"*, "*Breast Cancer*", "*Cancer Detection*", "*Logistic Regression*", "*Random Forest*", "*Decision Tree*", "*K Nearest Neighbors*", "*Classification*" and "*AI algorithms*". more than 1,000 papers were found but we focused on the papers which talked about Breast Cancer Detection with machine learning algorithms. We applied the following criteria on each paper:

1- The language is English.
2- The topic is related to breast cancer detection.
3- The paper discussed machine learning algorithms.
4- Only journals and conference publications are retaining.
5- Only medical engineering publications are kept.

We wrote the main information such as the title of the paper, the list of authors. Then, we included some information to conduct the systematic review, such as the algorithms used and whether the paper discusses only ML and Breast Cancer detection, the model accuracy and performance evaluation parameters, the dataset and the features.

A. Dataset Description:

The project is based on Wisconsin Diagnosis Breast Cancer dataset. It has 569 instances and 32 attributes shown in Table 1. And there is a column contains missing values but the other columns have no missing values. The output variable is either benign, or malignant.

Dataset	Number of Attributes	Number of Instances	Number of Classes
Wisconsin Diagnosis Breast Cancer	32	569	2

Table 1. Description of WDBC Dataset

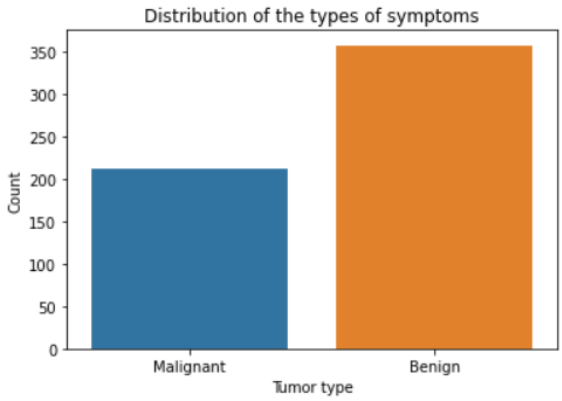

Figure 5. Distribution of all the types of symptoms

B. Performance Metrics:

This section describes the parameters that are used for measuring the performance of ML techniques. A confusion matrix for actual and predicted class is derived comprising of the standard four values True-Positive, False-Positive, True-Negative and False-Negative to evaluate the performance.

1. Accuracy:

It's a good predictor for performance of the trained model. It may defined as the measure of the correct prediction in correspondence to the wrong ones.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Recall:

- it asks the question "*how completely you find the label in question*".
- the goal is to find the most number of positive labels.
- also known as sensitivity, is the fraction of examples classified as positive, among the total number of positive examples. In other words, the number of true positives divided by the number of true positives plus false negatives.

$$Recall = \frac{TP}{TP + FN}$$

3. Precision:

- it asks the question "*How well you guess the label in question*".
- the goal is to minimize mistakes in guessing positive labels.
- is the fraction of true positive examples among the examples that the model classified as positive. In other words, the number of true positives divided by the number of false positives plus true positives.

$$Precision = \frac{TP}{TP + FP}$$

4. F1 Score:

- it asks the question "*how good and complete are the predictions*".
- the goal is to minimize both precision and recall.
- it's the harmonic mean.

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

IV. RESULTS AND DISCUSSION

Python 3.11, SKLearn, matplotlib, seaborn, pandas and imblearn are used in this experimental environment.

In order to understand the impact of test data on diagnostic results. It's important to observe the correlation between attributes. As shown in Figure 6.

From Figure 6. We can see that there are some attributes highly correlated with each other, for example perimeter_mean has a high correlation with area_mean, also perimeter_worst, radius_worst, area_worst, perimeter_se and area_se.

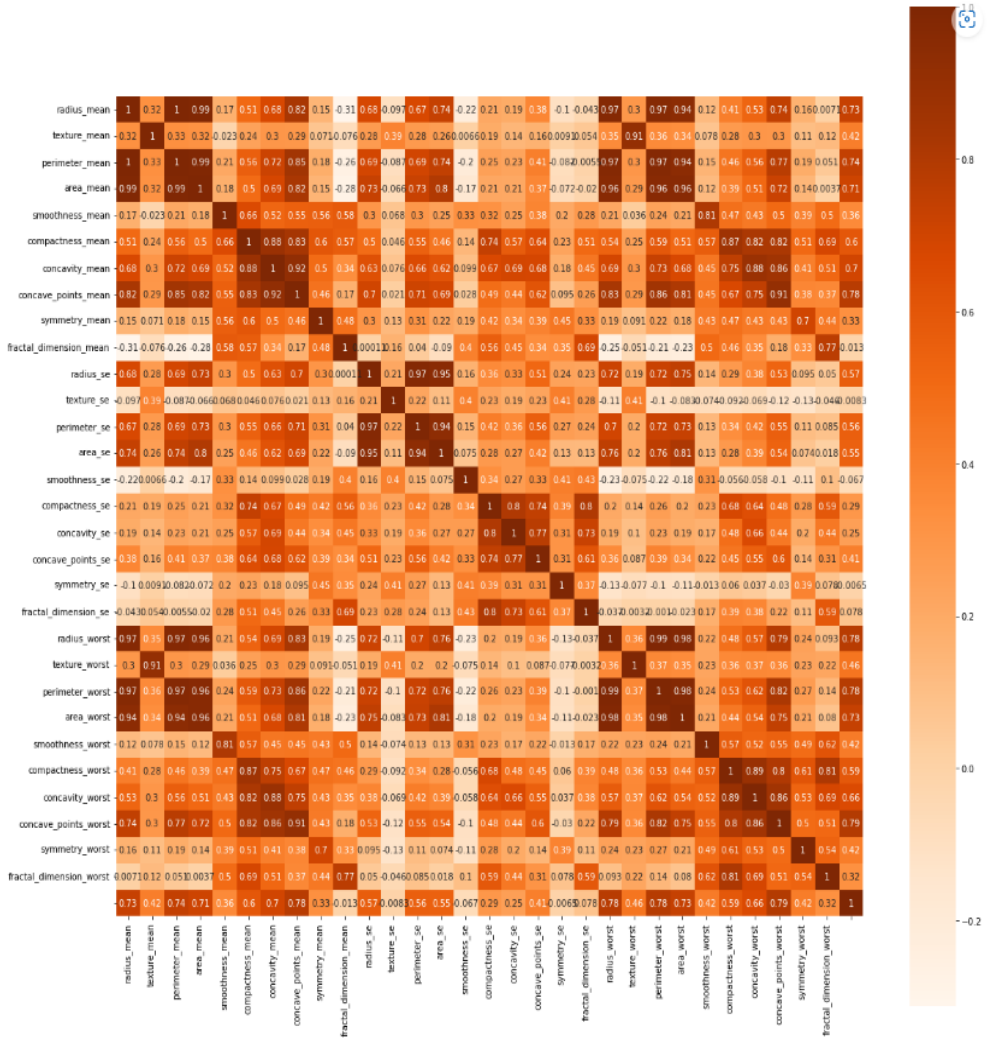

Figure 6. Correlation of features

After that we perform data cleaning such as removing the outliers, dropping some features, and normalizing the data, we trained four classification machine learning algorithms on 80% of the data and test it using 20% to detect breast cancer:

1. **Logistic Regression.**
   We used "*liblinear*" solver and Random state is 0, and it gave us 99% accuracy as shown in Table 1. we used other techniques to measure the performance of the model such as confusion metrics shown in Figure 7.

   From the confusion metric, this model has 100% Precision, 99% Recall, and 99% F1 score for the Benign class as shown in Table 4. And has 98% Precision, 100% Recall, and 99% F1 score for the Malignant class as shown in Table 3.

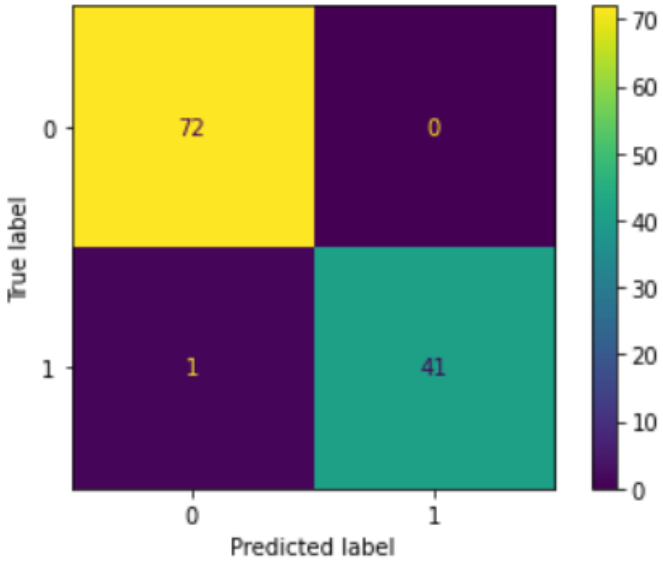

Figure 7. Logistic Regression Confusion Metric

2. **KNN.**
   We used the parameter number of neighbors is 5, and it gave us 99.3% accuracy as shown in Table 2. we used other techniques to measure the performance of the model such as confusion metrics shown in Figure 8.

   From the confusion metric, this model has 97% Precision, 100% Recall, and 99% F1 score for the Benign class as shown in Table 4. And has 100% Precision, 95% Recall, and 98% F1 score for the Malignant class as shown in Table 3.

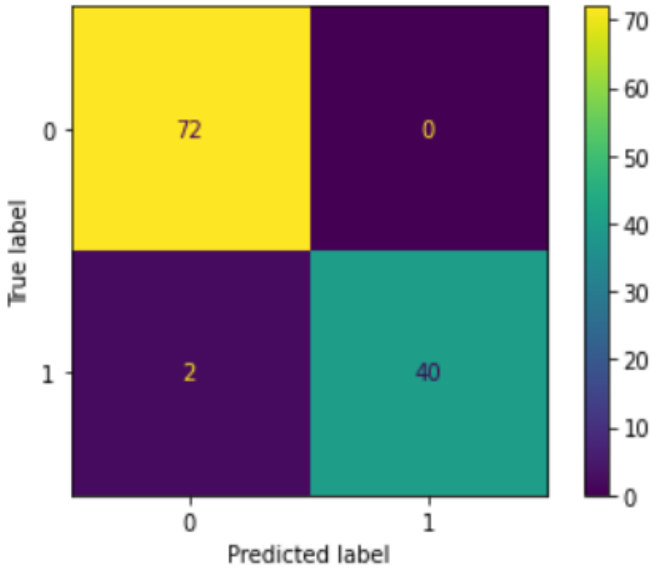

Figure 8. KNN Confusion Metric

3. Decision Tree.

We used these parameters gini by default and the minimum samples leaf is 5, and it gave us 100% accuracy as shown in Table 2. we used other techniques to measure the performance of the model such as confusion metrics shown in Figure 9.

From the confusion metric, this model has 100% Precision, 100% Recall, and 100% F1 score for the Benign class as shown in Table 4. And has 100% Precision, 100% Recall, and 100% F1 score for the Malignant class as shown in Table 3.

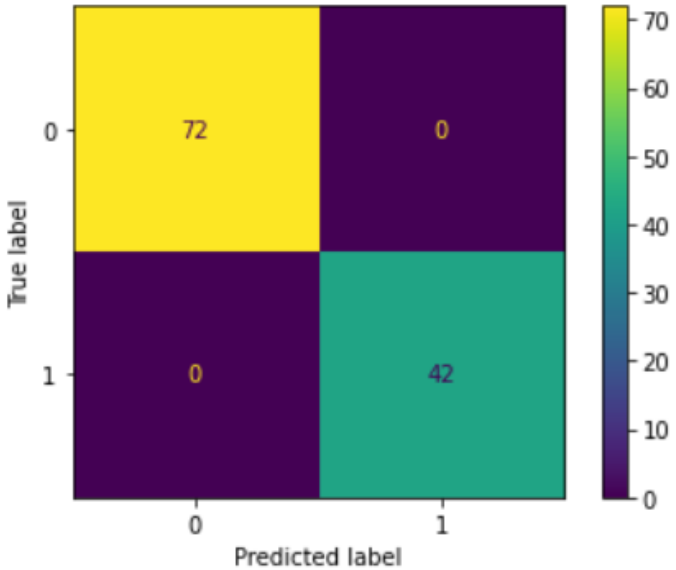

Figure 9. Decision Tree Confusion Metric

From Figure 10. We can see that decision tree is used to perform feature selection, the nodes on the top levels are the best features and the nodes on the bottom are less important.

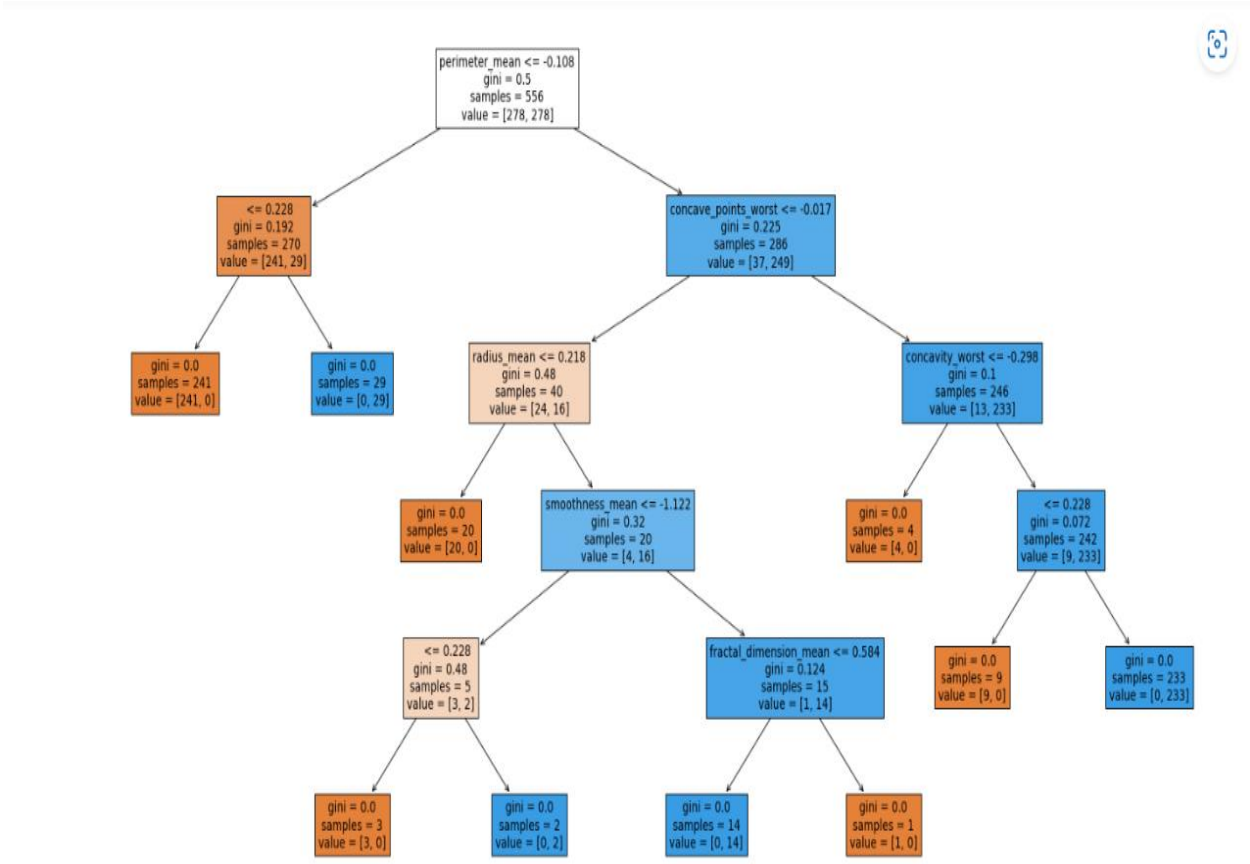

Figure 10. Decision Tree

4. Random Forest.

We used bootstrap equal false, and the max depth is 5 and the number of estimators is 5, and it gave us 97% accuracy as shown in Table 2. we used other techniques to measure the performa nce of the model such as confusion metrics shown in Figure 11.

From the confusion metric, this model has 99% Precision, 97% Recall, and 98% F1 score for the Benign class as shown in Table 4. And has 95% Precision, 98% Recall, and 96% F1 score for the Malignant class as shown in Table 3.

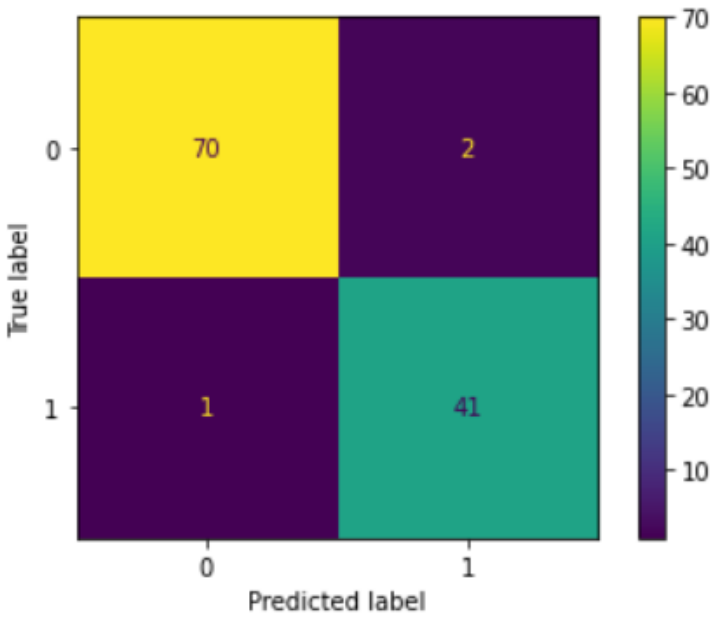

Figure 11. Random Forest Confusion Metric

Which algorithm has the highest accuracy?
From Table 2 and Figure 12. We can see that Decision Tree has the best accuracy.

	Logistic Regression	K Nearest Neighbors	Decision Tree	Random Forest
Accuracy (%)	98%	98%	100%	97%

Table 2. Model Accuracies

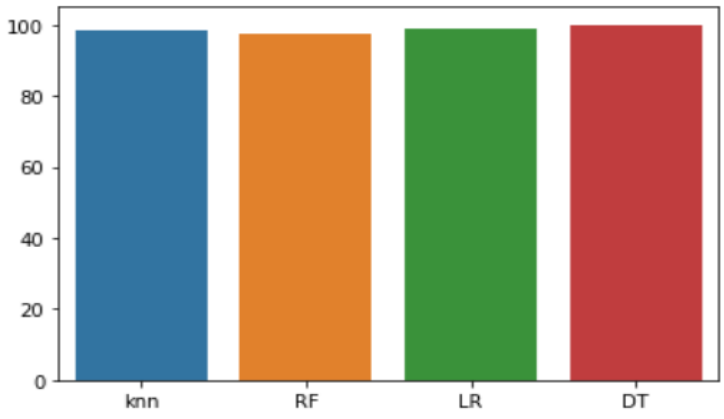

Figure 12. Bar plot for Models Accuracies

Which algorithm has the highest performance measurements with the Malignant class?
From Table 3. We can see that K Nearest Neighbors and Decision Tree have the best Precision (100%). Logistic Regression and Decision Tree have the best Recall (100%). Decision Tree has the best F1 Score (100%).

	Logistic Regression	K Nearest Neighbors	Decision Tree	Random Forest
Precision (%)	98%	100%	100%	95%
Recall (%)	100%	95%	100%	98%
F1 Score (%)	99%	98%	100%	96%

Table 3. Models Performance with Malignant class

Which algorithm has the highest performance measurements with the Benign class?
From Table 4. We can see that K Nearest Neighbors and Decision Tree have the best Recall (100%). Logistic Regression and Decision Tree have the best Precision (100%). Decision Tree has the best F1 Score (100%).

	Logistic Regression	K Nearest Neighbors	Decision Tree	Random Forest
Precision (%)	100%	97%	100%	99%
Recall (%)	99%	100%	100%	97%
F1 Score (%)	99%	99%	100%	98%

Table 4. Models Performance with Benign class

V. CONCLUSION AND FUTURE WORK

Most papers published in the field of breast cancer detection and subtype classification use machine learning techniques. In this paper we focused on four of machine learning algorithms (Logistic Regression, K-Nearest Neighbor, Decision Tree and Random Forest), each algorithm gave high accuracy on the data we used. And for performance measurements we used confusion metrics to assess the performance of each model.

Future studies should include AUC, ROC, Cross Validation, and use other machine learning algorithms such as Support Vector Machine (SVM) and XGBoost. Use large and different types of datasets such as MRI, and deep learning algorithms such as CNN, LSTM and GAN perform very well with these type of datasets, Future researchers may also focus on extracting significant features from genetic expression data to obtain better results.

REFERENCES

[1] [Bak82] Larry H Baker. Breast cancer detection demonstration project: five-year summary report.

[2] [CA21] Bahzad Charbuty and Adnan Abdulazeez. Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2(01):20–28, 2021.

[3] [KGG85] James M Keller, Michael R Gray, and James A Givens. A fuzzy k-nearest neighbor algorithm. IEEE transactions on systems, man, and cybernetics, (4):580–585, 1985.

[4] [Liu18] Lei Liu. Research on logistic regression algorithm of breast cancer diagnose data by machine-learning. In 2018 International Conference on Robots & Intelligent System (ICRIS), pages 157–160. IEEE, 2018.B. Smith, “An approach to graphs of linear forms (Unpublished work style),” unpublished.

[5] [Rie94] Martin Riedmiller. Advanced supervised learning in multi-layer perceptrons—from back-propagation to adaptive learning algorithms. Computer Standards & Interfaces, 16(3):265– 278, 1994.

[6] [SCQ17] Alessia Sarica, Antonio Cerasa, and Aldo Quattrone. Random forest algorithm for the classification of neuroimaging data in alzheimer’s disease: a systematic review. Frontiers in aging neuroscience, 9:329, 2017.

[7] [CRH05] Pamela Cowin, Tracey M Rowlands, and Sarah J Hatsell. Cadherins and catenins in breast cancer. Current opinion in cell biology, 17(5):499–508, 2005

[8] [SDS+10] Ganesh N Sharma, Rahul Dave, Jyotsana Sanadya, Piush Sharma, KK3255438 Sharma, et al. Various types and management of breast cancer: an overview. Journal of advanced pharmaceutical technology & research, 1(2):109, 2010.

[9] Dai, Bin, Rung-Ching Chen, Shun-Zhi Zhu, and Wei-Wei Zhang. "Using random forest algorithm for breast cancer diagnosis." In 2018 International Symposium on Computer, Consumer and Control (IS3C), pp. 449-452. IEEE, 2018.

[10] Liu, Lei. "Research on logistic regression algorithm of breast cancer diagnose data by machine learning. " In 2018 International Conference on Robots & Intelligent System (ICRIS), pp. 157-160. IEEE, 2018.

[11] Alarabeyyat, Abdulsalam, and Mohannad Alhanahnah. "Breast cancer detection using k-nearest neighbor machine learning algorithm." In 2016 9th International Conference on Developments in eSystems Engineering (DeSE), pp. 35-39. IEEE, 2016.

[12] Nassif, Ali Bou, Manar Abu Talib, Qassim Nasir, Yaman Afadar, and Omar Elgendy. "Breast cancer detection using artificial intelligence techniques: A systematic literature review." Artificial Intelligence in Medicine (2022): 102276.

[13] Sharma, Shubham, Archit Aggarwal, and Tanupriya Choudhury. "Breast cancer detection using machine learning algorithms." In 2018 International conference on computational techniques, electronics and mechanical systems (CTEMS), pp. 114-118. IEEE, 2018.