EasyChair Preprint

№ 8706

An Efficient Machine Learning Algorithm for Breast Cancer Prediction

Yousif A. Alhaj, Marwan M. Al-Falah, Abdullah M. Al-Arshy, Khadeja M. Al Nashad, Zain Alabedeen Ali Al Nomi, Badr A. Al Badawi and Mustafa S. Al Khayat

August 25, 2022

An Efficient Machine Learning Algorithm for Breast Cancer Prediction

Yousif A. Al Haj[a,*], Marwan M. Al Falah[b,*], Abdullah M. Al-Arshy[a], Khadeja M. Al-Nashad[b], Zain Alabedeen A. Al-Nomi[b], Badr A. Al-Badawi[b], Mustafa S. Al-Khayat[b]

[a] *Sana'a University, Faculty of Education, Humanities & Applied Science, Sanaa, Yemen; yalhag@gmail.com (Y.A.A);*
[b] *Knowledge & Modern Science University (KMSU), Faculty of Information Technology & Engineering, Sanaa, Yemen*
[c]

Abstract:

Cancer is a leading cause of death worldwide, with breast cancer (BC) being the most common and prevalent with 2.26 million cases each year, and the main cause of women's deaths, so early and correct detection to discover BC in its first phases, help to avoid death by describing the appropriate treatment and to maintain human life. Cancer cells are divided into two types Malignant and Benign. The first type is more dangerous and the second type is less dangerous. Due to the existence of artificial intelligence (AI) and the great direction to the use of machine learning in medicine, doctors get accurate results for diagnosis. In this paper, we tend to use the Wisconsin Breast Cancer Patients Database (WBCD) which has been collected from the UCI repository. In this paper, the WBCD dataset is divided into 75% for training and 25% for testing using a split test train. We addressed to research the performance of various well-known algorithms in the discovery of BC such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR) and Artificial Neural Networks (ANN). High results indicate that the RF algorithm is 98.2% superior to the rest of the machine learning algorithms.

***Keywords*:** **Machine Learning, Breast Cancer, Classification algorithms, WBCD**

I. INTRODUCTION

BC is one of the leading causes of death in the world. It is the most common dangerous cancer in women, exceeding lung cancer in the rate of new cases and increased mortality, according to statistics released by the International Agency for Research on Cancer (IARC) in December 2020. The number of new cases of cancer doubles insanely, reaching 2.26 million new cases and 685,000 deaths in 2020 [1]. When healthy celled change and begin to grow, a mass of cells called a tumor forms, and thus BC results, like any cancer. If it spreads in the body, it is a malignant tumor, and if it does not spread in the body, it is a benign tumor. It is liable for the loss of life of many girls across the globe. During our life, one in 5 people will catch cancer during their lifetime. It becomes necessary in the least time predicts BC. We can achieve the best way to save anyone even from

getting cancer with a timely diagnosis. It is possible to identify whether a cancer is malignant or benign by looking at a person's symptoms and comparing them with those of the victim, as well as extracting key features from complex data utilizing machine learning algorithms which are used for carcinoma prediction and identification [2]. Machine Learning algorithms have provided help in several fields such as medicine, early-stage cancer prediction, and varied diseases. Furthermore, machine learning techniques have helped in predicting the standard of cancer whether or not an individual has benign or malignant cancer, and this method is economical and with no errors [3].

In this paper, several machine learning algorithms are implemented to predict breast cancer by using the WBCD dataset. Furthermore, the comparison between their performance to evaluate the efficiency in terms of accuracy was discussed in this paper.

The rest of the paper divided as related work is obtainable in Section II. The essential technique is illustrated in Section III. Section IV represents the received results. At last, Section V concludes the total work.

II. RELATED WORK

Researchers have already done loads of analysis by applying machine learning algorithms to a medical dataset for classification to seek out a modality in an exceeding dataset for quicker diagnostic and prediction.

Comparative study of machine learning algorithms for breast cancer prediction [2]. Here they have got applied algorithms like LR and DT, each is in comparison as each has been expected to generate predictions with excessive accuracy. Their studies suggest selection tree classifier is selected because it had barely extra accuracy than the LR model.

Another paper called comparative analysis to predict breast cancer using machine learning algorithms[3]. They applied different machine learning algorithms such as SVM, KNN, NB, and DT. The applied models were compared supported Precision, Recall, and accuracy which might be calculated victimization specific equations. They conclude that the factitious neural network (ANN) provides higher prediction at 97.85% compared to the remainder of the algorithms.

Another analysis paper referred to as malignant and benign breast cancer classification victimization machine learning algorithms [4] uses several algorithms to diagnose BC. They used LR, SVM, RF, NB, DT, and KNN for the prediction of cancer. Their paper presented that SVM is the best algorithm for prediction.

For BC prediction exploitation machine learning in [5]. They used four main algorithms enforced during this study namely, LR, SVM, NB, and associated RF using BC dataset. It showed that the RF outperformed all alternative algorithms with an accuracy of 99.76%.

In a neuro-fuzzy inference model to classify diabetic retinopathy [6], the authors discussed the results of using the Adaptive Neuro-Fuzzy Inference System (ANFIS) to classify the retina's level of damage caused by DR into four classes (low, moderate, high, and not damaged), where we also benefited from the part of explaining the architecture of neural networks.

An analysis on SVM & ANN using breast cancer dataset [7], where they have provided explanations of different machine learning approaches and its one of the important applications in

BC diagnosis and prognosis which is used to analyze the data (WBCD). In this study, they demonstrate the modeling of breast cancer as a classification task and describe the implementation of ANN and SVM approaches which are used for classifying BC. It was observed that the ANN technique is more efficient than the SVM technique in breast cancer diagnostics.

Therefore this paper aims to work in comparison with the performance of some classifiers such as LR, DT, SVM, RF, NB, KNN, and ANN to evaluate the efficiency and effectiveness of those algorithms in terms of accuracy.

III. METHODOLOGY

First, we use the WBCD dataset after cleaning it and making it suitable for use. Second, we create some diagrams to understand and visualize the connections between each feature and to know the most important features. Third, we can divide the data into training data by 75% and test data by 25% that is 426 cases of training and 143 cases of testing. Fourth, we build a machine learning model for each algorithm to then test the models and choose the one that achieves the most accuracy in predicting cancer. Finally, the sort of most cancer is expected and as compared to the real values of our check dataset.

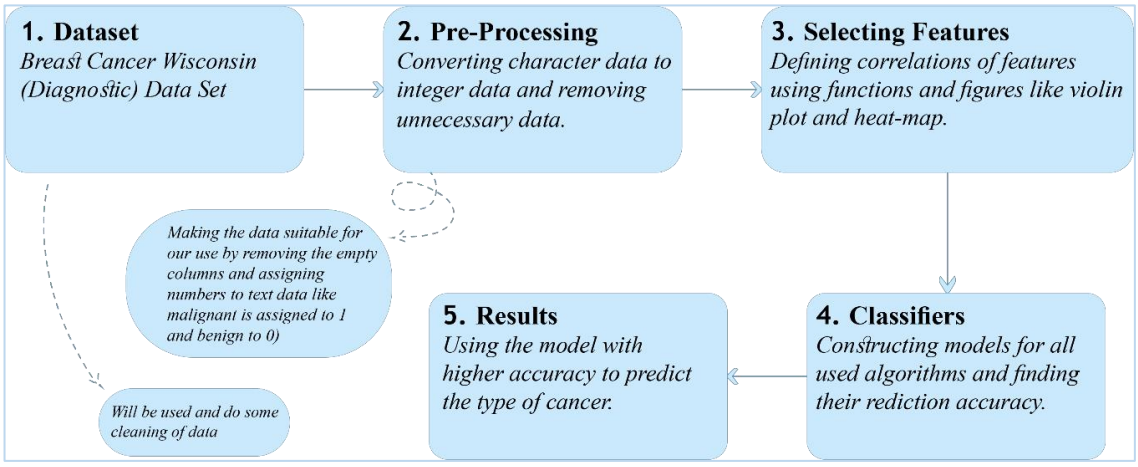

Fig. 1. Scheme of the model's steps

A. Dataset Description

To predict BC, we tend to use the (WBCD) collected from the UCI Machine Learning Repository [8]. In this dataset, there are 569 instances, 357 of whom are benign and 212 are malignant breast cancers respectively as proven in Fig. 2.

It contains 32 columns including the outcome column representing the diagnosis of breast cancer, 0 benign and 1 malignant.

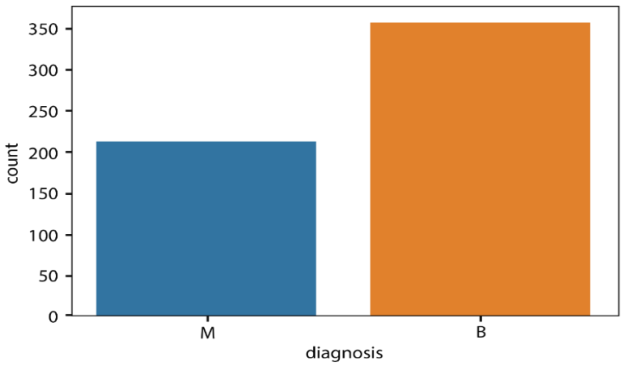

Fig. 2. Number of Benign and Malignant records

B. Data Pre-processing

We have made seven famous algorithms to evaluate the most accurate and useful in early disease detection to maintain human life. First, we collect and scrub the data by dropping empty instances in the features that are not relevant to breast cancer prediction. Also, we can remedy the missing values by finding the average/mean values of the features in our dataset. Then the data is represented through a Heat map to understand the association between the features with each other and the relationship of each feature to the diagnosis of breast cancer.

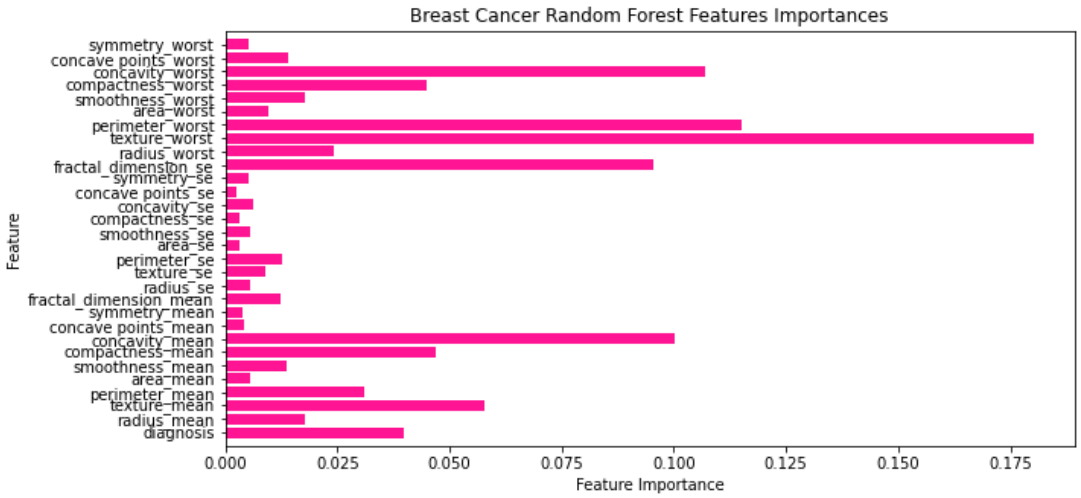

Fig. 3. Scheme of Feature Importance

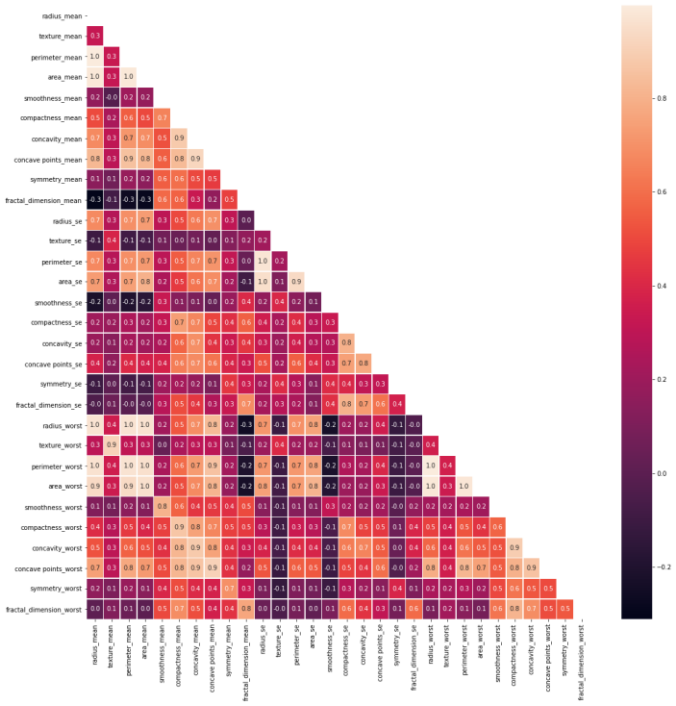

Fig. 4. Heat map

C. Train and Test

Our WBCD data is studied by CSV file, after it is prepared, it is completely free of missing values, we divided the data set into two parts, part for training 75% and part for testing 25%.
According to the step of data preparation, we chose the best features and dropped the least ones associated with the diagnosis of breast cancer to increase the efficiency of the prediction accuracy. After this step, the data is ready to be applied to our machine learning model to test the performance of the algorithms used in this paper.
Next, we completed the analysis of the performance of all algorithms by comparing the results of testing accuracy.

Violin Plot of diagnosis with other features depicting the amount of benign and malignant cancer cells and their relationship with the features, violin plot displaying all the mean features. The image below shows normal (Benign) and affected (Malignant) cells from our data set. At last, we build our model which contains the machine learning algorithms that we use to predict breast cancer.

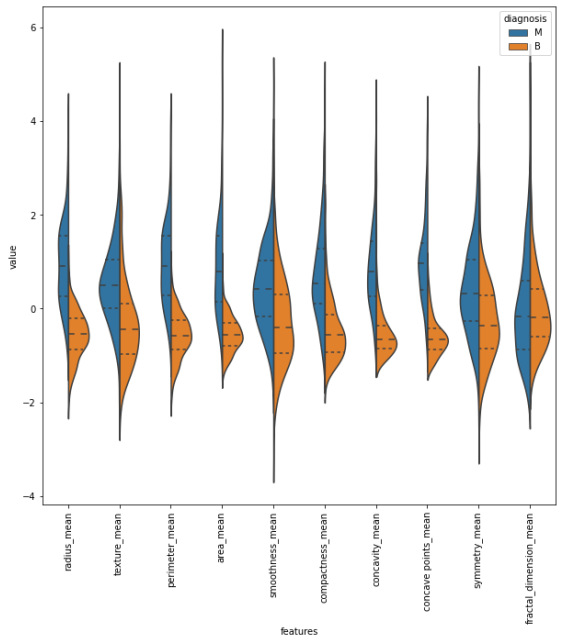

Fig. 5. Violin Plot Diagram

D. Classifiers

Logistic regression: The LR is a classification algorithm used to assign observations to a discrete set of classes. Logistic regression transforms its output using the logistic sigmoid function to return a probability value, it is a predictive analysis algorithm based on the concept of probability.

We may begin by assumptive p(x) be the linear function. However, the matter is that p is the probability that ought to vary from zero to one whereas p(x) is the associate degree infinite equation. to handle this downside, allow us to assume, that log p(x) could be a linear function of x, and in addition, to certain it between a variety of (0,1), we'll use logit transformation. Therefore, we'll take into account log p(x)/(1-p(x)). Next, we'll build this function to be linear:

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

After resolution for p(x):

$$p(x) = \frac{e^{\alpha_0 + \alpha}}{e^{\alpha_0 + \alpha} + 1}$$

To make the logistical regression a linear classifier, we tend to select an exact threshold, e.g. 0.5. Now, the misclassification rate is reduced if we tend to predict y=1 when $p \geq 0.5$ and y=0 when $p < 0.5$.

Here, one and zero are the categories. Since logistical regression predicts possibilities, we can match it to mistreatment probability. Therefore, for every coaching datum x, the anticipated category is y. The likelihood of y is either p if y=1 or 1-p if y=0. Now, the probability will be written as:

$$L(\alpha_0, \alpha) = \prod_{I=1}^{n} p(x_i)^{y_i}(1 - p(x_i)^{1-y_i}$$

The multiplication is often remodeled into a total by taking the log:

$$l(\alpha_0, \alpha) = \sum_{i=0}^{} \mathrm{y_i} \log p(x_i) + (1 - y_i) log 1 - p(x_i)$$

$$= \sum_{i=0}^{n} \log 1 - p(x_i) + \sum_{i=0}^{n} y_i log \frac{p(x_i)}{1 - p(x_i)}$$

Further, once swing the value of p(x):

$$l(\alpha_0, \alpha) = \sum_{i=0}^{n} - \log 1 + e^{\alpha_0 + \alpha} + \sum_{i=0}^{n} y_i(\alpha_0 + \alpha . x_i)$$

The next step is to require most of the on-top of probability function as a result of within the case of logistical regression gradient ascent is enforced (opposite of gradient descent).
n our experiment for this study, we used the default values of this algorithm to get accuracy [9]. We just changed the random_state parameter to 40 for initializing the internal random number generator.

Support Vector Machine: SVM has known a supervised machine learning algorithm rule which will be used for each classification and regression challenge. However, it is largely utilized in classification problems. within the SVM algorithm, we tend to plot each knowledge item to some extent in n-dimensional space (where n is many features you have) with the worth of every feature being the value of a specific coordinate. Next, we categorize by finding the super-level that differentiates the 2 categories (classes) well. We tend to use default values for all key parameters except the kernel parameter chosen linear, and random_state parameter we set the value to 40, to succeed in the required result [10].

Naïve Bayes: The NB classifier could be a probabilistic machine learning model that's used for classification tasks. The core of the classifier relies on Bayes Theorem. That is, the Naïve Bayes categorified assumes that the presence of a specific feature during a given class is unrelated to the existence of the other feature as mentioned within the following equation:

$$\mathrm{P(A\backslash B)} = \frac{\mathrm{P(B\backslash A)} * \mathrm{P(A)}}{P(B)}$$

Where P(A) is the previous Probability, P(B) is the Marginal Likelihood, P(B|A) is the Likelihood and P(A|B) is the Posterior Probability.

- **The math behind Naïve Bays Algorithm**

  Given a features vector X=(x1, x2, …, xn) and a category (class) variable y, Bayes Theorem states that:

$$P(y|X) = \frac{P(X|y) * P(y)}{P(X)}$$

We're fascinated by calculative the posterior probability P(y | X) from the likelihood P(X | y) and previous probabilities P(y), P(X). the exploitation of the chain rule, the likelihood P(X | y) may be rotten as:

$$P(X|y) = P(x_1, x_2, \ldots, x_n|y)$$

$$= P(x_1|x_2, \ldots, x_n, y) * P(x_2|x_3, \ldots, x_n, y) \ldots P(x_n|y)$$

but as a result of the Naïve's conditional independence assumption, the conditional possibilities are independent of every other.

$$P(X|y) = P(x_1|y) * P(x_2|y) \ldots . P(x_n|y)$$

Thus, by conditional independence, we have:

$$P(y|X) = \frac{P(x_1|y) * P(x_2|y) \ldots . P(x_n|y) * P(y)}{P(x_1) * P(x_2) \ldots . P(x_n)}$$

And because the divisor remains constant for all values, the posterior probability will then be:

$$P(y|x_1, x_2, \ldots, x_n) \propto P(y) \prod_{i=1}^{n} P(x_i|y)$$

The naïve Bayes algorithm combines this model with the concept of a decision rule. One common rule is to choose the foremost probable hypothesis; this is often called the utmost a posteriori or MAP decision rule. [11].

$$y = argmax_y P(y) \prod_{i=1}^{n} P(x_i|y)$$

Naïve Bayes (GaussianNB) parameters are priors (probabilities of the classes) and var_smoothing (part of the biggest contrast of all features) is not modified but the random_state parameter has been manipulated so we set it to 999 to get an acceptable result.

Random forest: The RF could be a supervised machine learning algorithmic program used wide in Classification and Regression issues. As its name suggests, a random forest consists of an oversized range of individual decision trees that act as a bunch. every tree within the random forest emerges from the category (class) prediction and therefore the class that gets the foremost votes becomes the prediction of our model. For the parameter values that we used in our experiment of the Random Forest algorithm, we set a value of 10 for the number of trees used for the n_estimators

parameter, chose the criterion “entropy” for the criterion parameter, and random_state equal 40 to get the desired result which we can see in Table III.

- **Problems Algorithm**
  When implementing random forests based on classification data, we should know that we often use a Gini index or the formula used to determine how to nodes in a branch of a decision tree.

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

  This formula uses class and probability in order to identify each Gini branch on a node and to identify which branches are likely to occur. Here, pi represents the relative frequency of the category we observe in the data set, and c represents the number of categories. We can also use entropy to determine how the nodes in the decision tree branch.

$$Entropy = \sum_{i=1}^{C} -p_i * \log_2(p_i)$$

  Entropy uses the probability of a given result of how the node is branching. But because of the logarithmic function used in calculating the Gini index, it is more mathematically intensive.

Decision Tree: The DT is the most powerful and popular tool for classification and prediction. It is a tool that has applications that cover several different areas. As its name suggests, it uses a tree-like flow diagram to indicate predictions that result from a series of feature-based splits. It begins with a root node and ends with a choice created by the leaves. We set the “entropy” criterion for the information gain as a function of measuring split quality and set random_state to 40 [10].

K-Nearest Neighbor: The KNN algorithmic program falls underneath the supervised Learning class and is employed for classification (most commonly) and regression. it's accustomed calculate missing values and reconstruct datasets. because the name suggests (K-Nearest Neighbor) it considers K the nearest neighbor (Datapoint) to predict the category (class) or continuous worth of the new Datapoint. We sufficed with implementing the default values for all parameters of this algorithm, due to the convergence of the results with each other and the appearance of slight differences in the test result.

For distance metrics, we will use the Euclidean metric.

$$d(x, x') = \sqrt{(x_1 - x_1')^2 + \ldots + (x_n - x_n')^2}$$

Finally, the input x gets assigned to the class with the largest probability.

$$P(y = j|X = x) = \frac{1}{K} \sum_{i \in \mathcal{A}} I(y^{(i)} = j)$$

For regression, the technique will be the same, instead of neighbor classes, we will take the target value and find the target value of the invisible data point by taking the mean, average, or any suitable function you want [13].

Artificial Neural Network: The ANN may be a cluster of algorithms that certify the underlying relationship during a set of knowledge almost like the human brain. The neural network helps to vary the input because the network provides the most effective doable result while not redesigning the output procedure.
ANN is formed of 3 layers specifically the input layer, an output layer, and hidden layer/s. There should be an affiliation between the nodes within the input layer with the nodes in the hidden layer and every hidden layer node with the nodes of the output layer. The input layer takes knowledge from the network. Then the hidden layer receives the raw information from the input layer and methods it. Then, the obtained value is sent to the output layer that successively will process the data from the hidden layer and provides the output. We will divide the interconnectivity of nodes between totally different layers into 2 main categories, namely, feedforward neural network and recurrent neural network. within the feedforward ANN, the movement of data from input to output is simply in one direction [6].

We used 30 neurons as the input layer as a count for all columns of our data set and 2 hidden layers for each layer of 16 neurons. The activation function “relu” is set in the case of the input layer and the first hidden layer, and we assign the same function to the second hidden layer while we assign a different activation function which is “sigmoid” to the output layer. At the end of our design, we used a single neuron to predict whether or not a patient had breast cancer. Figure 6 is just imaginary to illustrate and visualize the structure of the neural network components.

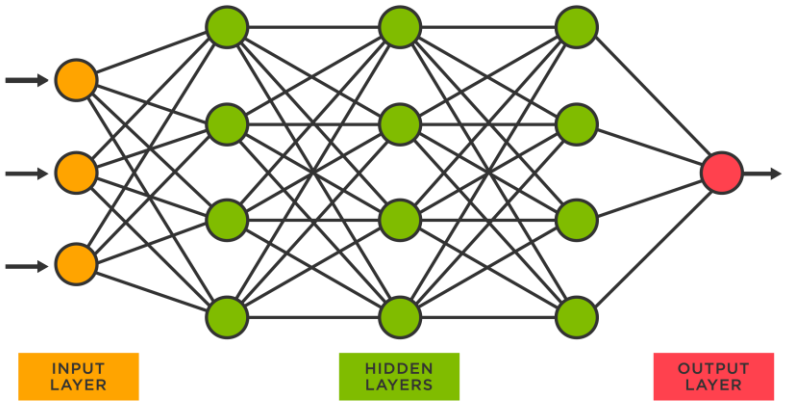

Fig. 6. Neural Network Diagram

Algorithm	Hyper-parameter	Value used
Logistic regression	random_state	40
Support Vector Machine	random_state	40
Naïve Bayes	GaussianNB	probabilities of the classes
	var_smoothing	part of the biggest contrast of all features
	random_state	999
Random forest	n_estimators	10
	random_state	40
Decision Tree	random_state	40
K-Nearest Neighbor	Default values	Due to the convergence of the results with each other
Artificial Neural Network	Input layer	30 Neurons
	Hidden layers	2 Layers for each layer of 16 neurons
	Output layer	1 Neuron

Table I. Hyper-parameter settings of the algorithms

IV. EXPERIMENTS RESULTS

A. System Specification

The model was trained on several devices with medium and high specifications in terms of CPU, RAM, and memory specifications (86% CPU), 492.7 MB / MAX 16 GB, the type of computer was (Laptop - Dell). The scientific computing platform is Kaggle, Jupyter Notebook, and the Python programming language while the associated libraries are (Scikit-Learn, Numpy, Pandas, Matplotlib, Seaborn, Missingno, and Warnings) and the accuracy was slightly different. The ANACONDA environment includes many applications such as the Spyder editor which was used in the work.

B. Results and Discussion

One of the lethal illnesses affecting ladies is breast cancer. In our work, the Wisconsin Breast Cancer Dataset changed into applied and numerous ML algorithms had been implemented to assimilate the efficacy and value of those algorithms to locate the best accuracy in classifying malignant and benign breast cancer. The correlation among specific functions of the dataset has been analyzed for characteristic selection.

Confusion matrix, sensitivity, specificity, ROC area (AUC), and accuracy metrics were used to measure the classification success of the methods. The following Equations show how these metrics are obtained. The confusion matrix is the matrix that represents the actual classes with the classes that are estimated in a classification system. Table II shows this matrix;

		Predicted Class	
		Positive	**Negative**
Factual Class	**Positive**	TF	FN
	Negative	FP	TN

Table II. Confusion matrices

- True Positive (TP): Data that is sick (Patient).
- True Negative (TN): Data that is not sick (Non-patient).
- False Positive (FP): Data that is sick and labeled as (Non-patient).
- False Negative (FN): Data that is not sick and labeled as (Patients).

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

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

The ROC value scale is plotted according to the true positive rate (TPR) and false positive rate (FPR). TPR is synonymous with sensitivity in the sensitivity equation as shown above. FPR is 1-specificity. The ROC curve for TPR and FPR values for different classification methods is plotted as in Fig. 7. The area below the blue line shows the area under the ROC curve value (AUC). The AUC is an effective, pooled measure of sensitivity and specificity that describes the inherent validity of diagnostic tests.

Our results showed the RF has the strongest technical predictor for breast cancer diagnosis with the RFC model having an accuracy of 98.23%, a sensitivity of 95.24%, a specificity of 100.00%, and an area under the curve (AUC) of 98%.

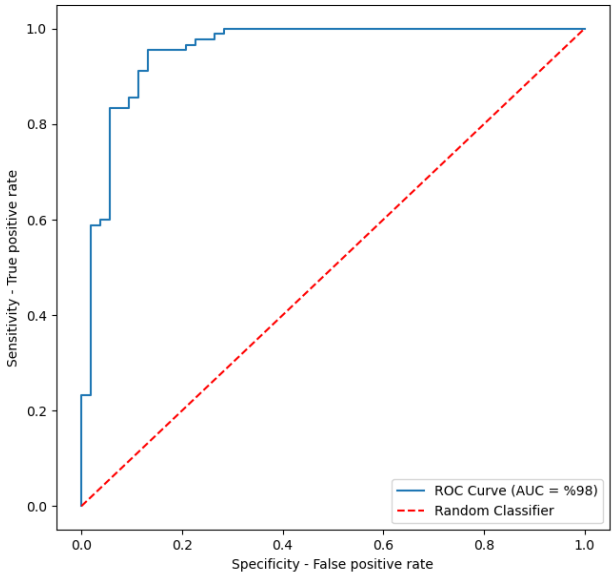

Fig. 7. ROC-AUC curve for Random Forest Classifier

The results helped select the best ML algorithm to build an automated breast cancer diagnosis system. The training and test values for the algorithms used are shown in Table III.

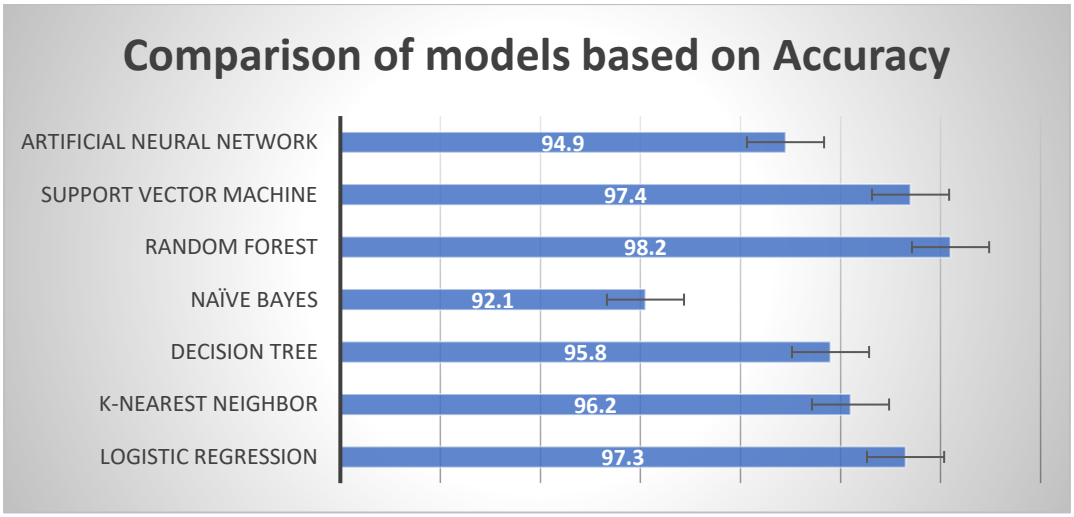

Table III. Comparison among Various Algorithms

We are aware that the most accurate is performed the usage of Random Forest with an accuracy of 98.2%, exceeding all other algorithms in this study. This indicates that Random Forest is the classifier of choice for the prediction and diagnosis of breast cancer pathological conditions.

C. CONCLUSION AND FUTURE WORK

Several machine learning algorithms were implemented to predict breast cancer using a public dataset namely the WBCD dataset. These algorithms namely, LR, DT, SVM, RF, NB, and KNN. When evaluating the effectiveness and efficacy of these algorithms in terms of different measures of accuracy, we can say that Random Forest has the most accuracy with an accuracy of 98.2%. This algorithm may be used to construct an automated diagnostic machine to predict breast cancer. In future work, we seek to deal with a relatively big dataset, optimize hyperparameters machine learning algorithms, and incorporate a few extra capabilities including breast most cancers segment detection and so on. We desire that this study will contribute to breast cancer treatment.

REFERENCES

[1] *Preventing cancer*. (n.d.). Retrieved March 13, 2022, from https://www.who.int/activities/preventing-cancer

[2] Sengar, P. P., Gaikwad, M. J., & Nagdive, A. S. (2020). Comparative study of machine learning algorithms for breast cancer prediction. *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020*, 796–801. https://doi.org/10.1109/ICSSIT48917.2020.9214267

[3] Thomas, T., Pradhan, N., & Dhaka, V. S. (2020). Comparative Analysis to Predict Breast Cancer using Machine Learning Algorithms: A Survey. *Proceedings of the 5th International Conference on Inventive Computation Technologies, ICICT 2020*, 192–196. https://doi.org/10.1109/ICICT48043.2020.9112464

[4] Ara, S., Das, A., & Dey, A. (2021). Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms. *2021 International Conference on Artificial Intelligence, ICAI 2021*, 97–101. https://doi.org/10.1109/ICAI52203.2021.9445249

[5] Breast Cancer Prediction using Machine Learning. (2019). *International Journal of Recent Technology and Engineering*, *8*(4), 4879–4881. https://doi.org/10.35940/ijrte.D8292.118419

[6] Imran, M., & Alsuhaibani, S. A. (2019). A neuro-fuzzy inference model for diabetic retinopathy classification. In *Intelligent Data Analysis for Biomedical Applications: Challenges and Solutions*. https://doi.org/10.1016/B978-0-12-815553-0.00007-0

[7] Divyavani, M., Kalpana, G., & Research Scholar, P. D. (2021). *AN ANALYSIS ON SVM & ANN USING BREAST CANCER DATASET*. https://www.researchgate.net/publication/348869189

[8] *UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set*. (n.d.). Retrieved March 14, 2022, from https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic)

[9] *1. Supervised learning — scikit-learn 1.0.2 documentation*. (n.d.). Retrieved March 23, 2022, from https://scikit-learn.org/stable/supervised_learning.html#supervised-learning

[10] Alhaj, Y. A., Dahou, A., Al-qaness, M. A. A., Abualigah, L., Abbasi, A. A., Almaweri, N. A. O., Elaziz, M. A., & Damaševičius, R. (2022). A Novel Text Classification Technique Using Improved Particle Swarm Optimization: A Case Study of Arabic Language. *Future Internet 2022, Vol. 14, Page 194*, *14*(7), 194. https://doi.org/10.3390/FI14070194

[11] *Naïve Bayes Algorithm. Exploring Naive Bayes: Mathematics, How… | by Bassant Gamal | Analytics Vidhya | Medium*. (n.d.). Retrieved July 27, 2022, from https://medium.com/analytics-vidhya/na%C3%AFve-bayes-algorithm-5bf31e9032a2

[12] *The math behind Logistic Regression | by Khushwant Rai | Analytics Vidhya | Medium*. (n.d.). Retrieved July 27, 2022, from https://medium.com/analytics-vidhya/the-math-behind-logistic-regression-c2f04ca27bca

[13] *The most insightful stories about Machine Learning - Medium*. (n.d.). Retrieved July 27, 2022, from https://medium.com/tag/machine-learning