



Journal of Applied Science

Biannual Peer Reviewed Journal Issued by Research
and Consultation Center , Sabratha University

Issue (12)
April 2024





Journal of Applied Science

Biannual Peer Reviewed Journal Issued by Research and Consultation Center,
Sabratha University

Editor

Dr. Hassan M. Abdalla

Associate Editors

Dr. Mabruk M. Abogderah

Dr. Jbireal M. Jbireal

Dr. Ali K. Muftah

Dr. Esam A. Elhefian

English Language Reviewer

Dr. Siham S. Abdelrahman

Arabic Language Reviewer

Dr. Ebrahim K. Altwade

Designed By

Anesa M. Al-najeh

Editorial

We start this pioneering work, which do not seek perfection as much as aiming to provide a scientific window that opens a wide area for all the distinctive pens, both in the University of Sabratha or in other universities and research centers. This emerging scientific journal seeks to be a strong link to publish and disseminate the contributions of researchers and specialists in the fields of applied science from the results of their scientific research, to find their way to every interested reader, to share ideas, and to refine the hidden scientific talent, which is rich in educational institutions. No wonder that science is found only to be disseminated, to be heard, to be understood clearly in every time and place, and to extend the benefits of its applications to all, which is the main role of the University and its scholars and specialists. In this regard, the idea of issuing this scientific journal was the publication of the results of scientific research in the fields of applied science from medicine, engineering and basic sciences, and to be another building block of Sabratha University, which is distinguished among its peers from the old universities.

As the first issue of this journal, which is marked by the Journal of Applied Science, the editorial board considered it to be distinguished in content, format, text and appearance, in a manner worthy of all the level of its distinguished authors and readers.

In conclusion, we would like to thank all those who contributed to bring out this effort to the public. Those who lit a candle in the way of science which is paved by humans since the dawn of creation with their ambitions, sacrifices and struggle in order to reach the truth transmitted by God in the universe. Hence, no other means for the humankind to reach any goals except through research, inquiry, reasoning and comparison.

Editorial Committee

Notice

The articles published in this journal reflect the opinions of their authors only. They are solely bearing the legal and moral responsibility for their ideas and opinions. The journal is not held responsible for any of that.

Publications are arranged according to technical considerations, which do not reflect the value of such articles or the level of their authors.


Journal Address:

Center for Research and Consultations, Sabratha University

Website: <https://jas.sabu.edu.ly/index.php/asjsu>

Email: jas@sabu.edu.ly

Local Registration No. (435/2018)

ISSN  2708-7301

ISSN  2708-7298

Publication instructions

The journal publishes high quality original researches in the fields of Pure Science, Engineering and Medicine. The papers can be submitted in English or Arabic language through the Journal email (jas@sabu.edu.ly) or CD. The article field should be specified and should not exceed 15 pages in single column.

All submitted research manuscripts must follow the following pattern:

- Title, max. 120 characters.
- Author Name, Affiliation and Email
- Abstract, max. 200 words.
- Keywords, max. 5 words.
- Introduction.
- Methodology.
- Results and Discussion.
- Conclusion.
- Acknowledgments (optional).
- References.

Writing Instructions:

Papers are to be submitted in A4 (200×285 mm) with margins of 25 mm all sides except the left side, which should be 30 mm. Line spacing, should also be 1.15.

Table 1. Font size and style

	Bold	English	Arabic
Font Style	✓	Times New Roman	Simplified Arabic
Article Title	✓	14 Capital	16
Authors Name	✓	12	14
Affiliation	×	11	13
Titles	✓	12	14
Sub-Title	✓	12	13
Text	×	12	14
Figure Title	✓	11	13
Table Title	✓	11	13
Equations	✓	12	14

Figures:

All figures should be compatible with Microsoft Word with serial numerals. Leave a space between figures or tables and text.

References:

The references should be cited as Harvard method, eg. Smith, R. (2006). References should be listed as follows:

Articles: Author(s) name, Year, Article Title, Journal Name, Volume and Pages.

Books: Author(s) name. Year. "Book title" Location: publishing company, pp.

Conference Proceedings Articles: Author(s) name. Year. "Article title". Conference proceedings. pp.

Theses: Author(s) name. Year. "Title". Degree level, School and Location.

Invitation

The Editorial Committee invites all researchers "Lectures, Students, Engineers at Industrial Fields" to submit their research work to be published in the Journal. The main fields targeted by the Journal are:

- Basic Science.
- Medical Science & Technology.
- Engineering.

Refereeing

The Editorial Committee delivers researches to two specialized referees, in case of different opinions of arbitrators the research will be delivered to a third referee.

Editorial Committee

Dr. Hassan M. Abdalla.
Dr. Mabruk M. Abogderah.
Dr. Jbireal M. Jbireal.
Dr. Ali K. Muftah.
Dr. Esam A. Elhefian.
Dr. Siham S. Abdelrahman.
Dr. Ebrahim K. Altwade.
Anesa M. Al-najeh.

CONTENTS

[1]	STUDY OF THE ACTIVE INGREDIENT OF FOUR DIFFERENT BRANDS OF COMMERCIAL DICLOFENAC SODIUM OF SELECTED PHARMACIES IN THE WESTERN REGION OF LIBYA.....	1
[2]	COMPARING SOLVING LINEAR PROGRAMMING PROBLEMS WITH APPLICATIONS OF THE MOORE-PENROSE GENERALIZED INVERSE TO LINEAR SYSTEMS OF ALGEBRAIC EQUATIONS	9
[3]	NITRATES IN MAN-MADE RIVER COMPARED TO GROUNDWATER WELLS IN EL-KUFRA AREA.....	19
[4]	DEGRADATION OF REACTIVE BLACK 5 DYE IN WATER FALLING FILM DIELECTRIC BARRIER DISCHARGE REACTOR (DBD).....	32
[5]	ANTIMICROBIAL RESISTANCE IN UROPATHOGEN ISOLATES FROM PATIENTS WITH URINARY TRACT INFECTIONS	44
[6]	SMOOTHING EFFECTS FOR A MODEL OF QUASI GEOSTROPHIC EQUATION.....	53
[7]	CLASSIFICATION OF BREAST CANCER USING MACHINE LEARNING ALGORITHMS.....	60
[8]	ENERGY-EFFICIENT INTRUSION DETECTION IN WSN: LEVERAGING IK-ECC AND SA-BILSTM	72
[9]	THE EFFECT OF ADDING POLYMER COMPOUNDS TO METALS ON ITS MECHANICAL PROPERTIES.....	89
[10]	EXPLORING DEADLOCK DETECTION ALGORITHMS IN CONCURRENT PROGRAMMING: A COMPARATIVE ANALYSIS AND EVALUATION	106
[11]	ISOLATION AND DETECTION OF BACTERIAL SPECIES CAUSING GINGIVITIS IN PATIENTS WITH TYPE 2 DIABETES.....	118

CLASSIFICATION OF BREAST CANCER USING MACHINE LEARNING ALGORITHMS

Rebah D. Sarreb

Dept. of Computer Engineering, Faculty of Engineering Regdalin, Sabratha University
Rebahdaw2018@gmail.com

Abstract

Breast Cancer is a common disease in women. Its early detection and classification using Machine Learning (ML) algorithms can effectively improve the patient's survival. In this study six Machine Learning Algorithms were applied to the dataset of Wisconsin Diagnostic Breast Cancer (WDBC) these are, Naïve Bayes (NB), K Nearest Neighbor (KNN), Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT). The data were preprocessed and split according to their size into 0.2 and 0.3. Each algorithm was tested and evaluated using performance metrics to classify the breast cancer into Malignant or Bingen tumors. Finally, the results of the algorithms are compared to each other. The results revealed that the LR and SVM achieved high accuracy (98%) at the size of the data test 0.3, whereas the lowest accuracy was NB (92%). Reducing the size of the data test to 0.2 improved the accuracy of all algorithms except DT, The LR had the best accuracy (99%). All the work was done in the Spyder environment based on Python programming language.

Keywords: Breast cancer; classification; accuracy; machine learning algorithm.

Introduction

Breast cancer is the most common disease among women, it is a major health issue globally affecting millions of women (Shruthi et al, 2020), and it usually leads to death if not diagnosed early. According to the World Health Organization (WHO), early diagnosis and accurate classification of breast cancer can improve patient survival rates (Mashudi et al, 2021). Generally, breast cancer is a result of a mutation in a single cell that divides irregularly, forming a mass in breast tissue known as a tumor (Mashudi et al, 2021). The tumors are classified into two types- malignant and benign tumors. The benign masse does not expand to other tissue, it can be removed by medical treatment (Ara et al, 2021). Whereas, the mass of a malignant tumor expands to other tissue and spreads at a higher rate (Singh and Raj, 2021). It is essentially to differentiate between the two types of tumors. However, the detection of breast cancer may be difficult at the beginning, due to the absence of symptoms, so a reliable diagnostic technique system is required to distinguish between them (Ara et al, 2021).

Machine learning algorithms have emerged as powerful tools for the classification and prediction of breast cancer, these algorithms have shown great potential for diagnosing and predicting different cancer diseases (Ozcan et al, 2022).

Numerous previous studies have been done in the field of breast cancer using machine learning algorithms. (Bokhare and Jha, 2023) Classified the breast cancer using K-nearest neighbor KNN, Support Vector Machine SVM, Random Forest RF, kernel SVM, Logistic Regression LR, and Naïve Bayes NB algorithms, its accuracy was evaluated using a Benchmark. That indicates the DT algorithm achieved the best performance. (Chen et al, 2023) Applied machine learning algorithms for early diagnosis of breast cancer using XGBoost, Logistic Regression LR, Random Forest RF, and K-nearest neighbor algorithms, the algorithms are evaluated by precision, accuracy, F1 score, and recall. The result revealed the XGBoost algorithm gives the best performance. (Kiran et al, 2023) Compared the accuracy between the Light GBM, Logistic Regression LR, Gradient Boosting, Random Forest RF, and the XGBoost algorithms, the Light GBM achieved the highest accuracy of 97.07%.

(Zhang and Li, 2022) compared the performance of five machine learning algorithms, namely Logistic Regression LR, Random Forest RF, K-nearest neighbor KNN, Support Vector Machine SVM, and Decision Tree DT for breast cancer prediction and classification using (WBCD), the performance of these algorithms was compared in terms of accuracy, F1 score, ROC curve, and PR curve. The LR in the study gives the best performance than others. (Ozcan et al, 2022) Classified and compared the performance of breast cancer using Support Vector Machine SVM, Naïve Bayes NB, Random Forest RF, Decision Tree DT, K Nearest Neighbor KNN, Logistic Regression LR, Multilayer Perceptron MLP, Linear Discriminant Analysis LDA, XGBoost XGB, Ada Boost ABC, and Gradient Boosting GBC machine learning algorithms. The GBC algorithm gave the highest accuracy of 99.12%. (Mayce et al, 2021) Compared three machine learning algorithms, Support Vector Machine SVM, Logistic Regression LR, and Neural Network (NN) to classify benign and malignant breast cancer. The performance of the algorithms was evaluated using the K-fold cross-validation technique and confusion matrix, the NN algorithm achieved the highest accuracy of 99.4%.

(Singh and Raj, 2021) Studied the performance of Support Vector Machine SVM, Logistic Regression LR, and Naïve Bayes NB algorithms. RF algorithm gave the highest accuracy rate (99.76%), with least error rate. (Houfani et al, 2020) Applied kernel and linear Support Vector Machines, Random Forest, Decision Tree, Multilayer perceptron, Logistic Regression, and K-nearest neighbors for breast cancer tumor classification. The Multilayer perceptron and Logistic Regression gave the best accuracy of 98% for breast cancer classification. (Parhusip et al, 2020) Compared and investigated different machine learning algorithms using K-nearest neighbor KNN, Support Vector Machine SVM, Random Forest RF, Logistic Regression LR, Naïve

Bayes NB algorithms, and Decision Tree DT. The data splitting into test sizes 0.3 and 0.25, by reducing the test data from 0.3 to 0.25 the accuracy improved in all algorithms except NB and the best performance given by RF. (Shruthi et al, 2020) Evaluated the performance of different machine learning algorithms for breast cancer prediction and classification using the (WBCD). Logistic Regression LR, Random Forest RF, K-nearest neighbor KNN, Support Vector Machine SVM, and Decision Tree algorithms, were implemented and their accuracies were measured. The RF achieved the highest accuracy of 97%, followed by the LR (95%), KNN (95%), DT (94%) and SVM (93%).

The results of algorithms that have been applied in this paper will be compared to predict and classify malignant and benign breast tumors in terms of accuracy, confusion matrix, precision, F1 score, and recall.

Material and Method

The main objective of this study is to identify an effective and predictive algorithm for classifying breast cancer by applying the following machine learning algorithms; Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), Decision Tree (DT), K Nearest Neighbor (KNN), and Logistic Regression (LR) on Wisconsin Diagnostic Breast Cancer (WDBC) dataset. The results obtained will be evaluated by performance metrics; accuracy, confusion matrix, precision, F1 score, and recall. The proposed architecture is detailed in Our methodology begins with data acquisition followed by pre-processing of the data. It contains four steps these are data cleaning, removing the null values, setting the target, and feature selection to obtain the optimal subset and reduce the redundancy that improves the accuracy of algorithms (Chen et al, 2023). The data was normalized from 0 to 1. Then it split into train data and test data with data test sizes 0.2 and 0.3. The data was trained by different algorithms and tested according to its size. Finally, the results of each algorithm are evaluated in terms of accuracy and performance metrics to determine the algorithm with the highest accuracy.

Dataset

The dataset used in this study was released from the Wisconsin Diagnostic Breast Cancer (WDBC), that obtained from the Kaggle dataset. The dataset consists of 569 patients and 30 integer-valued features, 10 of the 30 features are measured in the tumor cell. Feature selection is a process of removing irrelevant features or cleaning data from noise. This process directly affects classification success and performance. The features of the dataset have the properties as follows: radius, texture, perimeter, area, smoothness, compactness, concavity, concave, symmetry, and fractal dimensions. Each feature has a mean, standard error, and worst, which results in 30 features. The response class is divided into; benign (B) tumors with 357 instances and malignant (M) tumors with 212 instances (Mashudi et al, 2021).

The Experiment Environment

The algorithms are programmed in Python using the scikit learn library in the Spyder environment's statical data in this study is handled by using libraries such as Seaborn, NumPy, matplotlib, and Pandas.

Machine Learning Algorithms

It is a type of artificial intelligence that includes various statistical, probability analysis, and optimization techniques. It easily detects and classifies complex and large datasets. There are different ML algorithms and methods used in data analysis such as K Nearest Neighbor KNN, Decision Tree DT, Support Vector Machine SVM, Random Forest RF, Logistic Regression LR, and Naive Bayes NB etc. (Ozcan et al,2022).

Support Vector Machine (SVM)

Is used to classify two classes of classification problems by creating the best hyperplane that spreads the data. It was a set of mathematical functions to transfer the data to the desired form. The SVM kernel function is used to measure the distance between two data points and to choose the support vector to classify the data accurately. It is often used for binary classification problems and works precisely with high-dimensional data (Mashudi et al, 2021). SVM has four kernel functions, RPF, Linear, Sigmoid, and Polynomial (Zhang and Li, 2022). In this study, RPF Kernel has been used.

Decision Tree (DT)

Is used for both classification and regression algorithms. It is a tree-like structure starting with a root node and dividing into different branches based on the values of input features. The data is split into different branches, the process continues until reaches the leaf node, which represents the find decision or classification. DT algorithm analyzes the characteristics of breast tumors and classifies them as benign or malignant based on a set of rules derived from the training data (Bokhare and Jha, 2023).

Random Forest (RF)

Is a collection of decision trees that creates and combines multiple predictors to produce the final model. The predictors are typically decision-trained using the beginning method to build the average of numerous noises to reduce the variation (Chtouki et al, 2023), it classifies the data from the root node to offspring nodes to maintain similarity and consistency (Bokhare and Jha, 2023).

K Nearest Neighbor (KNN)

KNN is used to find the Nearest neighbors of the data in the dataset. It helps to classify a new data point to the nearest class (Bokhare and Jha, 2023). The KNN algorithm should be run several times with different K values to choose the best K value and reduce the number of errors. (Mashudi et al, 2021).

Logistic Regression (LR)

It is used to predict and classify categorical dependent values. It works by fitting a regression line of the data. This approach is based on the sigmoid function (S-shaped curve) (Bokhare and Jha, 2023), used to give real values between 0 to 1 representing the probability of the data that has been classified as either benign or malignant (Parhusip et al, 2020). The formula of Logistic Regression is:

$$F(X=1) = 1 / (1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p)}) \dots \dots \dots (1)$$

$F(X=1)$ is the chance of survival of the cancer patient, and the regression coefficients are b_0, b_1, \dots, b_p . (Chtouki et al, 2023). In this paper, Logistic Regression will be used to classify breast cancer into malignant or benign tumors based on input features. The Logistic Regression algorithm was applied to classify and evaluate breast cancer using metrics such as accuracy, F1 score, and specificity.

Naïve Bayes (NB)

This algorithm is based on the Bayesian theorem. The context of the Naïve Bayes Classifier (NBC) is assumed to be independent under certain conditions. NBC used data directly influence each other to determine the model, by utilizing known training compounds classified as active D and inactive H. The Bayes classifiers use the Bayes theorem, which is:

$$P(h|d) = p(d|h) * p(h)/p(d) \dots \dots \dots (2)$$

In Equation (2). $P(h)$ the prior probabilities of events h and the training data d , as well as the conditional probability of d given h . $P(d | h)$ represents the conditional probability of h given the training data d , while $P(h | d)$ denotes the probability of generating instance d for class (Amrane et al, 2018).

Performance Metrics

Confusion Matrix

Is the way to measure the performance of classified algorithms where the output is two or more types of classes. The confusion matrix is a table with two dimensions “Actual” and “prediction” (Singhal et al, 2022), that regulate the performance of the classification algorithm by comparing how many positive instances are true positive

TP, true negative TN, and how many negative instances are false positive FP, false negative FN (Shruthi et al, 2020).

Accuracy

Evaluate the classification algorithms. it is the fraction of prediction that determines the number of correct predictions to the total number of predictions (Singh and Raj, 2021). as in the following Equation.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \dots \dots \dots (3)$$

Precision

Represents how many of the positive predicted samples are positive samples (Shruthi et al, 2020). By using the following equation.

$$\text{Precision} = \frac{TP}{TP+FP} \dots \dots \dots (4)$$

Recall

It represents the true positive rate for the original samples that are predicted correctly by applying the following equation (Chen et al, 2023).

$$\text{Recall} = \frac{TP}{TP+FN} \dots \dots \dots (5)$$

TP- the number of True Positives classified

TN - the number of True Negatives classified.

FN -the number of False Negatives classified.

FP -the number of False Positives classified.

F1-Score

Obtained from the average of precision and recall, because of the contradiction between the two evaluation indexes, the higher value showed that the classification results are more effective. The equation applied for the index F1 score is (Chen et al, 2023).

$$\text{F1 Score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \dots \dots \dots (6)$$

Result and Discussion

In this work, six Machine Learning Algorithms are applied to the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. these are Naïve Bayes (NB), K Nearest Neighbor (KNN), Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT). The performance metrics that have been

used to evaluate and compare the algorithms to find the best algorithm for breast cancer classification are confusion matrix, accuracy, F1 score, recall, and precision.

Data Test Size (0.3)

The Comparison of train and test accuracy between the following algorithms with data test size 0.3 are; Naïve Bayes (NB), K Nearest Neighbor (KNN), Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT) show that the Logistic Regression and Support Vector Machine give the highest test accuracy (98%), while the random forest and Decision Tree have the highest train accuracy (100%) as shown in Table (1).

Table (1): Comparison of the Train and Test Accuracy for Applied Algorithms.

ML algorithms	Train Accuracy	Test Accuracy
Logistic Regression	99%	98%
Random Forest	100%	95%
Support Vector Machine	98%	98%
Decision Tree	100%	94%
K Nearest Neighbor	96%	94%
Naive Bayes	95%	92%

The comparison of the performance metrics between the applied algorithms at data test size (0.3) Table (2), revealed that the Logistic Regression and Support Vector Machine algorithms have the highest recall, accuracy, F1 score, and precision of (99%,98%,98%.97%) respectively. (Zhang, and Li, 2022) also found that at data test size 0.3, the Logistic Regression algorithm had the highest performance accuracy of 95%.

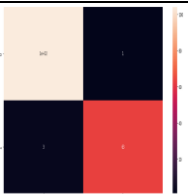
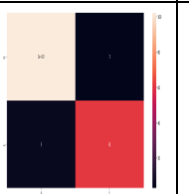
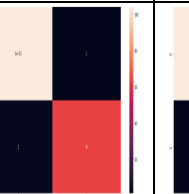
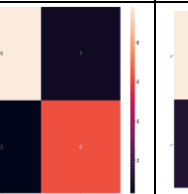
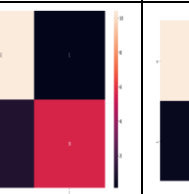
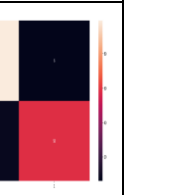
Table (2): Comparison of the Performance Metrics for the Applied Algorithms.

ML Algorithms	Malignant [1] Benign [0]		Precision	Recall	F1-Score	Accuracy	Support
Logistic Regression	104	1	0.97	0.99	0.98	98%	105
	3	63	0.98	0.95	0.97		66
Random Forest	102	3	0.95	0.97	0.96	95%	105
	5	61	0.95	0.92	0.94		66
Support Vector Machine	104	1	0.97	0.99	0.98	98%	105
	3	63	0.98	0.95	0.97		66
Decision Tree	98	7	0.97	0.93	0.95	94%	105
	3	63	0.90	0.95	0.93		66

K-nearest Neighbor	104	1	0.91	0.99	0.95	94%	105
	10	56	0.98	0.85	0.91		66
Naive Bayes	99	6	0.93	0.94	0.93	92%	105
	8	58	0.91	0.88	0.89		66

The comparison of the confusion Matrix of different algorithms at data size 0.3 indicates that the Logistic Regression and Support Vector Machine algorithms give the best accurate prediction of 105 Malignant cases 104 cases are correct and 1 case is incorrect. Whereas 66 benign cases 63 cases are correct and 3 cases are incorrect as shown in Table (3).

Table (3): Comparison of the Confusion Matrix for Applied Algorithms.

ML Algorithms	Logistic Regression	Random Forest	SVM	Decision tree	KNN	NB
Confusion Matrix	[104 1] [3 63]	[102 3] [5 61]	[104 1] [3 63]	[98 7] [3 63]	[104 1] [10 56]	[99 6] [8 58]
Heatmap of Confusion Matrix						

Data Test Size (0.2)

The Comparison of train accuracy and test accuracy for all the applied algorithms revealed that the maximum value of the accuracy test given by logistic regression of 99%, while the random forest and decision tree give the highest train accuracy of 100% Table (4).

Table (4): Comparison of Train and Test Accuracy of the Applied Algorithms.

ML algorithms	Train accuracy	Test accuracy
Logistic Regression	98%	99%
Random Forest	100%	97%
Support Vector Machine	98%	98%
Decision Tree	100%	93%
K Nearest neighbor	96%	95%
Naive Bayes	95%	94%

Comparing the results of performance metrics of the algorithms that have been used at data test size (0.2) shows that, the Logistic Regression gives the best performance of recall (100%), accuracy (99%), F1 score (99%), and precision 99% compared to other algorithms, and also has the highest performance for malignant and benign breast cancer classification Table (5). (Houfani et al, 2020) indicated that the Logistic Regression algorithm had the highest performance accuracy of 98%, at a data test size of 0.3.

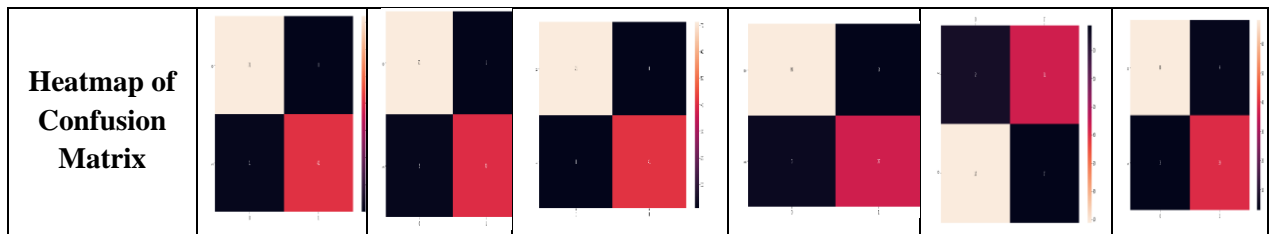
Table (5): Comparisons of the Performance Metrics of the Applied Algorithms.

ML Algorithms	Precision	Recall	F1-Score	Class	accuracy
Logistic Regression	0.99	1.00	0.99	Benign	99%
	1.00	0.98	0.99	Malignant	
Random Forest	0.97	0.99	0.98	Benign	97%
	0.98	0.95	0.96	Malignant	
Support Vector Machine	0.99	0.99	0.99	Benign	98%
	0.98	0.98	0.98	Malignant	
Decision Tree	0.93	0.96	0.95	Benign	93%
	0.93	0.88	0.90	Malignant	
K Nearest Neighbor	0.93	0.99	0.96	Benign	95%
	0.97	0.88	0.93	Malignant	
Naive Bayes	0.96	0.94	0.95	Benign	94%
	0.91	0.93	0.92	Malignant	

Comparisons of the results of the confusion matrix between the applied algorithms Table (6) indicated that the LR algorithm gives the best accurate prediction, from 72 malignant cases 72 cases are correct and 0 cases are incorrect, whereas of 42 benign cases, 41 cases are correct, and 1 is incorrect. (Houfani et al, 2020) revealed that LR obtained the best accurate prediction at data test size 0.25, of 54 Malignant cases, 53 cases were correct 1 case was incorrect, and 89 benign cases, 87 cases were correct and 2 cases were incorrect.

Table (6): Comparisons of the Confusion Matrix of the Applied Algorithms.

ML Algorithms	Logistic Regression	Random Forest	Support Vector Machine	decision tree	K-nearest neighbor	Naive Bayes
Confusion Matrix	[72 0] [1 41]	[71 1] [2 40]	[71 1] [1 41]	[69 3] [5 37]	[71 1] [5 37]	68 4] [3 39]



The Compression of the accuracy at data test sizes 0.2 and 0.3 using the six Machine learning algorithms Table (7) to predict and classify malignant and benign breast cancer, reveals that the accuracy of all algorithms except the DT algorithm is improved by reducing the data test size. also (Parhusip et al, 2020) found that all algorithms except the NB algorithm were improved by decreasing the data test size.

Table (7): The Accuracy of the Applied Algorithm at Different Data Test Sizes.

Machine learning Algorithms	Accuracy test size=0.2	Accuracy test size=0.3
Logistic Regression	99%	98%
Random Forest	97%	95%
Support Vector Machine	98%	98%
Decision Tree	93%	94%
K Nearest neighbor	95%	94%
Naive Bayes	94%	92%

Conclusion

Six machine learning algorithms are applied at data test sizes 0.2 and 0.3, these are Naïve Bayes (NB), K Nearest Neighbor (KNN), Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT). The results of the accuracy, F1 score, precision, recall, and confusion matrix are evaluated and compared to identify the best machine learning algorithms that are precise, reliable, and give the highest accuracy. At data test size 0.3 the LR and SVM achieved the highest accuracy of 98%, recall of 99%, F1 score of 98%, and precision of 97%. Whereas at data test size 0.2 the LR gives the highest recall of 100%, accuracy 99%, F1 score of 99%, and precision 99%. The compression between the algorithms used in this study at data test sizes 0.3 and 0.2 for the prediction and classification of malignant and benign breast cancer, revealed that the accuracy of all algorithms except the DT is improved by a reduction of data test size. Further work must apply these algorithms to

other databases to confirm our results, it also must be conducted by applying the machine learning algorithms used in this study and others by using different parameters with larger datasets of disease classes to improve the performance of accuracy.

Reference

- Amrane, M., Oukid, S., Gagaoua, I., and Ensar, T. (2018),” Breast Cancer Classification Using Machine Learning”. *International Journal of Advanced Technology and Engineering Exploration*, Vol. 8, pp; 2394-7454.
- Ara, S., Das, A., and Dey, A. (2021), “Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms”. *International Conference on Artificial Intelligence (ICAI)*.
- Bokhara., Jha, P. (2023).” Machine Learning Models Applied in Analyzing Breast Cancer Classification Accuracy”. *IAES International Journal of Artificial Intelligence*, Vol. 12, pp; 1370.
- Chen, H., Wang, N., Du, X., Mei, K., Zhou, Y., and Cai, G., (2023),” Classification Prediction of Breast Cancer Based on Machine Learning”. *Computational Intelligence and Neuroscience*.
- Chtouki, K., Rhanou, M.i, Mikram, M., Yousfi, S., and Amazian, K. (2023).” Supervised Machine Learning for Breast Cancer Risk Factor Analysis and Survival Prediction”.
- Houfani, D., Slatnia, S., Kazar, O., Zerhouni, N., Saouli, H.,and Remand, I. (2020). “Breast Cancer Classification using Machine Learning Techniques: A Comparative Study”. *Medical Technologies Journal*, Vol. 4, pp;535-544.
- Kiran R., Rajesh, T., Krishna, M., Gopal, N., and Kishan G. (2023).” Breast Cancer Classification Using Machine Learning”. *International Research Journal on Advanced Science Hub*. Vol. 05. pp;2582-4376.
- Mashudi, N., Rossi, S., Ahmad, N., and Noor, N. (2021),” Breast Cancer Classification: Features Investigation Using Machine Learning Approaches”. *International Journal of Integrated Engineering*. Vol. 13, pp;2229-8380.
- Mayce, K., Lomboy, R., Rowell M. and Hernandez (2021). “A Comparative Performance of Breast Cancer Classification using Hyper-Parameterized Machine Learning Models”.

- Ozcan, I., Aydin, H., and Cetinkaya, A. (2022). "Comparison of Classification Success Rates of Different Machine Learning Algorithms in the Diagnosis of Breast Cancer". *Asian Pacific Journal of Cancer Prevention*, Vol.23, pp; **3287- 3297**.
- Parhusip, H., Susanto, B., Linawati, L., Trihandaru, S., Sardjono, Y., and Mugirahayu, A., (2020)." Classification Breast Cancer Revisited with Machine Learning". *International Journal of Data Science*. Vol. 1, pp; 42-50.
- Shruthi, S., Binu, F., Ravi Kumar, A., Yeshwanth, S., and Mahalinga V., (2020)." Breast Cancer Classification using Python Programming in Machine Learning". *International Journal of Engineering Research & Technology*. Vol. 9, pp; 2278-0181.
- Singh, H., Raj, H. (2021). " Breast Cancer Analysis and Prediction by Using Machine Learning". *International Journal of Research in Engineering and Science*. Vol 9. pp; 69-73.
- Singhal, V., Chaudhary, Y., Verma, S., Agarwal, U., and Sharma, P. (2022)." Breast Cancer Prediction using KNN, SVM, Logistic Regression and Decision Tree". *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*. Vol 10. pp; 2321-9653.
- Zhang, Z., Li, Z. (2022)." Evaluation Methods for Breast Cancer Prediction in Machine Learning Field". *SHS Web of Conferences* 144.