How machine learning algorithms perform in breast cancer classification?

Md. Taslim[1], Md Shafiuzzaman[1], Md. Moradul Siddique[1], Md. Alam Hossain[1]*

1. *Department of Computer Science and Engineering, Jashore University of Science and Technology, Jashore, Bangladesh*

* **Corresponding Author:** Email: alam@just.edu.bd

Abstract

This paper experiments the performance of machine learning algorithms in breast cancer classification. There are several existing approaches that are intended to work on this motivation. However, no thorough evaluation exists, rather most of the works consider a small subset of algorithms with a single dataset. We come up with twelve off-the-shelf well-known algorithms, such as, Linear Discriminant Analysis (LDA), Logistic Regression (LR), Stochastic gradient descent Classifier (SGD), Decision Tree Classifier (DT), Support vector machine (SVM), K-Neighbors Classifier (KNN), Naive Bayes (NB), AdaBoost Classifier (AB), Random Forest Classifier(RF), Voting Classifier (VC), Bagging Classifier (BC), Gradient Boosting Classifier (GB) and explore their performance on two most popular breast cancer classification datasets, such as, Wisconsin Breast Cancer (Original) (WBC) dataset and the Wisconsin Breast Cancer Diagnosis (WBCD) dataset. In the testing phase, LR, SVM, SB, RF, VC, and BC achieved 100% accuracy on the WBC dataset, and NB, SVM, RF, and KNN achieved the highest accuracy with 96.49%. In addition, cross-validation accuracy of LR, SVM, LDA, KNN, AB, RF, VC, SGD, BC, and GB classifiers achieve the highest accuracy of 98.61%, 98%, 98.73%, 98.87%, 98.57%, 98.72%, 98.73%, 98.57%, 98.61%, 98.57% respectively in the testing phase on the WBC dataset and SGD, RF and LR achieved the highest accuracy with 98.87%, 96.67%, and 96.67% respectively in the testing phase on the WBCD dataset. The overall experiment signifies with a message that performance of the most of the algorithms can be improved with careful parameter tuning.

Key Words: Machine Learning, Breast Cancer, Classifier, Breast Cancer Diagnosis.

1. Introduction

According to the World Health Organization (WHO), about 9.6 million people died from cancer in 2018, and 70% of those happened in developing countries, where cancer diagnosis facilities are still scarce and expensive [1]. Among all the types of cancer, breast cancer is the most common [2]. Globally, it is expected to increase from 1.4 million in 2008 to more than 2.1 million by 2030 [27]. Every year, almost 1.5 million women are diagnosed with breast cancer [3]. Approximately 29.9% of cancer deaths in women are owing to breast cancer [4]. There are two types of breast cancer, such as benign and malignant. Benign represents the non-cancerous one with no threat to life, whereas the malignant one means the most cancerous one with a direct threat to life [5]. Identifying the type of breast cancer is imperative since it is the first step of the cancer diagnosis that determines the entire pathway of cancer treatment. Identifying these types through traditional laboratory-based methods such as physical syndromes, biopsies, and radiographic images requires extensive medical equipment and staff involvement [6]. The biopsy method is widespread; however, it highly depends on the doctor's expertise. Additionally, mammography, the standard diagnostic method for surgical biopsy, is not free from false positive results [7]. On the other hand, the effectiveness of diagnosis through radiographic images solely depends on the radiologists' explanation [8], and radiologists may miss up to 30% of the cases depending on the density of breasts [9]. So, the overall laboratory-based methods of breast cancer diagnosis are not free from errors, and at the same time, it is time-consuming and expensive.

Computer scientists have contributed several machine learning methods to automate breast cancer diagnosis. MF Aslan et al. [12] use four different Machine Learning (ML) algorithms to detect breast cancer, such as Artificial Neural Network (ANN), Extreme Learning Machines (ELM), SVM and KNN. The ELM achieved the highest accuracy (80%) on the Breast Cancer Coimbra dataset. In another study, Potdar et al. [20] used ANN, KNN, and Bayesian Classifiers to classify breast cancer. They use 3-fold cross validation to eliminate the imbalanced data problem. In 3-fold cross validation, ANN provides the highest accuracy of 97.4% accuracy on the WBCD dataset. Another ML technique for breast cancer classification is Convolutional Neural Networks (CNN). Y. J. Tan et al. [28] take the help of CNN to identify breast cancer through the Mammogram Imaging. They used three versions of CNN, where the first one uses

a raw classifier, and version 2 uses a TensorFlow-guided CNN model with raw input images. Version 3 also uses a TensorFlow-guided CNN model, but the input images are already preprocessed. In this study achieved the highest results of 82.71% of version 3. The result is obvious, and performance remains poor. Moreover, Agarap et al. [16] provide a comparison of six ML algorithms: GRUSVM [29], LR, MLP, Nearest Neighbor (NN) search, SoftMax Regression, and SVM on the WDBC dataset. Among them, the MLP algorithm achieved the highest accuracy (99.04%). Bayrak et al. [19] compared SVM and ANN models on WBC dataset, where SVM obtained the best result with 96.997% accuracy. Existing studies on breast cancer detection using ML techniques do not provide any clear winner. The majority of them report their analysis result depending on a single dataset in a single perspective. To conduct a thorough investigation, we use twelve well-known ML algorithms and report their performance in both WBC and WBCD datasets.

The rest of the paper is arranged as follows: Section II explains the literature review, section III describes the methodology of the comparative study, section IV shows the comparative results and discussions, section V compares with the previous study, and finally, section VI draws the conclusion.

2. LITERATURE REVIEW

This section provides an overview of significant studies conducted in breast cancer classification.

MM Islam et al. [10] use SVM and KNN on WBCD dataset with 10-fold cross-validation. They reported 98.57% accuracy for SVM and 97.14% for KNN which are significant. However, result on a single dataset without any justification does not seem to be impactful.

Besides SVM and KNN, MF Aslan et al. [12] added ANN and ELM to measure impact of ML algorithms on breast cancer classification. Their objective was to process the routine blood analysis results and determine how well these techniques work to find breast cancer. The blood analysis dataset was collected from the work of Patr´ıcio et al. [28]. From their reported result, we can see ELM methods obtained the best results (80%) with the shortest training time (0.0075s) compared to the other three. Despite the fact that this study uses a different dataset, the performance is in lower side considering the fact of safety critical nature of breast cancer classification.

AA Bataineh et al. [7] focuses on nonlinear ML methods and uses NB, Classification and Regression Trees (CART), MLP, KNN, and SVM on WBCD dataset. Finally, they evaluate the model performance concerning the effectiveness and efficiency of each algorithm in terms of precision, recall, and accuracy. The MLP model outperformed the others in terms of accuracy (96.70%), precision (100%), and recall (97%). They used 10-fold cross-validation for more accurate results.

S Sharma et al. [17] compare the performance of RF, KNN, and NB on the WBCD dataset. To compare the model performances, they have considered accuracy and precision. The KNN (95.90%) outperforms RF and NB. Furthermore, KNN had the highest precision (98.27%) and f1-Score (94.20%).

In another study, K Sivakami et al. [11] present a disease status prediction strategy. This strategy is divided into two parts. 1. Information Treatment and Option Extraction, and 2. DT-SVM Hybrid Model for predictions. They used Weka Software tools for data preparation, data analysis, and result comparison. In this study, three classification techniques are compared, and DT-SVM (91%) outperforms Instance-based Learning (IBL), Sequential Minimal Optimization (SMO), and NB classifiers. They also used the WBCD dataset from the UCI machine learning repository. However, their dataset distribution (60% training and 40% testing) is not perfect because the sample dataset is limited.

Mohammed et al. [15] use three different approaches that improve the accuracy and enhance the performance of three different classifiers: DT (J48), NB, and SMO. They conducted two different datasets, the WBC and the Breast Cancer dataset. The authors mainly focus on dealing with imbalanced data. Data imbalance is a big problem in the classification field. The resampling techniques are used to deal with imbalanced data. Besides, they also used the 10-fold cross-validation method to solve this data imbalance issue. A resample filter was added to improve classifier performance. SMO outperformed the other two classifiers in terms of accuracy.

Some of the researchers also work for optimizing the model to improve performance. Assegie et al. [18] propose an optimized KNN model. They use grid search techniques to find the best value of k for the KNN model. This study also compares the effect of the hyper-parameter tuning model to the effect of the default hyper-parameter model. Hyperparameter tuning has a significant impact on the performance of the KNN model. The optimized hyper-parameter tuning model then

achieved 94.35% accuracy, while the default hyper-parameter achieved 90.10% accuracy.

Table I summarizes the existing ML works on Breast Cancer Classification. From the table, we can observe that researchers are usually confined into a set of ML algorithms, however, different researchers use different subset of algorithms and the reported results are also confusing in terms of identifying a clear winner. Therefore, we use a larger set of ML algorithms (more specifically 12 in number) and evaluated their performance in 2 different datasets.

3. METHODOLOGY

Our experiment comprises with following five steps:

- Experimental Methods with Parameters
- Environment Setup and Dataset Description
- Data Preprocessing
- Performance Evaluation Metrics
- Data Visualization

Figure 1 depicted our experiment steps. According to the figure, we describe the following steps.

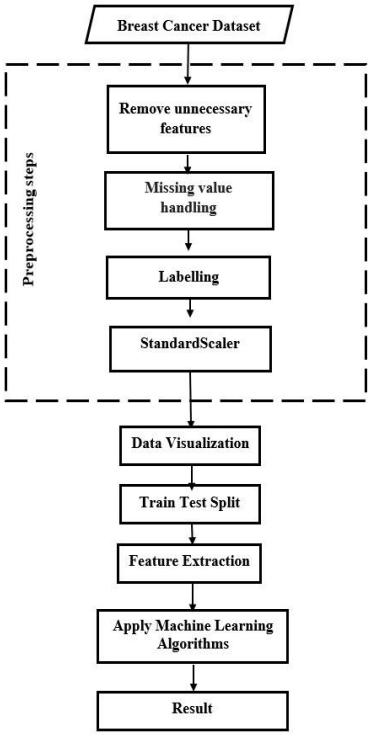

Fig. 1: A typical workflow diagram of our experiments.

3.1 Experimental Methods with Parameters

We use WBCD and WBC datasets for our experiment. This two datasets are not homogeneous feature sets and samples. Therefore, we use different parameters for different classifiers to improve performance. Furthermore, both the datasets are imbalance and they treat different classifiers in different ways, such as KNN is misclassified when k=1 because KNN works based on majority voting. In addition, NB classifiers are also affected by the imbalanced datasets because NB works based on target class probability. LR also gives false alarms because LF performs based on distance from the hyperplane. In the same way, the imbalanced datasets affect the other classifiers. Table II summarizes the parameter tuning of all twelve algorithms used in this experiment.

3.2 Environment Setup and Dataset Description

To compare the outcomes of the ML techniques basis on the two breast cancer datasets, we apply twelve ML techniques that are implemented on a machine with AMD Ryzen 9 3900X 12-Core Processor with 64GB RAM. We use opensource platforms such as Scikit-learn for the ML library. An integrated development environment named Jupyter Notebook is used to run the program. WBC and WBCD, both the datasets are collected from Kaggle. WBC dataset contains 699 samples with 11 features. They are two classes, Malignant and Benign, that denote 4 and 2, respectively. Following the figure 2 shows the statistics of WBC datasets. The datasets have 16 missing values that represent the question mark (”?”) symbol. In addition, the WBCD dataset contains 569 samples with 32 features where two classes are denoted as’ M’(Malignant) and ’B’(Benign), following the figure 3, which shows the statistics of WBCD datasets.

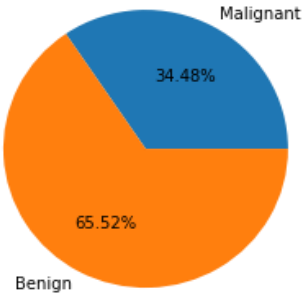

Fig. 2: Statistics of Wisconsin Breast Cancer (Original) (WBC) dataset.

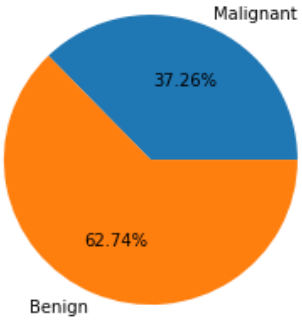

Fig. 2: Statistics of Wisconsin Breast Cancer Diagnosis (WBCD) dataset.

3.3 Data Preprocessing

In this prepossessing stage, we first drop the" Sample code number" feature from the WBC dataset and" id"," Unnamed: 32" from the WBCD dataset. The missing value ("?") of WBC data was replaced by the mode using the mode () function. The mode is a statistical term that returns the most frequently occurring value entered in a dataset. For labeling and scaling, we use LabelEncoder () and StandardScaler () functions.

3.4 Evaluation

The efficiency of ML algorithms is assessed using a set of performance measures. To evaluate the performance, TP, FP, TN, and FN are used to create a confusion matrix for the actual and predicted classes. The meanings of the terms are listed below.

- TP stands for True Positive (Correctly Classified)
- TN stands for True Negative (Incorrectly Classified)
- FP stands for False Positive (Correctly misclassified)
- FN stands for False Negative (Incorrectly misclassified)

The following formulas are used to evaluate the proposed system's performance [30].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Sensitivity or Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = 2 * \frac{\text{precision} * recall}{precision + recall} \quad (4)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (5)$$

$$\text{False Discovery Rate} = \frac{FP}{FP + TP} \quad (6)$$

$$\text{False Omission Rate} = \frac{FN}{FN + TN} \quad (7)$$

3.5 Data Visualization

Data visualization is the most important part of any ML application. Through data visualization, we find out the characteristics of data and how to correlate features to features. There are two classes as Malignant and Benign. In figure 4 and figure 5, we see that the two classes are almost separated. As a result, the machine learning algorithm is easily classified into two separate categories and achieve better performance.

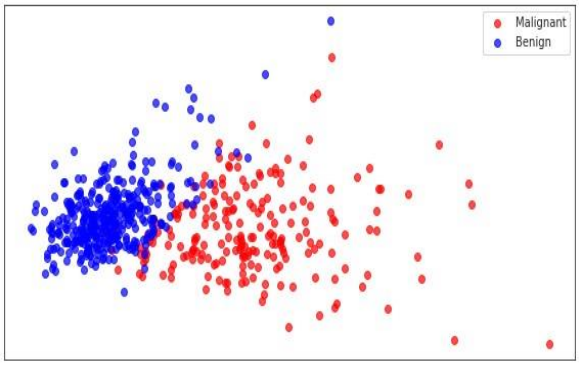

Fig. 4: Represent two classes are almost separated on the WBCD datasets.

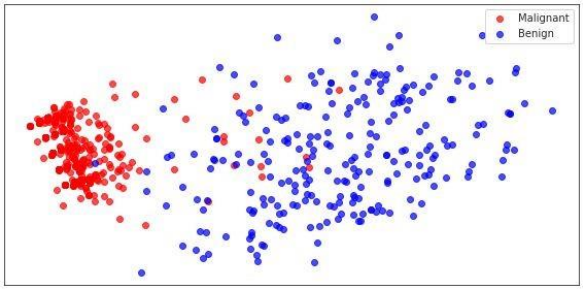

Fig. 5: Represent two classes are almost separated on the WBC datasets.

The WBCD dataset has 32 features, and the WBC dataset has 11 features. The following figure shows the correlation among features on the two datasets. In figures 6 and figures 7, we have seen that there are no correlated features.

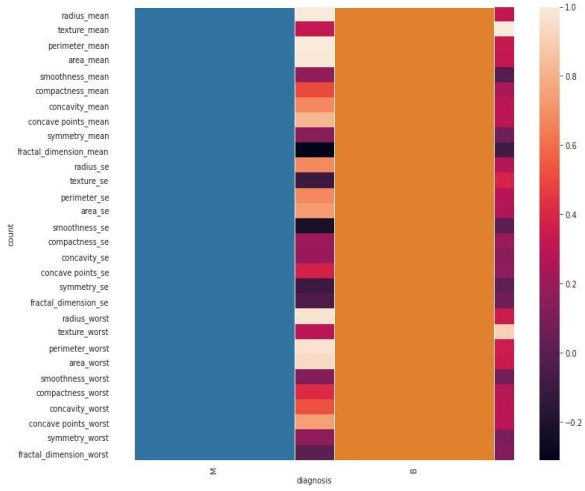

Fig. 6: Heat map for checking correlated features on the WBCD dataset.

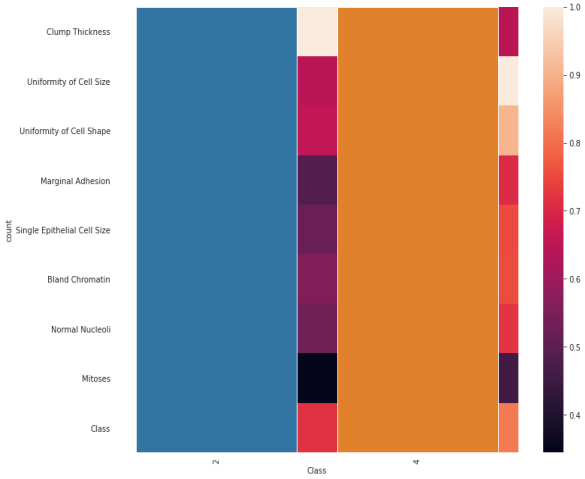

Fig. 7: Heat map for checking correlated features on the WBC dataset.

4. RESULTS AND DISCUSSION

This section is divided into two parts: (i) performance of the model without cross-validation and (ii) performance of the model with cross-validation. On the WBC dataset, 512 samples (90%) are used for training, and 57 samples (10%) are used for testing, while 629 samples (90%) are used for training and 70 samples (10%) were used for testing from WBCD dataset. Furthermore, the performance of all classifiers has been evaluated in terms of accuracy, precision, f1-measures, specificity, sensitivity, false discovery rate (FDR), and false omission rate (FOR).

4.1 The performance of the model without cross-validation

4.1.1 Performance of Training Phase

Table III displays the false negative value of twelve algorithms. False negative value is a critical issue in the classification of breast cancer.

The reason for giving zero false negative value of these classifiers is that these are ensemble and boosting classifiers. Table IV summarizes the evaluation result on WBC and Table V depict the result on WBCD.

In WBCD dataset AB, RF, and GB provide the promising results. In addition, DT, SVM, LDA, KNN, and BC also generate less false discovery rate with a high false omission rate. Furthermore, f1-score, specificity, and precision are almost nearly 1. Also, we investigate the model performance metrics on the WBC dataset.

Table V shows the all-performance metrics on the WBC dataset. From this table, it is clear that RF and GB models are the best model according to the whole metrics. Moreover, recall or sensitivity and f1-Score are also high in SVM, DT, SGD, and BC models.

Table VI displays the accuracy of different classifiers on the WBCD dataset. In this case, most classifiers provide accuracy greater than 95% except NB. Three classifiers, AB, RF, and GB, among them, provide 100% accuracy. Furthermore, VC, SGD and BC classifiers achieve the accuracy 99.02%, 99.02%, and 99.41%, respectively.

Table VII displays the accuracy of different classifiers on the WBC dataset. Here almost every classifier provides 95% accuracy, but the range of accuracy is less than shown in table IV. Also, AB, VC, and BC classifiers obtain 98.092%, 98.25%, and 98.73% accuracy.

4.1.2 Performance of Testing Phase

For both datasets, it can be observed without a doubt from Table VIII that the seven algorithms such as NB, LR, SVM, AB, RF, VC and BC do not provide any false negative values, indicating that they accurately identified each sample. Moreover, the three algorithms (DT, SGD, and GB) do not provide any false negative values on the WBCD dataset.

In Table IX, several classifiers, such as NB, LR, LDA, KNN, and RF show better performance in terms of all metrics. Hence, On the other performance parameters of classifiers are standard.

Overall, in Table X, the performance of all the classifiers is incredibly high. Additionally, NB, DT, and GB classifiers provide well-standard result.

From table XI, we can observe that the performance of each classifier varies, but NB, KNN, and RF consistently achieve the highest accuracy on the testing set (96.49%).

In addition, table XII, owing to cross-validation on the WBC dataset, boosts the accuracy of model where appropriate accuracy achieves in LR (98.6%), LDA (98.57%), KNN (98.97%), AB (98.57%), RF (98.7%), SGD (98.57%), BC (98.6%) and GB (98.57%).

4.2 The performance of the model with cross-validation

4.2.1 Performance of Training Phase

Since the datasets are imbalanced, for that some classifiers provide high accuracy. Because the dataset distribution is not uniform, the probabilistic model produced good results for certain test sets. Cross-validation is the most commonly used technique to solve this problem.

The table XII presents the cross-validation accuracy of twelve classifiers in the training phase on the WBCD dataset. Almost every classifier's cross-validation accuracy is lower than its original accuracy. It indicates that the imbalanced dataset had an effect on the classifier.

Table XIV also displays the cross-validation accuracy of twelve classifiers during the training phase on the WBC dataset. In this case, two classifiers, NB and LR, outperform the original accuracy. Furthermore, the other two classifiers, LDA and KNN, provide nearly the same accuracy as the original accuracy.

4.2.1 Performance of Testing Phase

The cross-validation accuracy of twelve classifiers during testing on the WBCD dataset is shown in Table XV. The results of LR, DT, AB, RF, and SGD are superior to the original results, as can be seen from the table. For this, we argued that these classifiers are not overfitted.

Likewise, the table XVI represents the cross-validation accuracy of twelve classifiers in the testing phase on the WBC dataset. Also, here three classifiers, LDA, KNN, and GB, outperform the original result. In addition, other classifiers also perform well. We see that LR, AB, RF, VC, SGD, and BC classifiers provide a 98.5% accuracy. So, we conclude that these classifiers perform well on the test dataset.

5. COMPARISON WITH PREVIOUS STUDY

Table XVII represents the comparison with the previous study. Some previous studies do not mention that this accuracy is train, test or validation. For that, we do not mention their accuracy in train, test or validation. We compare their result with our proposed study. We see that in study [10] SVM (98.57%) and KNN (97.14%) where our experimental result in testing phase SVM (100%) but cross-validation accuracy SVM (98%), which is more than their accuracy and also KNN outperform than their accuracy in the cross validation. However, our experimental methods outperform than others' previous studies.

6. Conclusion

Use of machine learning algorithms in breast cancer classification shows promising results. However, all the existing literature provides equivocal results in terms of accuracy and effectiveness. In this paper, we conduct a comparative study on most of the widely used ML algorithms and evaluate their performance on two well-reputed breast cancer classification datasets. This contribution provides a thorough evaluation of all the well-known ML algorithms and we can see most of the algorithm can outperform with proper parameter tuning.

7. References

[1] ttps://www.who.int/en/news-room/fact-sheets/detail/cancer. Last Access: 06.02.2022.

[2] Sivakami, K., and Nadar Saraswathi. "Mining big data: breast cancer prediction using DT-SVM hybrid model." International Journal of Scientific Engineering and Applied Science (IJSEAS) 1, no. 5 (2015): 418-429.

[3] Mojrian, Sanaz, Gergo Pinter, Javad Hassannataj Joloudari, Imre Felde, Akos Szabo-Gali, Laszlo Nadai, and Amir Mosavi. "Hybrid machine learning model of extreme learning machine radial basis function for breast cancer detection and diagnosis; a multilayer fuzzy expert system." In 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), pp. 1-7. IEEE, 2020.

[4] H. You and G. Rumbe, "Comparative study of classification techniques on breast cancer FNA biopsy data," International Journal of Artificial Intelligence and Interactive Multimedia, vol. 1, no. 3, pp. 6-13, 2010.

[5] Al Bataineh, Ali. ”A comparative analysis of nonlinear machine learning algorithms for breast cancer detection.” International Journal of Machine Learning and Computing 9, no. 3 (2019): 248-254.

[6] A. A. Ardakani, A. Gharbali, and A. Mohammadi, “Classification of breast tumors using sonographic texture analysis,” J. Ultrasound Med., vol. 34, no. 2, pp. 225–231, 2015.

[7] Al Bataineh, Ali. ”A comparative analysis of nonlinear machine learning algorithms for breast cancer detection.” International Journal of Machine Learning and Computing 9, no. 3 (2019): 248-254.

[8] B. L. Sprague, E. F. Conant, T. Onega, M. P. Garcia, E. F. Beaber, S. D. Herschorn, C. D. Lehman, A. N. A. Tosteson, R. Lacson, M. D. Schnall, D. Kontos, J. S. Haas, D. L. Weaver, and W. E. Barlow, “Variation in Mammographic Breast Density Assessments among Radiologists in Clinical Practice: A Multicenter Observational Study,” Ann. Intern. Med., vol. 165, no. 7, pp. 457–464, 2016.

[9] T. M. Kolb, J. Lichy, and J. H. Newhouse, “Comparison of the Performance of Screening Mammography, Physical Examination, and Breast US and Evaluation of Factors that Influence Them: An Analysis of 27,825 Patient Evaluations,” Radiology, vol. 225, no. 1, pp. 165–175, 2002.

[10] Islam, Md Milon, Hasib Iqbal, Md Rezwanul Haque, and Md Kamrul Hasan. ”Prediction of breast cancer using support vector machine and K-Nearest neighbors.” In 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), pp. 226-229. IEEE, 2017.

[11] Sivakami, ”Mining Big Data: Breast Cancer Prediction using DT-SVM Hybrid Model”, 2015

[12] Aslan, Muhammet Fatih, Yunus Celik, Kadir Sabancı, and Akif Durdu. ”Breast cancer diagnosis by different machine learning methods using blood analysis data.” (2018).

[13] M. Hussain, S. K. Wajid, A. Elzaart, and M. Berbar, “A comparison of SVM kernel functions for breast cancer detection.” IEEE Eighth International Conference Computer Graphics, Imaging and Visualization, pp. 145-150, 2011.

[14] Al Bataineh, Ali. ”A comparative analysis of nonlinear machine learning algorithms for breast cancer detection.” International Journal of Machine Learning and Computing 9, no. 3 (2019): 248-254.

[15] Mohammed, Siham A., Sadeq Darrab, Salah A. Noaman, and Gunter Saake. ”Analysis of breast cancer detection using different machine learning techniques.” In International Conference on Data Mining and Big Data, pp. 108-117. Springer, Singapore, 2020.

[16] Agarap, Abien Fred M. ”On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset.” In Proceedings of the 2nd international conference on machine learning and soft computing, pp. 5-9. 2018.

[17] 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.

[18] Assegie, Tsehay Admassu. ”An optimized K-Nearest Neighbor based breast cancer detection.” Journal of Robotics and Control (JRC) 2, no. 3 (2021): 115-118.

[19] Bayrak, Ebru Aydındag, Pınar Kırcı, and Tolga Ensari. ”Comparison of machine learning methods for breast cancer diagnosis.” In 2019 Scientific meeting on electrical-electronics biomedical engineering and computer science (EBBT), pp. 1-3. Ieee, 2019.

[20] Potdar, Kedar, and Rishab Kinnerkar. ”A comparative study of machine learning algorithms applied to predictive breast cancer data.” International Journal of Science and Research 5, no. 9 (2016): 1550-1553.

[21] Gayathri, B. M., and C. P. Sumathi.” Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer.” In 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 15. IEEE, 2016.

[22] Siegel, R. L., K. D. Miller, and A. Jemal. ”Cancer statistics, 2020 CA Cancer J Clin.” American Cancer Society 70 (2020): 7-30.

[23] Byrne, Dallan, Martin O’Halloran, Edward Jones, and Martin Glavin. ”A comparison of data-independent microwave beamforming algorithms for the early detection of breast

[24] Samsuzzaman, Md, Mohammad T. Islam, Md T. Islam, Abdullah AS Shovon, Rashed I. Faruque, and Norbahiah Misran. ”A 16-modified antipodal Vivaldi antenna array for microwave-based breast tumor imaging applications.” Microwave and Optical

Technology Letters 61, no. 9 (2019): 2110-2118.

[25] Alqahtani, Wedad Saeed, Nawaf Abdulrahman Almufareh, Dalia Mostafa Domiaty, Gadah Albasher, Manal Abduallah Alduwish, Huda Alkhalaf, Bader Almuzzaini et al. "Epidemiology of cancer in Saudi Arabia thru 2010–2019: a systematic review with constrained metaanalysis." AIMS Public Health 7, no. 3 (2020): 679.

[26] Brown, Martin L., Sue J. Goldie, Gerrit Draisma, Joe Harford, and Joseph Lipscomb. "Health service interventions for cancer control in developing countries." Disease Control Priorities in Developing Countries. 2nd edition (2006).

[27] Story, H. L., R. R. Love, R. Salim, A. J. Roberto, J. L. Krieger, and O. M. Ginsburg. "Improving outcomes from breast cancer in a lowincome country: lessons from Bangladesh." International journal of breast cancer 2012 (2012).

[28] Tan, Y. J., K. S. Sim, and F. F. Ting. "Breast cancer detection using convolutional neural networks for mammogram imaging system." In 2017 International Conference on Robotics, Automation and Sciences (ICORAS), pp. 1-5. IEEE, 2017.

[29] Abien Fred Agarap. 2017. A Neural Network Architecture Combining Gated Recurrent Unit (GRU) and Support Vector Machine (SVM) for Intrusion Detection in Network Traffic Data. arXiv preprint arXiv:1709.03082 (2017).

[30] Alam BR, Khatun MS, Taslim M, Hossain MA. Handling Class Imbalance in Credit Card Fraud Using Various Sampling Techniques. American Journal of Multidisciplinary Research and Innovation. 2022 Oct 3;1(4):160-8.

Table I. Analysis of The Previous Study

Authors	Models	Dataset	Results (Best)
Islam et al. [10]	1. SVM 2. K-NN	WBCD	SVM (98.57%)
K Sivakami et al. [11]	1. DT-SVM 2. Instance-based Learning (IBL) 3. Sequential Minimal Optimization (SMO) 4. Nave-based classifiers.	WBCD	DT-SVM (91%)
AA Bataineh et al. [14]	1. Multilayer Perceptron (MLP) 2. KNN 3. Classification and Regression Trees (CART) 4. NB 5. SVM	WBCD	MLP (99.12%)
Mohammed et al. [15]	1. DT(J48) 2. NB 3. SMO	1. WBC 2. Breast Cancer dataset	1. For Wisconsin Breast Cancer (WBC) SMO (99.56%) 2. For the Breast Cancer dataset J48 (98.20%)
S Sharma et al. [17]	1. RF 2. K-NN 3. NB	WBCD	K-NN (95.90%)
Assegie et al. [18]	K-NN	WBC	K-NN (94.35%)
MF Aslan et al. [12]	1. ANN 2. Extreme Learning Machine (ELM) 3. SVM 4. k-NN.	Blood analysis Dataset	ELM (80%)
AA Bataineh et al. [7]	1. NB 2. CART 3. MLP 4. SVM	WBCD	MLP (96.70%)

Table II. Experimental Methods with Parameters

Dataset	Classifier	Cross-validation function	Cross-validation Parameter	Tuning Parameter
WBCD	NB	KFold	nsplits=10, shuffle=True, random state=101	default
	LR	RepeatedKFold	nsplits=8, random state=51	default
	DT	StratifiedKFold	nsplits=5, shuffle=True, random state=101	criterion=" entropy", max depth=3
	SVM	StratifiedKFold	nsplits=8, shuffle=True, random state=101	kernel='linear'
	LDA	KFold	nsplits=6,shuffle=True, random state=101	default
	K-NN	StratifiedKFold	nsplits=5,shuffle=True, random state=51	n neighbors=3
	AD	KFold	nsplits=5,shuffle=True, random state=101	default
	RF	RepeatedKFold	nsplits=10, randomstate=101	default
	VC	RepeatedStratifiedKFold	nsplits=10, randomstate=101	estimators= [AdaBoostClassifier(), Logistic Regression (), RandomForestClassifier(), SVC()]
	SGD	StratifiedKFold	nsplits=10, shuffle=True, random state=51	max iter=500, random state=51
	BC	KFold	nsplits=10,shuffle=True, random state=101	base estimator = (RandomForestClassifier(), n estimators = 100)
	GB	KFold	nsplits=10, shuffle=True, random state=51	default
WBC	NB	KFold	nsplits=10, shuffle =True, random state=101	default
	LR	KFold	nsplits=10, shuffle=True, random state=101	default
	DT	StratifiedKFold	nsplits=10, shuffle=True, random state=101	criterion="entropy", max depth=3
	SVM	KFold	nsplits=10,shuffle=True, random state=101	default
	LDA	KFold	nsplits=10, shuffle=True, random state=101	default
	K-NN	KFold	nsplits=10, shuffle=True, random state=101	default
	AD	KFold	n splits=10, shuffle=True, random state=101	default
	RF	KFold	nsplits=10, shuffle=True, random state=101	default
	VC	KFold	nsplits=10, shuffle=True, random state=101	estimators= [AdaBoostClassifier (), Logistic Regression (), RandomForestClassifier(), SVC ()]
	SGD	RepeatedStratifiedKFold	nsplits=10,random state=101	max iter=100
	BC	KFold	nsplits=10,shuffle=True, random state=101	baseestimator= (RandomForestClassifier( ), n estimators = 100)
	GB	KFold	nsplits=10,shuffle=True, random state=101	default

Table III. Model Performance Metrics on The WBCD Dataset in The Training Phase

Parameters	NB	LR	DT	SVM	LDA	KNN	AB	RF	VC	SGD	BC	GB
Precision	0.96	0.99	0.99	0.99	0.99	0.99	1.0	1.0	1.0	0.99	1.0	1.0
Recall or Sensitivity	0.94	0.98	0.95	0.98	0.95	0.96	1.0	1.0	0.98	0.99	0.99	1.0
Specificity	0.94	0.99	0.99	0.99	0.99	0.99	1.0	1.0	1.0	0.99	1.0	1.0
F1-Score	0.95	0.99	0.97	0.99	0.97	0.98	1.0	1.0	0.99	0.99	0.99	1.0
False Discovery Rate (FDR)	0.038	0.003	0.003	0.003	0.003	0.003	0.0	0.0	0.0	0.006	0.0	0.0
False Omission Rate (FOR)	0.095	0.025	0.08	0.025	0.08	0.05	0.0	0.0	0.025	0.015	0.015	0.0

Table IV. Model Performance Metrics on The WBC Dataset in The Training Phase

Parameters	NB	LR	DT	SVM	LDA	KNN	AB	RF	VC	SGD	BC	GB
Precision	0.949	0.974	0.945	0.969	0.978	0.974	0.983	1.0	0.986	0.964	0.983	1.0
Recall or Sensitivity	0.985	0.974	0.997	0.990	0.958	0.985	0.988	1.0	0.988	0.993	0.998	1.0
Specificity	0.907	0.948	0.902	0.941	0.956	0.949	0.967	1.0	0.972	0.933	0.968	1.0
F1-Score	0.967	0.974	0.970	0.979	0.968	0.979	0.986	1.0	0.987	0.978	0.990	1.0
False Discovery Rate (FDR)	0.050	0.026	0.055	0.031	0.022	0.026	0.017	0.0	0.014	0.035	0.017	0.0
False Omission Rate (FOR)	0.028	0.052	0.005	0.019	0.085	0.028	0.024	0.0	0.023	0.014	0.004	0.0

Table V. Accuracy on The WBCD Dataset in The Training Phase

Parameters	NB	LR	DT	SVM	LDA	KNN	AB	RF	VC	SGD	BC	GB
Accuracy	93.94	98.83	96.68	98.83	96.68	97.85	100	100	99.02	99.02	99.41	100

Table VI. Accuracy on The WBC Dataset in The Training Phase

Parameters	NB	LR	DT	SVM	LDA	KNN	AB	RF	VC	SGD	BC	GB
Accuracy	95.71	96.50	96.18	97.29	95.70	97.29	98.09	100	98.25	97.14	98.73	100

Table VII. False Negative Value in The Testing Phase

Dataset	NB	LR	DT	SVM	LDA	KNN	AB	RF	VC	SGD	BC	GB
WBCD	0	0	0	0	2	1	0	0	0	0	0	0
WBC	0	0	1	0	1	2	0	0	0	1	0	1

Table VIII. Model Performance Metrics on The WBCD Dataset in The Testing Phase

Parameters	NB	LR	DT	SVM	LDA	KNN	AB	RF	VC	SGD	BC	GB
Precision	0.95	0.93	0.88	0.93	0.95	0.98	0.88	0.95	0.93	0.84	0.93	0.93
Recall or Sensitivity	1.0	1.0	1.0	1.0	0.95	0.98	1.0	1.0	1.0	1.0	1.0	1.0
Specificity	0.87	0.82	0.73	0.82	0.86	0.93	0.74	0.88	0.82	0.67	0.82	0.82
F1-Score	0.97	0.96	0.93	0.96	0.95	0.98	0.94	0.98	0.96	0.91	0.96	0.96
False Discovery Rate (FDR)	0.04	0.06	0.11	0.06	0.05	0.02	0.12	0.04	0.07	0.16	0.07	0.07
False Omission Rate (FOR)	0.0	0.0	0.0	0.0	0.14	0.07	0.0	0.0	0.0	0.0	0.0	0.0

Table IX. Model Performance Metrics on The WBCD Dataset in The Testing Phase

Parameters	NB	LR	DT	SVM	LDA	KNN	AB	RF	VC	SGD	BC	GB
Precision	0.976	1.0	0.976	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.976
Recall or Sensitivity	1.0	1.0	0.976	1.0	0.976	0.953	1.0	1.0	1.0	0.976	1.0	0.976
Specificity	0.967	1.0	0.966	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.966
F1-Score	0.988	1.0	0.976	1.0	0.988	0.976	1.0	1.0	1.0	0.988	1.0	0.976
False Discovery Rate (FDR)	0.024	0.0	0.024	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.024
False Omission Rate (FOR)	0.0	0.0	0.034	0.0	0.034	0.069	0.0	0.0	0.0	0.034	0.0	0.034

Table X. Accuracy on The WBCD Dataset in The Testing Phase

Parameters	NB	LR	DT	SVM	LDA	KNN	AB	RF	VC	SGD	BC	GB
Accuracy	96.49	94.74	91.23	96.49	92.98	96.49	91.23	96.49	94.74	87.72	94.74	94.74

Table XI. Accuracy on The WBC Dataset in Testing Phase

Parameters	NB	LR	DT	SVM	LDA	KNN	AB	RF	VC	SGD	BC	GB
Accuracy	98.57	100	97.14	100	98.57	97.14	100	100	100	98.57	100	97.14

Table XII. Cross Validation Accuracy on The WBCD Dataset in The Training Phase

Parameters	NB	LR	DT	SVM	LDA	KNN	AB	RF	VC	SGD	BC	GB
Cross validation	93.15	97.91	93.75	97.85	96.49	96.68	96.29	96.03	97.93	96.88	95.51	96.29

Table XIII. Cross Validation Accuracy on The WBC Dataset in The Training Phase

Parameters	NB	LR	DT	SVM	LDA	KNN	AB	RF	VC	SGD	BC	GB
Cross validation	95.69	96.50	95.23	96.50	95.23	97.14	95.71	97.14	96.66	95.95	97.30	96.19

Table XIV. Cross Validation Accuracy on The WBCD Dataset in The Testing Phase

Parameters	NB	LR	DT	SVM	LDA	KNN	AB	RF	VC	SGD	BC	GB
Cross validation	95.0	96.67	91.67	93.07	90.0	93.07	91.48	96.67	94.67	98.33	94.6	91.67

Table XV. Cross Validation Accuracy on The WBC Dataset in The Testing Phase

Parameters	NB	LR	DT	SVM	LDA	KNN	AB	RF	VC	SGD	BC	GB
Cross validation	95.71	98.61	95.71	98.00	98.73	98.87	98.57	98.72	98.73	98.57	98.61	98.57

Table XVI. Comparison With The Previous Study

Previous Study			Proposed Methods		
Ref.	Methods	Dataset	Methods	Testing	Cross-Validation
[10]	SVM – 98.57% KNN – 97.14%	WBC	SVM KNN	100% 97.143%	98.00% 98.872%
[2]	DT – SVM – 91%	WBC	DT SVM	97.143% 100%	95.714% 98.00%
[5]	NB – 94.73 KNN – 95.61	WBCD	NB KNN	96.49% 96.49%	95% 93.07%
[14]	SVM (RBF) - 96.28	WBC	SVM(RBF)	100%	96.503%