Vol 7 | Issue 1 | Pages 106-112

Trends in Artificial Intelligence
ISSN: 2643-6000

Research Article

DOI: 10.36959/643/310

Evaluating the Generalizability of Support Vector Machine for Breast Cancer Detection

Oluwaseyi Ezekiel Olorunshola, Okeh Dominic Ebuka and Adeniran Kolade Ademuwagun*

Computer Science Department, Air Force Institute of Technology, Kaduna

Abstract

Breast cancer is caused by the abnormal growth of cells in the breast. Malignant breast cancer is an aggressive type that, if not diagnosed and treated early, can be fatal, While the benign breast cancer can be removed without recurring when detected early. Recent research has shown that Breast cancer detection relies heavily on accurate and reliable methods. Machine learning models, particularly Support Vector Machines (SVMs), have shown promise in this area. However, concerns exist regarding their generalizability across real-world scenarios with varying software environments and data processing techniques. This research proposes to investigate this gap by comparing SVM performance with other classifiers such as Naïve Bayes, Random Forest, Multilayer Perceptron and Decision Tree. These classifiers were tested on the Wisconsin Breast Cancer dataset using both the Waikato Environment for Knowledge Analysis and Jupyter Notebook. The study recorded performance metrics such as accuracy, precision, recall, and f1_score. After the analysis, it was observed that in WEKA, Support Vector Machine under the 10-fold cross-validation and 70% split, had the highest accuracies of 0.981 and 0.977 respectively. Interestingly, Multilayer Perceptron also achieved an accuracy of 0.977 under the 70% split. In the Jupyter Notebook, the Support Vector Machine also produced the highest accuracy value of 0.99 under the 70% split. Random Forest produced the highest accuracy of 0.97 while SVM had a value of 0.96 in the 10-fold cross-validation. It can be noted that the result generated were for just two environments and techniques, whereas there are more techniques and environments available.

Keywords

Malignant, Benign, Support vector machine, Classifiers, Evaluation metrics, Breast cancer

Introduction

Breast cancer occurs when abnormal cells in the breast grow uncontrollably, forming tumors that can potentially spread throughout the body and become life-threatening [1]. The year 2020 saw 2.3 million women diagnosed with breast cancer and 685,000 deaths worldwide. By the end of 2020, 7.8 million women who had been diagnosed with breast cancer in the previous 5 years were still alive, making it the most common cancer globally [1]. From recent research, it was stated that timely detection of the disease can lead to a positive prognosis and a high chance of survival. In North America, patients with breast cancer have a 5-year relative survival rate of over 80% due to the early detection of the disease [2]. The traditional method for detecting cancer relies on a gold-standard approach involving three tests: Radiological imaging, clinical examination, and pathology testing. This conventional method relies on regression to determine the presence of cancer. The effective incorporation of Information and Communication Technologies (ICT) into medical practice has become a crucial factor in the modernization of the healthcare system, particularly in the realm of cancer treatment [3]. The latest machine learning (ML) techniques and algorithms are developed based on model design. ML is a computational approach that can be used to find the best solutions to a problem without requiring explicit programming by a computer programmer or an experimenter [4]. The utilization of ML models, particularly Support Vector Machines (SVMs), has displayed notable potential in the realm of breast cancer detection through the analysis of mammograms and other imaging modalities. However, concerns exist regarding the generalizability of these models when applied in real-world settings. Current research on SVM models for breast cancer detection often evaluates them in single environments and with limited

***Corresponding author:** Oluwaseyi Ezekiel Olorunshola, Computer Science Department, Air Force Institute of Technology, Kaduna

Accepted: September 25, 2024

Published online: September 27, 2024

Citation: Olorunshola OE, Ebuka OD, Ademuwagun AK (2024) Evaluating the Generalizability of Support Vector Machine for Breast Cancer Detection. Trends Artif Intell 7(1):106-112

Copyright: © 2024 Olorunshola OE, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

SCHOLARS.DIRECT

variations in training and testing methodologies. This raises questions about whether the reported accuracy and precision translate well to different software platforms and data splitting techniques. This research aims to investigate the generalizability of SVM models for breast cancer detection by comparing its performance with other classifiers such as Naïve Bayes (NB), Random Forest (RF), Multilayer Perceptron (MLP) Neural Network (NN), and Decision Tree (DT) under various programming environments and data splitting techniques. The performance of each classifier will be compared across both environments and data-splitting techniques using the chosen evaluation metrics. Statistical tests will be conducted to assess the significance of any observed differences. This research hypothesizes that while SVM models might achieve high accuracy in specific environments, their performance may deteriorate when applied to different software platforms or with alternative data-splitting methods. This research is expected to reveal the generalizability of SVM models for breast cancer detection. By comparing SVM with other algorithms under various conditions, the study will provide valuable insights into the robustness and reliability of these models in practical applications. The analysis of this research was carried out using the Waikato Environment for Knowledge Analysis (WEKA), a tool that provides Classification, Clustering, Association Mining, Feature Selection, and Data Visualization [5]. Additionally, Python's Jupyter Notebook was used to evaluate the performance of these five classifiers using the four performance metrics: accuracy, precision, recall, and F1-score. This was done to validate the results obtained from WEKA. The remaining part of this paper is arranged as follows; Section 2 contains a literature review while Section 3 contains the methodology. Results are discussed and analyzed in Section 4 while Section 5 concludes the paper.

Literature Review

Hoque, et al. [6] utilized the Extreme Gradient Boosting (XGBoost) ML technique to detect and analyze breast cancer. The breast cancer Wisconsin (diagnostic) dataset was used in the study and it comprised of 569 rows, where each row denoted a distinct digitized image of a breast mass, and 33 columns. Out of 569 rows, no column had missing data besides the "Unnamed: 32" column which only had null values. It was stated in the study that in contrast to linear regression models, ML models like XGBoost and RF models were generally resistant to multicollinearity between features. Hence, for this problem, the researchers refrained from using a linear regression model. The result stated that XGBoost provided an accuracy of 94.74% and a recall of 95.24%.

Islam, et al. [7] evaluated and compared the classification accuracy, precision, recall, and F1 scores of five different ML methods using a primary dataset (500 patients from Dhaka Medical College Hospital). It was stated that ML and Explainable Artificial Intelligence (AI) were crucial in classification as they not only provide accurate predictions but also offered insights into how the model arrived at its decisions, aiding in the understanding and trustworthiness of the classification results. Five different supervised ML techniques, including decision tree, random forest, logistic regression, naive Bayes, and XGBoost, were used to achieve optimal results on the dataset. The study applied SHAP analysis on the XGBoost model to interpret the model's predictions and understand the impact of each feature on the model's output. After the final evaluation, the XGBoost achieved the best model accuracy score, which was 97%.

Dinesh, et al. [8] carried out a study to compare the efficacy of the state-of-the-art SVM method for image prediction with that of KNN, Logistic Regression, RF, and DT. The study made use of the UCI ML Laboratory which provided a total of 569 samples. The maximum acceptable error was set at 0.5, and the minimum power of analysis was set at 0.8. Predictions made using Logistic Regression appeared to have a higher accuracy (95%) than those made using SVM, KNN, DT, or RF (92%, 90%, 85%, and 91%). This proposed system had a probability importance of 0.55.

Elsadig, et al. [9] selected eight classification algorithms that had been used to predict breast cancer to be under investigation. These classifiers include single and ensemble classifiers. A trusted dataset has been enhanced by applying five different feature selection methods to pick up only weighted features and neglect others. Accordingly, a dataset of only 17 features has been developed, SVM is ranked at the top by obtaining an accuracy of 97.7% with classification errors of 0.029 False Negative (FN) and 0.019 False Positive (FP). Therefore, it was noteworthy that SVM was the best classifier and outperformed even the stack classier.

Using the Wisconsin Breast Cancer Diagnosis Dataset, Strelcenia and Prakoonwit [10] presented an effective feature engineering method to extract and modify features from data and the effects it has on different classifiers. The feature was used to compare six popular ML models for classification. The models compared were Logistic Regression, RF, DT, K-Neighbors, MLP, and XGBoost. The results showed that the DT model, when applied to the proposed feature engineering, was the best performing, achieving an average accuracy of 98.64%.

Chaurasiya and Rajak [11] carried out an experiment to compare the accuracy measures of four prominent classification models considering their performance qualitatively on Wisconsin Diagnostic Breast Cancer (WDBC) dataset. RF, SVM, KNN and Logic Regression ML algorithms were analyzed on a classification technique that generally contains two different steps. In the first step the training dataset which contains labelled classes was used to build classification model by selecting a suitable classification algorithm. In the later step which is predictive phase, the accuracy of the built classification model was evaluated on the validation dataset. RF classifier was experimentally observed to be the best algorithm with accuracy of 95% and precision of 90.9% as compared to the other three classifiers.

Guleria, et al. [12] research was based on the prediction and diagnosis of the classes of breast cancer (benign or malignant) by using supervised learning techniques in WEKA. The research made use of KNN (83.41% precision, 90.04% recall, 80.42% accuracy, and 0.86 F-Measure), NB (88.37% precision, 94.53% recall, 87.41% accuracy, and 0.91

F-Measure), LR (81.65% precision, 88.55% recall, 77.97% accuracy, and 0.84 F-Measure), and DT (85.71% precision, 92.53% recall, 83.91% accuracy, and 0.88 F-Measure). It was observed that the prediction model built-up by NB provided the higher accuracy as well as higher F-measure among all the algorithm.

Ibeni, et al. [13] made use of three classifiers NB, BN, and Tree Augmented Naïve Bayes (TAN). This paper presented the fully Bayesian approach to assess the predictive distribution of all classes using three datasets: Breast cancer, breast cancer Wisconsin, and breast tissue dataset. The prediction accuracies of Bayesian approaches were also compared with K-NN, DT (J48) and SVM. The results of the performance metrics evaluated on the algorithms were arranged according to accuracy, precision, recall, and F-measure for the breast cancer dataset Algorithm: KNN: 94.992%, 96.94%, 95.483%, and 96.207%. SVM: 96.852%, 97.161%, 98.017% and 98.591%. DT (J48): 94.992%, 95.633%, 96.688% and 96.157%. BN: 97.28%, 96.506%, 99.325% and 97.895%. NB: 95.994%, 95.196%, 98.642%, and 96.888%. TAN: 96.280%, 95.851%, 98.430%, and 97.123%. The result showed that BN was the best performing algorithm.

Akbuğday [4] investigated the accuracies of three different ML algorithms; k-NN, NB, and SVM using WEKA. The values of the report were as follows; K-NN had 96.85% accuracy, NB had 95.99% accuracy and C-SVM a sub-classifier of SVM had 96.85%. It was observed that k-NN and SVM algorithms were the most accurate ones with identical confusion matrices and accuracy values.

Keleş [14] research was aimed at the prediction and detection of breast cancer early with non-invasive and painless methods that use data mining algorithms. In this study, an antenna was designed to operate in the 3-12 GHz UWB frequency range and a 3D breast structure consisting of skin layer, fat layer, and fibro glandular layer was designed. A separate model was also designed by adding a tumor layer to the breast structure. The dataset that was created had 6006 rows/values, 5405 of which were used as the training dataset, while 601 were used as the test data set. The dataset was then converted to the arff format, which was the file type used by the Weka tool. The 10-fold cross-validation technique was then used to obtain the most accurate results using the Knowledge Extraction based on Evolutionary Learning (Keel) data mining software tool. The results indicated that Bagging, IBk, Random Committee, RF, and Simple CART algorithms were the most successful algorithms, with over 90% accuracy in detection.

Methodology

The research design, environment and dataset are described in this Section. And also, the algorithm and performance metrics are also examined. Two different environments were used to analyze the datasets in order to determine the best performing classifier out of the 5 classifiers against the 4 performance metrics for breast cancer prediction. These environments are WEKA and Python's Juypiter Notebook. The 10-fold cross-validation and 70% split was carried out in each of the 2 environments.

The following methods were used to compare WEKA's user-friendly platform, which is great for initial exploration and rapid prototyping, with Python's power and flexibility for building and deploying advanced ML models for breast cancer analysis. Each environment has its advantages; for instance, Python offers advanced techniques, scalability, integration, deployment, and sharing, whereas WEKA provides rapid prototyping, testing, and data processing tools. Akbuğday [4] stated that due to WEKA's Java-based nature and comprehensive built-in algorithm library, employing the use of another platform with better-implemented algorithms environments such as Python or R may lead to more accurate classifiers with better programming practices and platform-specific advantages. Hence, this research was carried out using the 2 environments.

Research design

This research utilizes a quantitative research methodology to conduct a comparative analysis of various ML algorithms, assessing their performance using specific metrics, to predict breast cancer mortality.

Environment description

The analysis in this study was conducted using WEKA version 3.8.6 and Python Jupyter Notebook version 7.1.2. WEKA, developed by Holmes, Donkin, and Witten in 1994, is an open-source ML software that offers a comprehensive collection of tools for data preprocessing, classification, regression, clustering, association rules mining, and visualization. Jupyter Notebook is a project Spun off IPython in 2014 by Fernando Perez and Brian Granger. It is a non-profit, open-source project born out of the IPython in 2014 as it evolved to support interactive data science and scientific computing across all programming language. Jupyter notebook was created based on Python Programming Language developed by Guido van Rossum, a Dutch programmer in the late 1980s.

Data description

The dataset; Breast Cancer Wisconsin (Diagnostic) dataset used in this study was sourced from the UCI Repository. It consists of clinical and demographic features of breast cancer patients. The dataset comprises features extracted from digitized Fine Needle Aspirate (FNA) biopsies images. This dataset is a multivariate dataset which consists of 569 instances and 33 features, and it has 0 mismatches and 0 missing values.

Performance metrics:

i. Accuracy: The ratio between the correctly classified samples and the total number of samples in the evaluation dataset [15].

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

ii. The recall: Also known as the sensitivity or True Positive Rate (TPR), and is calculated as the ratio between correctly classified positive samples and all

Table 1: Result of the performance metrics of the algorithms in WEKA.

	Accuracy		Recall		F1-Score		Precision	
	10-fold split	**70% split**	**10-fold split**	**70% split**	**10-fold split**	**70% split**	**10-fold split**	**70% split**
DT	0.933	0.935	0.933	0.927	0.929	0.931	0.926	0.936
SVM	**0.981**	**0.977**	**0.976**	0.973	**0.98**	**0.975**	**0.983**	**0.978**
MLP	0.976	**0.977**	0.963	**0.975**	0.965	**0.975**	0.966	0.975
NB	0.926	0.918	0.92	0.917	0.921	0.914	0.922	0.912
R F	0.965	0.971	0.956	0.968	0.962	0.969	0.967	0.971

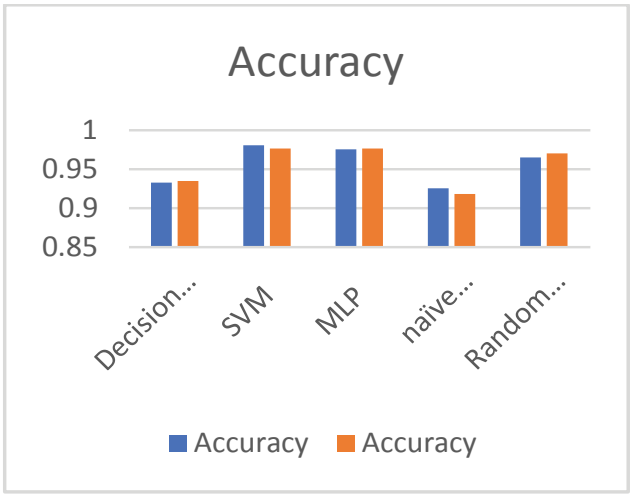

Figure 1: Comparison of Accuracy in WEKA.

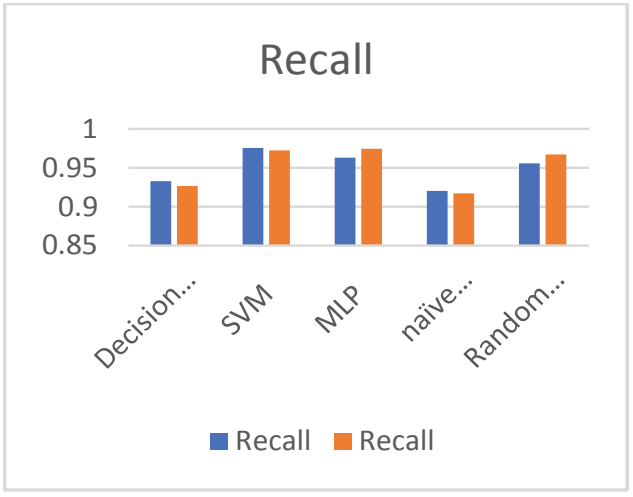

Figure 2: Comparison of Recall in WEKA.

samples assigned to the positive class [15].

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

The Precision: Is the calculated as the ratio between correctly classified samples and all samples assigned to that class [15].

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

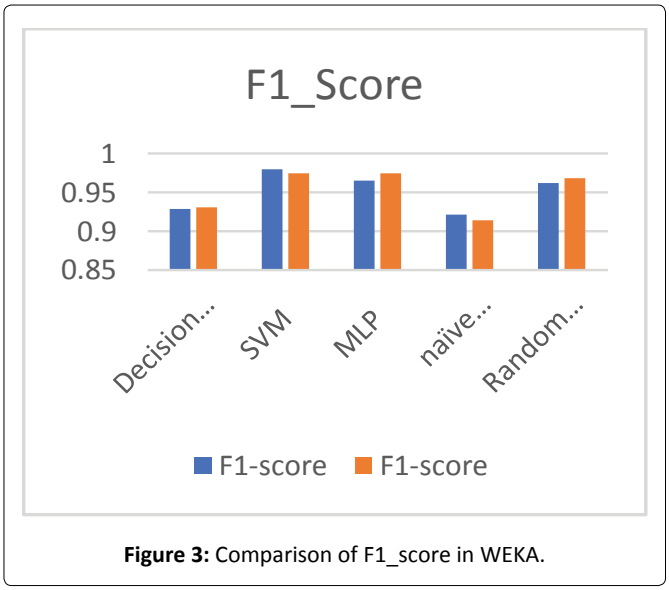

Figure 3: Comparison of F1_score in WEKA.

iii. F1-Score: The F1 score is the harmonic mean of precision and recall, meaning that it penalizes extreme values of either [15].

$$F1 = 2 * \frac{precision * recall}{precision + recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

Results and Discussion

This Section discusses and analyzes the values of the performance metrics obtained on each classifier when 10-fold cross-validation and 70% split was applied on the dataset in both WEKA and Jupyter Notebook environments. Table 1 and Figure 1, Figure 2, Figure 3 and Figure 4 show the results from WEKA while Table 2 and Figure 5, Figure 6, Figure 7 and Figure 8 show the results from Python's Jupyter Notebook for accuracy, recall, F1 score and precision [16-18].

Discussion on WEKA environment

Figure 1 shows the results obtained after the models implemented accuracy. SVM achieved accuracies of 0.981 and 0.977 under the 10-fold validation and 70% split respectively. MLP also achieved an accuracy of 0.977 under the 70% split, matching the SVM accuracy score. In Figure 2, the Recall of SVM is observed to be 0.976 in the 10-fold cross-validation, while MLP achieved the highest Recall of 0.975 under the 70% split. In Figure 3, SVM produced 0.98 and 0.975 F1_score for both the 10-fold cross-validation and 70% split respectively

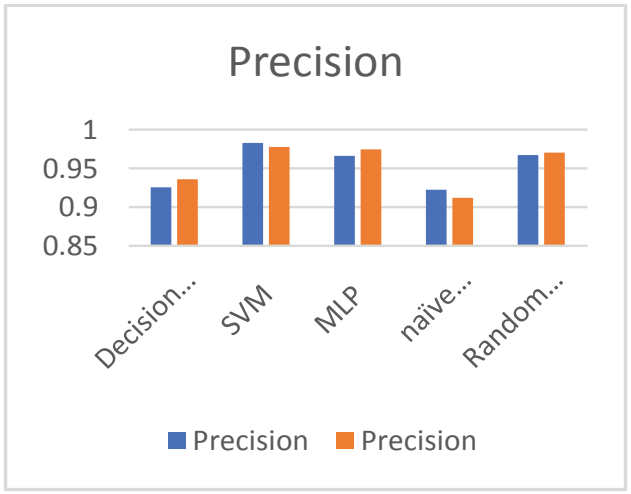

Figure 4: Comparison of Precision in WEKA.

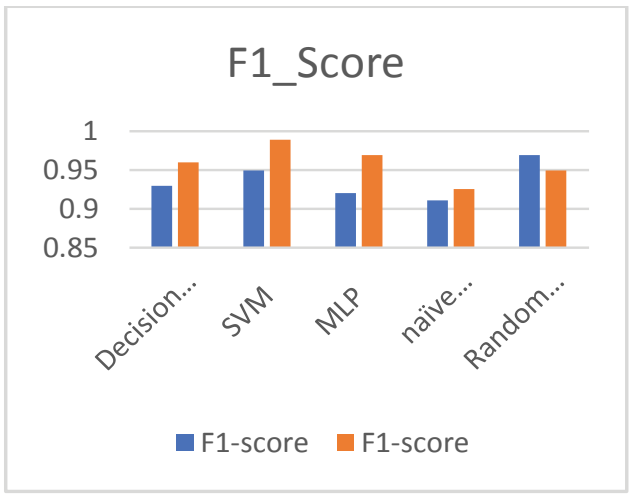

Figure 7: Comparison of F1_score in Python.

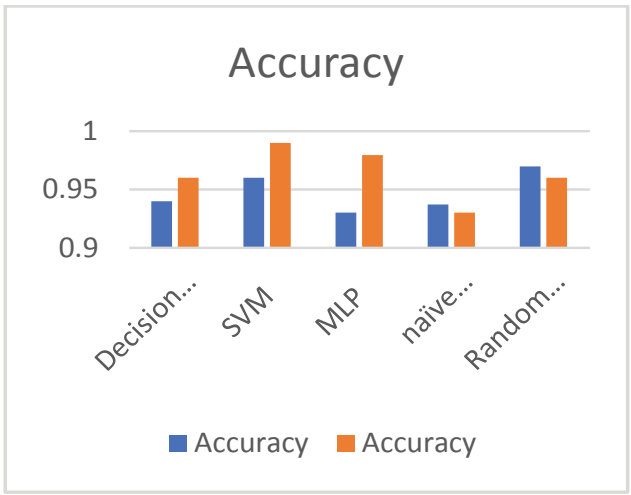

Figure 5: Comparison of Accuracy in Python.

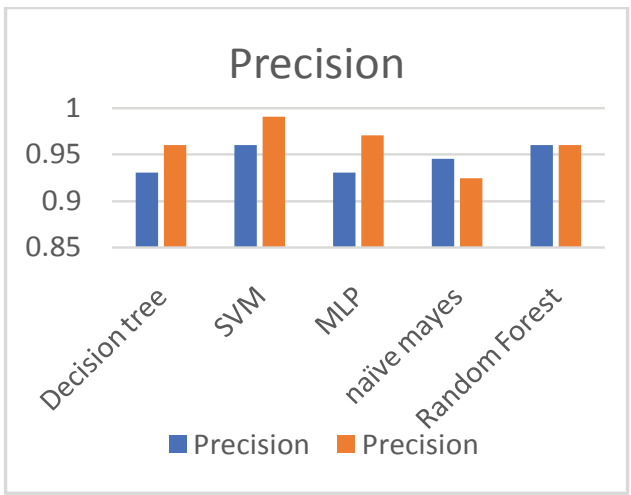

Figure 8: Comparison of Precision in Python.

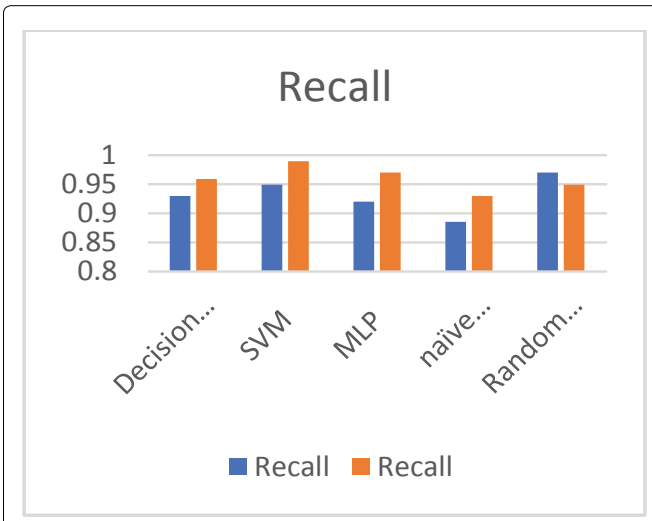

Figure 6: Comparison of Recall in Python.

but it's worth noting that MLP with score of 0.975 rivals the 70% split score of SVM. In Figure 4 which shows Precision results, SVM had the highest value of 0.983 and 0.978 for both 10-fold cross-validation and 70% split, respectively. Therefore, it can be concluded that SVM produced the overall best results in predicting breast cancer.

Discussion on Python's Jupyter notebook

In Figure 5, SVM outperformed all the other metrics and had an accuracy value of 0.99 in the 70% split, while RF produced the highest accuracy value of 0.97 in the 10-fold cross-validation. In Figure 6, SVM produced the highest Recall value of 0.99 in the 70% split while in the 10-fold cross-validation, RF outperformed other classifiers by having a value of 0.97. Figure 7 shows the comparison of F1_score where SVM outperformed other classifiers with a value of 0.99 in the 70% split while RF produced the best value of 0.97 in the 10-fold cross-validation. In Figure 8, SVM produced a Precision value of 0.99 in the 70% split while RF produced the best value of 0.97 in the 10-fold cross-validation.

Table 2: Result of the performance metrics of the algorithms in Python.

	Accuracy		Recall		F1-Score		Precision	
	10-fold split	**70% split**	**10-fold split**	**70% split**	**10-fold split**	**70% split**	**10-fold split**	**70% split**
DT	0.94	0.96	0.93	0.96	0.93	0.96	0.93	0.96
SVM	0.96	**0.99**	0.95	**0.99**	0.95	**0.99**	**0.96**	**0.99**
MLP	0.93	0.98	0.92	0.97	0.92	0.97	0.93	0.97
NB	0.937	0.93	0.886	0.93	0.911	0.926	0.945	0.924
RF	**0.97**	0.96	**0.97**	0.95	**0.97**	0.95	**0.96**	0.96

Summary of the results

Table 1 and Table 2 show the summary of all the values gotten from the analysis of the dataset under the 10-fold cross-validation and 70% split in the WEKA and Jupyter Notebook environments.

Conclusion

The early detection of breast cancer stands as a leading factor that has significantly increased the survival rate in patients. The successful integration of ICT into the field of medical science has heralded the arrival of innovative technologies such as ML, deep learning, and AI. These technologies have unequivocally demonstrated their ability to provide faster and more efficient methods for detecting and predicting breast cancer, ultimately resulting in a marked increase in the survival rate of patients. This research aims to investigate the generalizability of SVM models for breast cancer detection by comparing its performance with other classifiers such as NB, RF, MLP, and DT under various programming environments and data splitting techniques. it was discovered that SVM performed the best under the 10-fold cross-validation and percentage split techniques in WEKA with accuracy values of 0.981 and 0.977 respectively. In Jupyter Notebook, SVM had the highest accuracy value of 0.99 in the 70% split, but in 10-fold cross-validation, RF outperformed all other algorithms with an accuracy value of 0.97. It is worth noting that the SVM was not too far off, as it had an accuracy value of 0.96. Hence, SVM is recommended to be leveraged as a built-in algorithm for medical applications, which would be helpful to medical practitioners or clinicians for the early detection of breast cancer. Future research may explore the impact of additional factors like dataset size, class imbalance, and feature engineering on the generalizability of the models. Additionally, the research could be extended to incorporate more cutting-edge deep learning architectures for a more comprehensive evaluation of model performance in breast cancer detection. This research utilizes stronger languages to emphasize the importance of the research and the potential impact of the research findings.

References

1. World Health Organization (2023) WHO & World Health Organization: WHO. Breast cancer.
2. Sun Y, Zhao Z, Zhang Y, et al. (2017) Risk factors and preventions of breast cancer. International Journal of Biological Sciences 13: 1387-1397.
3. Naji MA, Filali SE, Aarika K, et al. (2021) Machine learning Algorithms for breast cancer prediction and diagnosis. Procedia Computer Science 191 487-492.
4. Akbugday B (2019) Classification of breast cancer data using machine learning algorithms. 2019 Medical Technologies Congress (TIPTEKNO).
5. Shah C, Jivani A (2013) Comparison of data mining classification algorithms for breast cancer prediction. 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT).
6. Hoque NR, Das NS, Hoque NM, et al. (2024) Breast Cancer Classification using XGBoost. World Journal of Advanced Research and Reviews 21: 1985-1994.
7. Islam T, Sheakh MA, Tahosin MS, et al. (2024) Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI. Scientific Reports 14.
8. Dinesh P, Vickram AS, Kalyanasundaram P (2024) Medical image prediction for diagnosis of breast cancer disease comparing the machine learning algorithms: SVM, KNN, logistic regression, random forest and decision tree to measure accuracy. AIP Conference Proceedings.
9. Elsadig MA, Altigani A, Elshoush HT (2023) Breast cancer detection using machine learning approaches: A comparative study. International Journal of Power Electronics and Drive Systems/International Journal of Electrical and Computer Engineering 13: 736-745.
10. Strelcenia E, Prakoonwit S (2023) Effective feature engineering and Classification of breast cancer diagnosis: A Comparative study. BioMedInformatics 3: 616-631.
11. Chaurasiya S, Rajak R (2022) Comparative analysis of machine learning algorithms in breast cancer classification. Research Square (Research Square).
12. Guleria K, Sharma A, Lilhore UK, et al. (2020) Breast cancer prediction and classification using supervised learning techniques. Journal of Computational and Theoretical Nanoscience 17: 2519-2522.
13. Ibeni WNLWH, Salikon MZ .M, Mustapha A, et al. (2019) Comparative analysis on Bayesian classification for breast cancer problem. Bulletin of Electrical Engineering and Informatics 8.
14. Keleş MK (2019) Breast cancer prediction and detection using data mining classification algorithms: A comparative study. Tehnicki Vjesnik-technical Gazette 26.
15. Hicks SA, Strümke I, Thambawita V, et al. (2022) On evaluation metrics for medical applications of artificial intelligence. Scientific Reports 12.

16. Fatima N, Liu L, Sha H, et al. (2020) Prediction of breast cancer, comparative review of machine learning techniques, and their analysis. IEEE Access 8: 150360-150376.

17. Mohammed SA, Darrab S, Noaman SA, et al. (2020) Analysis of breast cancer detection using different machine learning techniques. In: Communications in Computer and Information Science 108-117.

18. Mosayebi A, Mojaradi B, Naeini AB, et al. (2020) Modeling and comparing data mining algorithms for prediction of recurrence of breast cancer. PLoS One 15: e0237658.

DOI: 10.36959/643/310

Copyright: © 2024 Olorunshola OE, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

SCHOLARS.DIRECT