Sistemasi: Jurnal Sistem Informasi
Volume 13, Nomor 3, 2024: 1178-1187
ISSN:2302-8149
e-ISSN:2540-9719

Analysis and Visualization of Breast Cancer Prediction through Machine Learning Models

[1]Ayepeku .O. Felix*, [2]Omosola .J. Olabode, [3]Ayeni .J. Kehinde
[1-2]Dept. of Mathematical and Computing Science, Thomas Adewumi University Oko-Irese
[3]Dept. Of Computer Science, Kwara State Polytechnic, Ilorin
email: olukayode.ayepeku@tau.edu.ng

(***received***: 17 April 2024, ***revised***: 22 April 2024, ***accepted***: 22 April 2024)

Abstract

This research presents an in-depth exploration of breast cancer prediction through the application of machine learning models, specifically focusing on Logistic Regression, K-Nearest Neighbors, Support Vector Classifier, 'Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier, AdaBoost Classifier, and XGBoost Classifier. The study utilizes a comprehensive dataset comprising clinical features extracted from Kaggle. Various algorithms are employed, and a meticulous analysis of precision, recall, F1-score, and accuracy is conducted to assess model performance. Through advanced visualization techniques and statistical analysis, the research provides insights into the effectiveness of machine learning models in predicting breast cancer. The outcomes of this study aim to contribute valuable knowledge to the field of medical diagnostics, emphasizing the importance of machine learning methodologies in enhancing breast cancer prediction and classification.

Keywords: Breast Cancer, Efficiency, Performance measurement, Analysis, Visualization

1. Introduction

A malignant tumor discovered in breast tissue provides the basis for the diagnosis of breast cancer. A malignant tumor is a specific kind of tumor that can spread to neighboring cells or potentially across the body. Men and women can both get breast cancer, although women are more likely to have it. [reddy 2022). In 2020, 685 000 people worldwide died from breast cancer, accounting for 2.3 million new diagnoses. Breast cancer is the most common cancer worldwide, with 7.8 million women alive as of the end of 2020 who had received a diagnosis during the previous five years. All across the world, breast cancer affects women at any age after adolescence; however, its prevalence rises with age. From the 1930s through the 1970s, when surgery alone was the main form of therapy, there was minimal improvement in the death rate from breast cancer (radical mastectomy). Survival rates started to rise in the 1990s as nations implemented early detection systems for breast cancer that were connected to all-encompassing treatment regimens that included efficient (Rautalin, Jahkola, and Roine 2022) of the various oncology case categories; 11.6% involved breast cancer, with women accounting for 24.2% of those cases. Any new hard mass or lump in the breast tissue is an indication of breast cancer. But not every protrusion is malignant. Cancerous masses can be seen using mammography. Only 78% of women with cancer receive a correct diagnosis from a mammogram. (Garg and Gupta, 2020). Breast cancer is still a major public health issue, which is driving the demand for sophisticated diagnostic methods and therapeutic approaches. The nexus of technology and healthcare has made novel ways possible in response to this problem, such as the use of machine learning algorithms in the study of breast cancer. These algorithms show promise in improving tailored medicines, forecasting treatment results, and improving diagnostic accuracy through the use of computational approaches. This use of machine learning in breast cancer research represents a viable direction for enhanced patient care, providing an analytical viewpoint to support conventional approaches. In light of this, investigating the function of machine learning algorithms in the analysis of breast cancer becomes crucial for the development of precision medicine in oncology.

This study compares the results and interprets the performance evaluation of eight methods, which resulted in Random Forest and XGBoost as the highest-ranked methods for diagnosing breast cancer patients as compared to the other six methods. The rest of the content of this paper is organized in the following order: In Section 2, we discussed related work about machine learning and breast cancer. We discussed research methodology in Section 3, and in Section 4, we discussed data analysis and interpretation. In the last section, we conclude this paper and present possible future work.

2. Literature Review

Breast cancer affects most women worldwide and is the second-most common cause of death for them. However, if the disease is detected early and adequately treated, the condition can be recovered. By allowing patients to get prompt medical care, early identification of breast cancer can greatly improve prognosis and survival rates. Precise benign growth characterization can also assist patients in avoiding needless medical procedures. Support Vector Machine (SVM), Random Forest, XGBoost, K-Nearest Neighbor (KNN), and Logistic Regression are the main topics covered in this article. As part of the engagement, the dataset will be evaluated and analyzed, and a model will be created. In order to identify BC, ML (machine learning) techniques have been used in a lot of research in the healthcare industry recently. Considering that the algorithms provide Other scientists have used the algorithms to solve challenging issues since they produce good results. A CNN algorithm was utilized to recognize and diagnose invasive ductal carcinoma in breast cancer images, and it achieved an accuracy of about 88%. In 2020, Mutasa et al. Moreover, it is often used in the medical field to forecast and identify anomalous occurrences in order to better understand terminal diseases like cancer. (2020, Dong and Inoue). Many studies have looked at the use of genetics and imaging in breast cancer screening. Furthermore, no study that we are aware of has integrated the application of these two techniques. The authors of Sunardi, Yudhana, and WindraPutri (2022) gave a summary of the many approaches used in histological image analysis (HIA) for breast cancer identification. Numerous convolutional neural network (CNN) designs serve as the foundation for these techniques.

The kind of dataset that each author used helped to classify their work. Everything was set up with the most current occurrence at the top and in reverse chronological order. The results of this investigation suggest that ANNs were first used in the field of HIA approximately around the middle of 2012. PNNs and ANNs were the two types of algorithms that were used most frequently. However, morphological and textural characteristics were important in the process of extracting features. It is clear that using deep convolutional neural networks for early diagnosis and detection of breast cancer improves treatment outcomes for patients. Several different algorithms were used in the process of making NCD forecasts. An overview of the several techniques for histological image analysis (HIA) in breast cancer diagnosis was given by the authors in Fatima et al. 2020. Numerous convolutional neural network (CNN) designs serve as the foundation for these techniques. The kind of dataset that each author used helped to classify their work. Everything was set up with the most current occurrence at the top and in reverse chronological order. The results of this investigation suggest that ANNs were first used in the field of HIA approximately around the middle of 2012. PNNs and ANNs were the two types of algorithms that were used most frequently. However, morphological and textural characteristics were important in the process of extracting features. It is clear that using deep convolutional neural networks for early diagnosis and detection of breast cancer improves outcomes. Authors Wang et al. (2023) discussed the possible applications of insect-based NIC diagnostic algorithms in the diagnosis of cancer and diabetes. The authors stated that ovarian, lung, prostate, and breast cancers were all effectively identified. Directed ABC combined with neural networks improves a breast cancer diagnosis. Additionally, the authors created a highly useful method for determining leukemia and diabetes. They reasoned that combining NICs with traditional categorization methods produces more dependable and positive outcomes. They emphasized the necessity for more study on diabetes and the identification of illnesses at various stages. Khan et al. (2020) contain

According to the authors' findings, NNs might be a helpful tool for classifying cancer diagnoses, especially in the early stages of the illness. Their findings indicate that certain NNs have shown promise in the identification of cancerous cells. However, the imaging technique requires a significant amount of computer power to prepare the pictures. In this part, we look at how AI and CNNs can make things easier. By boosting low-contrast features, cutting noise, eliminating artifacts, and improving image registration, CNNs and AI can enhance the quality of medical images. Additionally, they can help in ROI recognition, imaging, and segmentation to enable accurate evaluation and identification of anatomical features or lesions. In addition to employing contrast-limited adaptive histogram equalization (CLAHE) approaches to enhance image quality, artificial intelligence (AI) algorithms may also modify levels of brightness, contrast, and intensity. CNNs may also identify and eliminate common visual artifacts, guaranteeing precise interpretation. AI techniques enhance image alignment, and segmentation and ROI detection allow for accurate diagnosis and study of certain areas. Finally, beyond the point of initial acquisition, super-resolution imaging employing CNNs can enhance the quality and resolution of pictures.

AI-driven super-resolution methods leverage deep learning models to generate high-resolution images from low-resolution inputs, resulting in enhanced performance. In the subject of smart health, evolutionary computing techniques like genetic algorithms, classifiers, and support vector machines, as well as computational intelligence techniques like fuzzy systems, artificial neural networks, and swarm intelligence, are helpful tactics. Al-Masni, Al-Antari, and others (2018). Research in Khan et al. (2020) indicates that the proposed CNN Improvements for Breast Cancer Classification (CNNI-BCC) model aids physicians in detecting breast cancer. The recommended approach makes use of a deep-learning neural network system that has been trained to categorize different forms of breast cancer. Based on information from 221 actual patients, the results had a 90.50 percent accuracy rate. This model has the ability to classify and identify breast lesions. The examination of this model shows that it can assess the conditions of afflicted individuals during the detection stage, suggesting that it is an improvement over earlier techniques. (Matsuo & Associates, 2020). To ascertain the similarities and differences between SVM, logistic regression, naive Bayes, and random forest, Mallika and Suresh Babu (2023) conducted a comparison. The breast cancer dataset from Wisconsin is utilized for comparison. (Allugunti 2022) Based on the assessment findings, the random forest approach yielded the highest accuracy (99.76%) with the lowest amount of error. Every experiment may be carried out repeatedly thanks to the use of the Anaconda Data Science Platforms. A strategy for grouping breast cancer patients into distinct subgroups was proposed by the study's authors, Meenalochini and Ramkumar (2021). Data on prognostic breast cancer from the Wisconsin Diagnosis and Analysis. Next, a neural network approach is used to classify the various forms of breast cancer, with particular focus on the multilayer perceptron (MLP) and the back-propagation neural RBF. The input layer of the neural network is represented by the nine attributes in this dataset. The input data will be divided into two categories by the neural network: benign and malignant cancer. The algorithm created and tested on the database obtained 97% repeatability of classification using the RBF neural network. The authors (Twala and Molloy 2022) assessed and contrasted two distinct Bayesian classifiers, namely tree-augmented naive Bayes and Markov blanket estimating networks, in order to construct an ensemble model for the prediction of breast mass severity. The authors have demonstrated that these methods, which are based on Bayesian classifiers, are a competitive substitute for alternative approaches in the field of medicine. The authors (Altinok and Guvenis 2023) have chosen to use Bayesian networks (BN) in the field of emergency medicine, where it is a useful approach due to its potent symbolism and management of ambiguity and where several alternatives are feasible depending on the data that has been provided. The reason why Bayesian networks are so effective is because of their symbolic representation. Using a variety of classification approaches, the random forest (RF) classifier is an ensemble approach. A decision tree may be used to implement each of these. Using several decision trees leads to improved classification accuracy (Jinbo et al., 2023). To put it simply, the RF is an ensemble classifier that combines many decision trees to increase efficiency and prediction accuracy. Researchers developed an RF-based classifier in Macaulay 2020. They trained their algorithm on two datasets, and the results are promising: they obtained good performance and high accuracy.

The WDBC has been used to compare three classifiers: k-closest neighbor (KNN), radial basis function (RF), and nearest neighbor (NB). Training and evaluating these classifiers on the previously given dataset determines how accurate they are in predicting breast cancer tumors. The study's authors discovered that while every classifier they evaluated produced detection accuracy rates greater than 94%, KNN performed the best. It outperforms both the NB and RF classifiers in terms of accuracy. The KNN classifier has better precision and F1-score in addition to its greater accuracy (Rixen et al., 2023). Price and Lindquist state that when feature selection approaches are used, the ANN classifier performs better. Using a small dataset of 275 samples, the authors of Al-Azzam and Shatnawi (2021) assess two machine learning classifier models: RF and extreme gradient boost (XGBoost). The authors contend that a large dataset is required to confirm their conclusions since using a restricted dataset may reflect inaccurate results. This is true even if their results indicate that RF has surpassed XGBoost in terms of accuracy in diagnosing breast cancer. Nine classification models, including LR, Gaussian naive Bayes, RBF SVM, linear SVM, DT, RF, XGBoost, KNN, and gradient boosting, were studied in a recent study.

The models are trained and tested using the Wisconsin Diagnosis Cancer Dataset. We may infer from the data that the best approach for supervised learning is KNN, while the best approach for semi-supervised learning is LR (Huang and Chen 2022). The ensemble learning methodology is one of the least complete ways to offer a trade-off between variance and bias. Numerous studies have demonstrated that combining separate classifiers to create an aggregated classification model can enhance classification performance. The three basic techniques for ensemble classification are stacking, boosting, and bagging. According to Kim, Kang, and Sohn (2021), the stacking approach combines the output of many classification models into a single one.

3. Research Methodology

The research strategy and datasets utilized in the statistical analysis and performance assessment of patients with breast cancer are presented in this chapter. In order to accomplish the study's goal, it also illustrates the data analysis methods, processes, and statistical handling of the data.

3.1 Proposed Methodology

These eight classification models—Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier, AdaBoost Classifier, XGBoost Classifier, K-Nearest Neighbors, and Logistic Regression—were applied in this study to diagnose breast cancer. Remember that the goals chosen and the dataset used may affect the values of some hyperparameters and other implementation-specific choices. Blood tests, magnetic resonance imaging, computed tomography, mammography, positron emission tomography, and genetic analysis are all restricted in terms of early diagnosis and prognosis. Sophisticated techniques could require invasive operations, erroneous diagnosis, and specific training.

A few advantages of machine learning technology include early diagnosis, improved accuracy, customized therapy, risk assessment, data integration, predictive prognosis, and drug development, all of which may improve cancer detection and prognosis. These methods aid in speeding up the arduous and prone-to-mistakes present process of identifying promising drugs and treatments for cancer treatment.

Differential X-ray absorption is used in mammography, a form of X-ray imaging, especially for breast tissue. It is mostly used for breast cancer screening, diagnosis, and early detection. Additionally, it detects smaller anomalies like cancer and microcalcifications. However, ultrasonography uses high-frequency sound waves to create images of breast tissue. It is commonly used in combination with mammography to improve its diagnostic performance by providing additional information about the type of breast abnormalities and helping to distinguish between solid masses and fluid-filled cysts. Breast biopsies can also be guided by ultrasonography.

Thermography, sometimes called thermal imaging, is the process of recording the patterns of heat emitted by the body's surface. Using a specialized camera, it analyzes the surface temperature of the skin to reveal temperature variations in areas with greater blood flow, such as the presence of tumors. While thermography has been studied as a non-invasive screening method for breast cancer, its use as a stand-alone method is restricted due to problems with accuracy and heterogeneity in result interpretation. It can be used as an adjuvant tool in certain situations or for monitoring breast health. Mammography is the gold standard for breast cancer screening and is strongly recommended in many countries. As an adjuvant method, ultrasonography is widely used, especially in situations with dense breast tissue. Due to problems with specificity and sensitivity. The ideal screening method is chosen by medical specialists in collaboration with patients, taking into account the patient's age, clinical conditions, breast density, and risk factors. Two examples of ensemble classifiers that use different ensemble approaches are the RF classifier and the KNN classifier. The bagging method is crucial for RF. There are tree-based classifiers in this collection. The stack classifier is a stacking-based classifier that receives the results of earlier classification models as input. Figure 1 shows our recommended methodology. This graphic's data is taken from the dataset. To make the breast cancer dataset usable, load the data into it. Sort the data according to features (X) and labels (Y). To ensure that each factor influences the algorithms in the same manner, normalize or standardize the features before conducting a quantitative analysis of them. Utilizing the data, create training and test sets. Thirty percent goes toward testing, and seventy percent goes toward training, as an illustration of a typical distribution. The techniques utilized for model evaluation and training are random forest (RF), decision tree (DT), logistic regression (LR), k-nearest neighbors (KNN), AdaBost (ADB), XGBoost (XGB), Gradient Boosting Classifier (GBM), and linear support vector classifier (linear SVC). Each approach's effectiveness is evaluated.

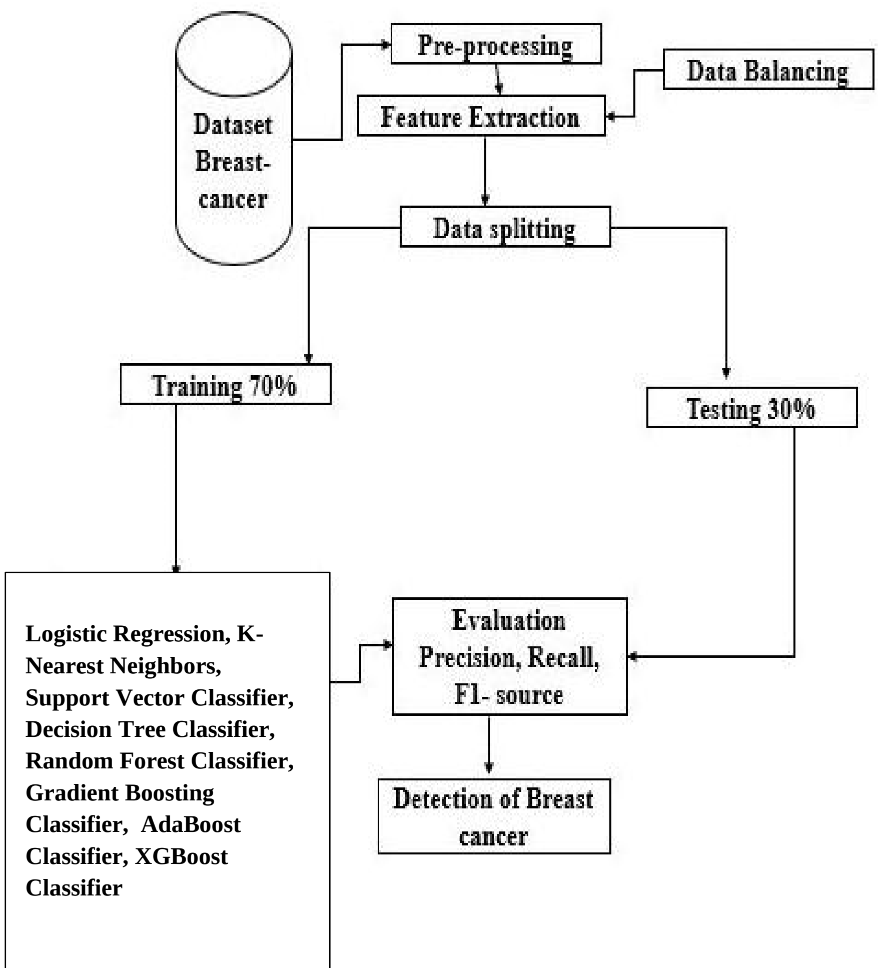


Figure 1. Proposed methodology

In figure 1 above, the Method of modeling study aims to suggest eight machine learning techniques: LR, KNN, SVC, RF, GBM, ADB, DT, XGB in that order. Finding an appropriate main model with the highest possible prediction accuracy is the study's goal. We included accuracy and the index of F measure metric while choosing the main model. As a classification model, we employed the entire eight Algorithms model to determine the best to predict whether breast cancer will be benign or malignant. The ratios of the datasets are dispersed in smaller groups, such 70% to 30%. The study suggests using 30% of the data for testing and 70% of the data for the training set. First, we used training data to train a classification model, and then we used test data on a learned

3.2 Description of Dataset:

This study uses dataset relevant to Breast Cancer from Kaggle. Attributes of the datasets are relevant to the breast cancer patient. The dataset is comprises of 563 given a separate ID for each subject in the dataset.

S/No.	Characteristics	Code
1	radius mean	R
2	texture mean	T
3	area mean	AR
4	parameter mean	PM
5	smoothness mean	SM
6	compactness means	CM
7	concave mean	C
8	concavity mean	CN
9	Symmetry means	SM
10	Fractal dimension mean	FD

Table 1. Characteristics of dataset

Each characteristic represents a different aspect of the dataset, such as measurements or statistical properties. The corresponding code provides a shorthand reference for these characteristics, making it easier to work with and analyze the dataset. For example, if we were discussing the radius mean, we could simply refer to it as "R" in our analysis, thanks to the assigned code. This table, labeled as "Table 1: Characteristics of dataset," serves as a quick reference guide for understanding and interpreting the dataset's attributes.

3.2 Experiments and Interpretations

We will go into more detail about the comparative study of the experimental outcomes of eight machine learning models in this part. There are 569 counts altogether in the dataset categorization, with 212 counts classified as cancerous and 357 counts as benign. The numbers are spread throughout several classes and sets. The goal is to analyze the database using all eight machine learning techniques. We integrated accuracy and the F-measure metric index to choose the main model. The classifier's measures determine accuracy, which is determined by averaging true positive and false positive items. The F-measure matrix is determined by averaging the harmonic sum of recall and precision.

 ISSN:2302-8149
e-ISSN:2540-9719

Classification mode	LR	KNN	SVC	RF
Accuracy	0.91	0.92	0.91	**0.93**
precision	0.91	0.92	0.92	**0.93**
Recall	0.90	0.91	0.90	**0.92**
F1-score	0.90	0.91	0.91	**0.92**

Table 2. Interpretations

Classification mode	GBM	ADB	DT	XGB
Accuracy	0.91	0.91	0.91	**0.93**
precision	0.92	0.91	0.91	**0.93**
Recall	0.90	0.90	0.90	**0.92**
F1-score	0.91	0.90	0.90	**0.92**

Table 3. Interpretations

In discussing the performance metrics of different classification models, we can reference Table 2 and Table 3 for interpretations. Comparing the performance of Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), and Random Forest (RF) as shown in Table 2, we observe varying accuracies ranging from 0.91 to 0.93. The precision, recall, and F1-score also display fluctuations across these models. Similarly, Table 3 provides insights into the performance of Gradient Boosting Machine (GBM), AdaBoost (ADB), Decision Tree (DT), and XGBoost (XGB). Here, we can observe similarities and differences in accuracy, precision, recall, and F1-score compared to the models in Table 2.

These tables aid in understanding the relative strengths and weaknesses of each classification model, facilitating informed decisions regarding model selection for the task at hand.

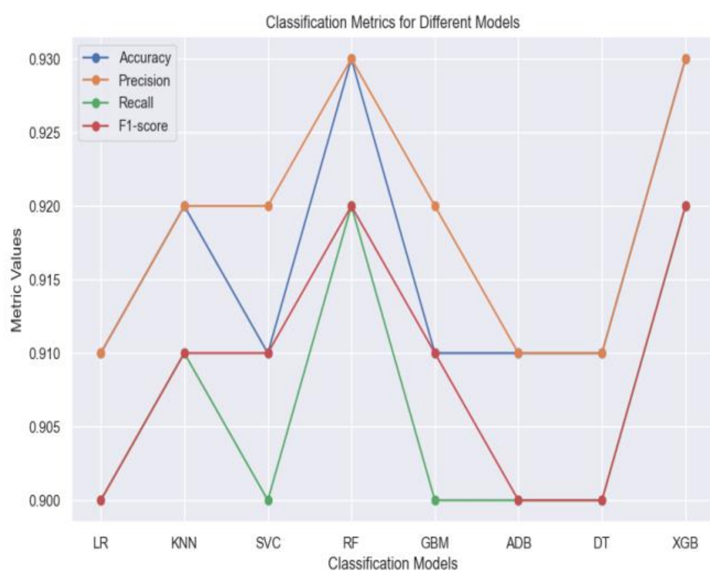
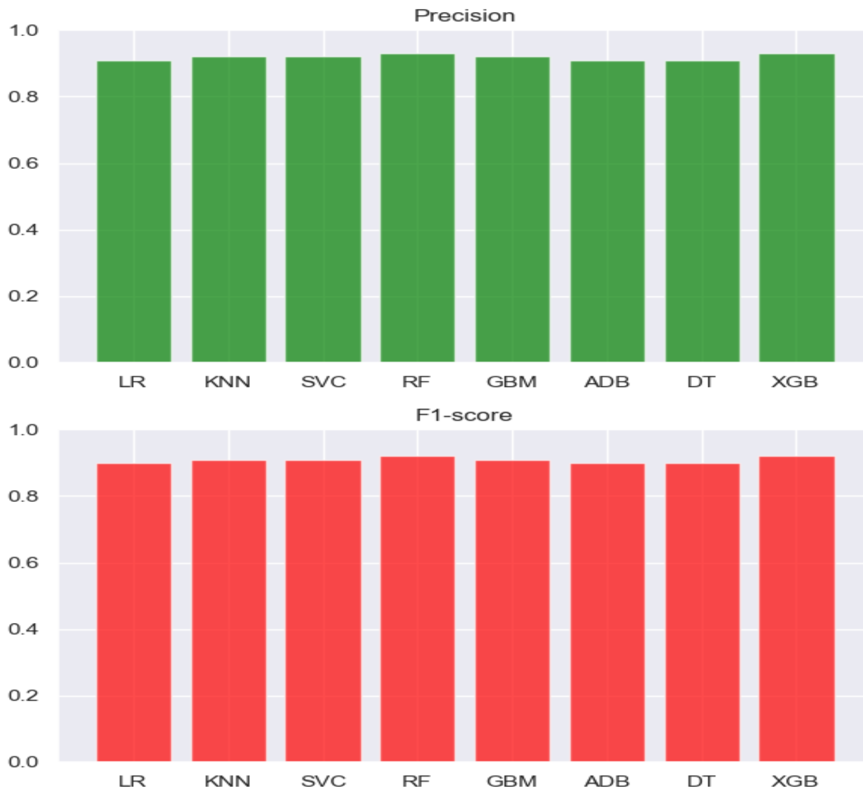


Figure 3. Accuracy and F-measure Metric values

ISSN:2302-8149
e-ISSN:2540-9719



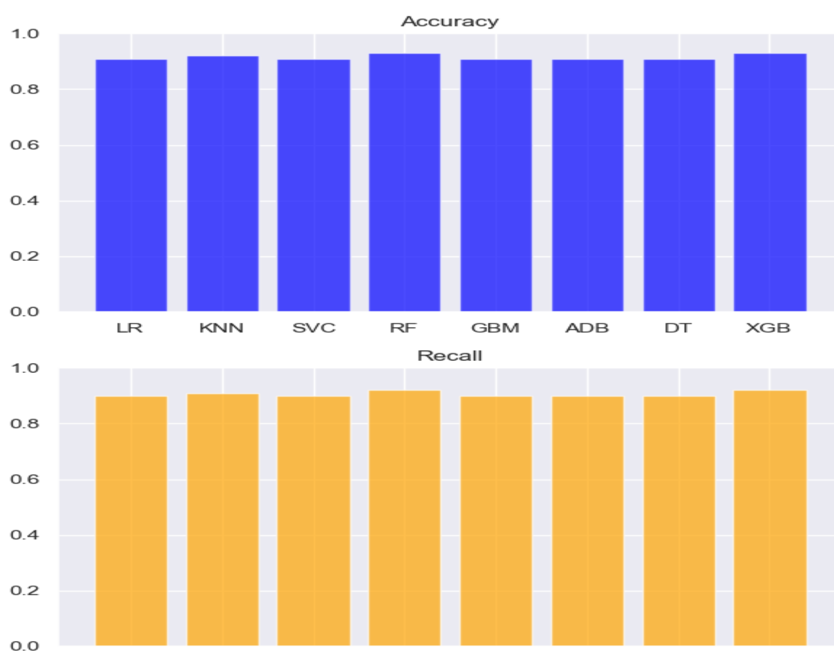


Table 2 and Figure 3 demonstrate that, out of the eight algorithms, Random Forest and XGBoost have been the most successful in forecasting the number of accurately diagnosed patients with malignant and benign tumor

ISSN:2302-8149
e-ISSN:2540-9719

cancer from a sample of 569 cases. This suggests that these two algorithms are the most suitable for patient diagnosis prediction. Overall, it is determined that Random Forest and XGBoost are superior to the other 6 methods—GBM, ADB, KNN, DT, LR, and SVC—and effective Natural Artificial Language that can diagnose cancer patients in the categories of Malignant and Benign. This is true even though the accuracy measurement of the Decision Tree algorithm is nearly equal.

4. Conclusion

Eight classifiers were used in this investigation once the data set's nominal values were converted to numerical values. Logistic Regression, K-Neighbors Classifier, SVC, Random Forest Classifier, Decision Tree Classifier, Gradient Boosting Classifier, AdaBoost Classifier, and XGBoost are the eight approaches. The goal of applying machine learning algorithms to datasets was to assess how well natural language functions and determine whether approach is more effective in diagnosing breast cancer. One class was designated as "Benign," and zero as "Malignant." The information, which was obtained from a reliable source, has been utilized in many research to determine the incidence of breast cancer. This study included 32 characteristics and 569 participants in its datasets. To compare the outcomes, it also used cross validation and the F-measureEven if the study has limitations due to the paucity of data on breast cancer, the results are more beneficial since they not only advance humankind but also emphasize the critical role that technology plays in medical research. According to the study's findings, machine learning algorithms are a reliable method of generating breast cancer detection results that can also be used for other medical problems. The study also analyzes and interprets the performance evaluation of the eight approaches, comparing and ranking Random Forest and XGBoost as the best techniques for breast cancer patient diagnosis. Furthermore, it's likely that we may obtain more precise and effective findings if we apply similar machine learning techniques to larger data sets.

REFERENCE

[1] Reddy, A. (2022). Support Vector Machine Classifier For Prediction Of Breast Malignancy Using Wisconsin Breast Cancer Dataset. *Journal of Artificial Intelligence, Machine Learning and Neural Network*, *21*(January), 1–8. https://doi.org/10.55529/jaimlnn.21.1.8.

[2] Rautalin, M., Jahkola, T., & Roine, R. P. (2022). Breast Reconstruction– Prospective Follow up on Breast Cancer Patients’ Health-Related Quality of Life. *World Journal of Surgery*, *46*(4), 836–844. https://doi.org/10.1007/s00268-021-06426-4.

[3] Gupta, P., & Garg, S. (2020). Breast Cancer Prediction Using Varying Parameters of Machine Learning Models. *Procedia Computer Science*, *171*, 593–601. https://doi.org/10.1016/j.procs.2020.04.064.

[4] Mutasa, S., Chang, P., Nemer, J., Van Sant, E. P., Sun, M., McIlvride, A., Siddique, M., & Ha, R. (2020). Prospective Analysis Using a Novel CNN Algorithm to Distinguish Atypical Ductal Hyperplasia From Ductal Carcinoma in Situ in Breast. *Clinical Breast Cancer*, *20*(6), e757–e760. https://doi.org/10.1016/j.clbc.2020.06.001.

[5] Dong, L., & Inoue, K. (2020). Diagnosis of Breast Cancer from Mammogram Images Based on CNN. *Journal of the Institute of Industrial Applications Engineers*, *8*(4), 117–121. https://doi.org/10.12792/jiiae.8.117.

[6] Sunardi, S., Yudhana, A., & WindraPutri, A. R. (2022). Mass Classification of Breast Cancer Using CNN and Faster R-CNN Model Comparison. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, *August*. https://doi.org/10.22219/kinetik.v7i3.1462.

 ISSN:2302-8149
e-ISSN:2540-9719

[7] Fatima, N., Liu, L., Hong, S., & Ahmed, H. (2020). Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis. *IEEE Access*, *8*, 150360–150376. https://doi.org/10.1109/access.2020.3016715.

[8] Wang, Y., Tang, L., Chen, P., & Chen, M. (2023). The Role of a Deep Learning-Based Computer-Aided Diagnosis System and Elastography in Reducing Unnecessary Breast Lesion Biopsies. *Clinical Breast Cancer*, *23*(3), e112–e121. https://doi.org/10.1016/j.clbc.2022.12.016.

[9] Khan, F., Khan, M. A., Abbas, S., Athar, A., Siddiqui, S. Y., Khan, A. H., Saeed, M. A., & Hussain, M. (2020). Cloud-Based Breast Cancer Prediction Empowered with Soft Computing Approaches. *Journal of Healthcare Engineering*, *2020*(May), 1–16. https://doi.org/10.1155/2020/8017496.

[10] Matsuo, K., Tanabe, K., Hayashi, M., Ikeda, M., Yasaka, M., Machida, H., Shida, M., et al. (2020). Utility of Comprehensive Serum Glycopeptide Spectra Analysis (CSGSA) for the Detection of Early Stage Epithelial Ovarian Cancer. *Cancers*, *12*(9), 2374. https://doi.org/10.3390/cancers12092374.

[11] Mallika, M., & Babu, K. S. (2023). Breast Cancer Prediction Using Machine Learning Algorithms. *International Journal of Science and Research (IJSR)*, *12*(10), 1235–1238. https://doi.org/10.21275/sr231015173828.

[12] Allugunti, V. R. (2022). Breast Cancer Detection Based on Thermographic Images Using Machine Learning and Deep Learning Algorithms. *International Journal of Engineering in Computer Science*, *4*(1), 49–56. https://doi.org/10.33545/26633582.2022.v4.i1a.68.

[13] Meenalochini, G., & Ramkumar, S. (2021). Survey of Machine Learning Algorithms for Breast Cancer Detection Using Mammogram Images. *Materials Today: Proceedings*, *37*, 2738–2743. https://doi.org/10.1016/j.matpr.2020.08.543.

[14] Twala, B., & Molloy, E. (2022). On Effectively Predicting Autism Spectrum Disorder Using an Ensemble of Classifiers. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4215300.

[15] Altinok, O., & Guvenis, A. (2023). Interpretable Radiomics Method for Predicting Human Papillomavirus Status in Oropharyngeal Cancer Using Bayesian Networks. *Physica Medica*, *114*(October), 102671. https://doi.org/10.1016/j.ejmp.2023.102671.

[16] Jinbo, K., Fujita, T., Kasahara, R., Jinbo, R., Kisara, S., Onobe, J., Kimijima, I., et al. (2023). The Effect of Combined Risk Factors on Breast Cancer-Related Lymphedema: A Study Using Decision Trees. *Breast Cancer*, *30*(4), 685–688. https://doi.org/10.1007/s12282-023-01450-9.

[17] Al-Azzam, N., & Shatnawi, I. (2021). Comparing Supervised and Semi-Supervised Machine Learning Models on Diagnosing Breast Cancer. *Annals of Medicine and Surgery*, *62*(February), 53–64. https://doi.org/10.1016/j.amsu.2020.12.043.

[18] Huang, Z., & Chen, D. (2022). A Breast Cancer Diagnosis Method Based on VIM Feature Selection and Hierarchical Clustering Random Forest Algorithm. *IEEE Access*, *10*, 3284–3293. https://doi.org/10.1109/access.2021.3139595.

[19] Kim, J., Kang, J., & Sohn, M. (2021). Ensemble Learning-Based Filter-Centric Hybrid Feature Selection Framework for High-Dimensional Imbalanced Data. *Knowledge-Based Systems*, *220*(May), 106901. https://doi.org/10.1016/j.knosys.2021.106901.