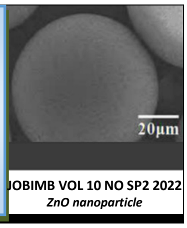
JOBIMB, 2022, Vol 10 No 2, 29-34
https://doi.org/10.54987/jobimb.v10i2.754

HIBISCUS PUBLISHER

JOURNAL OF BIOCHEMISTRY, MICROBIOLOGY AND BIOTECHNOLOGY

Website: https://journal.hibiscuspublisher.com/index.php/JOBIMB



Evaluation of Some Selected Breast Cancer Classification Algorithms in Nigeria

Aminu Suleiman[1]*, Eviwiekpaefe, A.E.[1], Abdulhadi Yakubu[2], Garba Uba[3] and Zainab Yahaya[4]

[1]Department of Computer Science, Nigerian Defence Academy, No 6 Ahmadu Bello Way, Sabon Gari 200283, Kaduna, Nigeria.
[2]Department of Public Health, College of Health Sciences, Jigawa State Polytechnic, Dutse, PMB 7040, Nigeria.
[3] Department of Science Laboratory Technology, College of Science and Technology, Jigawa State Polytechnic, Dutse, PMB 7040, Nigeria.
[4] National Biotechnology Development Agency (NABDA), Lugbe, Abuja. P.M.B 5118, Wuse, Abuja, Nigeria.

*Corresponding author:
Aminu Suleiman
Department of Computer Science,
Nigerian Defence Academy.
No 6 Ahmadu Bello Way,
Sabon Gari 200283,
Kaduna,
Nigeria.
Email: suleimanaminu58@gmail.com

HISTORY

Received: 25th May 2022
Received in revised form: 14th of May 2022
Accepted: 18th of June 2022

KEYWORDS

Breast Cancer
Pre-Malignant
Support Vector Machine (SVM)
K- nearest neighbour (KNN)
Decision Tree (DT)

Abstract

Breast Cancer (BC) is a prevalent disease that affects mostly women in the world. According to the World Health Organization (WHO), BC represent about 25 percent of all cancers in women with 685 000 deaths in 2020. An early detection of this disease can greatly increase the chances of taking the right decision on a successful treatment plan. This resulted in the need of new research avenues most especially in a country like Nigeria where there is low awareness of the disease and late presentation of BC by patients is normal. To achieve this, Support Vector Machine (SVM), KN Neighbor (KNN) and Decision Tree (DT) was used on a local dataset obtained from Ahmadu Bello University Teaching Hospital Zaria to provide some effective diagnostic capabilities. The dataset was classified into three classes (Benign, Pre-malign and Malign) and the SVM obtained a good classification accuracy of (99.2%). Late presentation of breast cancer is normal because of low awareness of the disease in the country therefore more awareness of the disease is highly recommended and women above the age of 34 years should always go for the breast cancer screening at least once a year with or without sign, sickness or symptoms.

INTRODUCTION

Breast cancer (BC) is a disease that happens when cells in the breast get uncontrollably large. Breast cancer is the second leading cause of mortality among women, after lung cancer. According to the World Health Organization (WHO), it is the most common illness in women worldwide, with almost 2.3 million new cases diagnosed and 685 000 deaths in 2020, accounting for approximately 25% of all malignancies in women [1]. Breast cancer is a form of cancer that arises from breast tissue; it is most frequent in women and is one of the most researched diseases, owing to its prevalence and its history of high mortality rate (second to lung cancer). However, it occurs in males also [2].

However, in this research, Machine Learning will be introduced using Support Vector Machine (SVM), K Nearest Neighbor (KNN) and Decision Tree (DT) classifier on a local dataset obtained from Ahmadu Bello University Teaching Hospital Zaria (ABUTH) ecology department for Breast Cancer classification.

Many researchers used Machine Learning approach on Breast Cancer prediction with the majority using a Wisconsin Breast Cancer Dataset (WBCD) for their classification which classifies Breast Cancer into two classes (benign and malignant), the WBCD differs from our local data due to social, physical, and mental factors that can lead to either good or poor performance of their model if adopted in Nigeria [3]. Subsequently, Momenimovahed et al., [4], used SPSS to conduct their research. They concluded that significant epidemiological and biological differences between breast cancer patients in Nigeria and in western populations and suggest the importance for the establishment of an effective breast cancer screening program

uniquely tailored for the Nigerian population and no machine learning technique used in their research [5].
A remarkable amount of research has been done in the area of building models that assist in predicting different types of diseases and health related problems especially those related to Breast Cancer, using different machine learning algorithms. This section presents some of the works done in this area and their various outcomes [6].

In order to obtain breast cancer classifications and lead pathologists to find a systematic and objective prognosis [7], used Wisconsin Breast Cancer Dataset (WBCD) and divide it into two binary classes (benign cancer and malign cancer). They conducted a comparison between the two new implementations and evaluate their accuracy using cross-validation. They use two different classifiers: The NB (Naive Bayes) classifier and KNN (k-nearest neighbor) for breast cancer classification. Results show that KNN gives the highest accuracy (97.51%) with the lowest error rate then NB classifier (96.19 %). Also Wisconsin Diagnostic Breast Cancer (WDBC) was employed by [6,8–11], for their classification of breast cancer using machine learning techniques with different classifiers such as SVM, KNN, DT, ANN etc and a hybrid technique such as SVM and Computer Aided Design (CAD) and more.

There are considerable epidemiological and biological variations between blacks and whites with breast cancer, which has far-reaching consequences for the establishment of an efficient cancer screening program in Nigeria. Ntekim et al [12], did a study with the goal of describing the clinicopathologic aspects of breast cancer identified at their hospital and discussing the implications for cancer screening. They conclude that there are major epidemiological and biological differences between breast cancer in Lagos and in western populations that must be addressed in order to build an effective breast cancer screening program that is specifically adapted to the Nigerian population.

Nigerian researchers such as Johnson et al., [13–18], and Akinyemiju [19], used statistical packages such as SPSS to conduct research on the correlation between Nigerians patient's presentation of breast cancer and other countries they discovered that late presentation of breast cancer in Nigeria was normal, therefore, breast cancer continues to carry a poor prognosis in the country. They proposed the need for effective breast cancer screening programme uniquely designed for the Nigerian population. However, Adebayo et al [20], used Machine learning technique to identify if a patient is likely or unlikely to contract breast cancer based on previous patient history.

The goal of this research is to develop a model to accurately predict the class of breast cancer in people at an early stage, and to introduce additional parameter (i.e Pre-Malignant) in the breast cancer classification. The objectives of this study are to import and perform preprocessing on the dataset, to employ Machine Learning approach for the classification of breast cancer in Nigeria. And to classify breast cancer in three (3) classes' i.e. Benign, Pre-malign, and Malign.

MATERIALS AND METHODS

This methodology depicts the design of the method to be exploited to the experiment. It incorporates data collection, data preprocessing, data modeling, model evaluation and feature importance selection, as shown in **Fig. 1**.

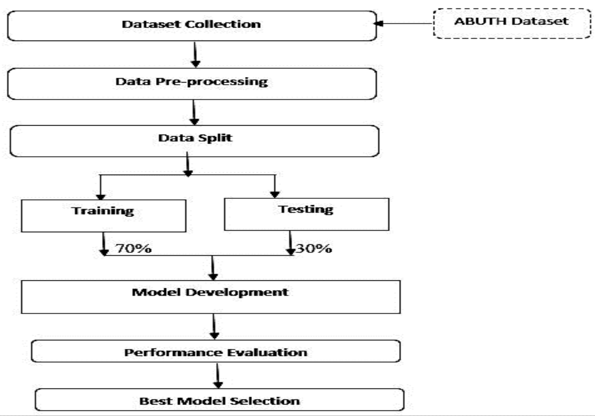


Fig. 1. Research methodology.

Data Collection

Breast cancer categorization can be divided into several schemata. Each of these factors influences therapy response and prognosis. Breast cancer categorization should contain all of these classification features, as well as other discoveries, such as physical exam indicators. The illness is classified fully by histopathological type, grade, stage (TNM), receptor status, and the presence or absence of genes as confirmed by DNA testing. Breast cancer records were selected as the source of data for this work. This dataset was collected from ABUTH Zaria Local Government Area of Kaduna State. The data was not in a digital form; instead, the information was raw data. We had to input the records into Microsoft Excel and saved the data in CSV format. The dataset considered contains: ID, Age, Gender, Marital Status, Education Level, Occupation, Cancer Location, Smoothness, Smoking History, Herp II, Histology, Estrogen Receptor (ER), Progesterone Receptor (PR), Family History, Classification including a target variable classified into a binary classification of Benign, Premalignant and Malignant as shown in **Table 1**.

Table 1. The training dataset.

S/N	Parameters	Values
1	Age	18-80 years
2	Gender	Male, Female
3	Marital Status	Single, Married, Divorced, Widowed
4	Education level	SSCE, OND, HND, Bsc, Msc, Phd
5	Cancer Location	Left, Right
6	Occupation	House wife, Civil Servant, Retired
7	Smoothness	0,1,2,3
8	Histology	Yes, No
9	Estrogen Receptor (ER)	Reactive, Non:Reactive
10	Progesterone Receptor (PR)	Reactive, Non:Reactive
11	Family History	Yes, No
12	Smoking History	Never, Former, Current
13	HERP II	Reactive, Non:Reactive
14	Class	Benign, Pre-malign, Malignant

Age

Age is an important risk factor in detecting/classifying breast cancer patient, it has been suggested that age at diagnosis is related to breast cancer survival, and it signifies the type of investigation that will be done for a patient (like mammography that can only be done for women above 34 years). It is of numerical value in the data set.

https://doi.org/10.54987/jobimb.v10i2.754

Gender
Gender plays a role in development since women are more likely than males to get breast cancer. Treatment options for male breast cancer include surgery with or without radiation therapy, chemotherapy, hormonal therapy, and others.

Marital Status
Previous studies have demonstrated that marital status could affect the survival outcome of several kinds of cancers, and it might act as an independent prognostic factor for overall survival in patients with breast cancer.

Education level/Occupation
Socioeconomic status is also connected with education level and employment, both of which can dramatically influence patients' perception of the tumor, affecting the degree of early identification, diagnosis, and therapy.

Cancer Location
This affects the biology and behavior of a *breast cancer patient* and sometimes the *treatment* plan. A small group of *patients* with *tumors located* in specific places are on high-risk. Some tumors are smaller but grow quickly, while others are larger and grow slowly this can also be determined by the location.

Smoothness or Breast Mass: Patients frequently accidentally recognize their breast tumor, which is solitary and firm with uneven edges and a less smooth surface. Most breast cancers appear as a painless lump, with just a minority exhibiting varying degrees of discomfort or tingling.

Histology
This is a very important factor in breast cancer classification, the histologic grade of a tumor can supply prognostic information in addition to that provided by LRD (localized, regional and distant) stage which is very important in determining the patient receptor status.

Estrogen Receptor (ER) & Progesterone Receptor (PR)
Breast cancer's estrogen receptor and progesterone receptor status is critical in determining therapy choices. Understanding the ER/PR status of the main tumor as well as any distant or recurring cancers can assist clinicians in ensuring that patients receive the proper treatment while avoiding the negative effects of a treatment that may not work.

Family History
Aside from old age, the most important risk factor for the occurrence of breast cancer in women is a family history. This study looks into how women's experiences with early-stage illness detection, diagnosis, and treatment differ.

Smoking History
This is a critical component in determining the risk of developing breast cancer. Women who started smoking before the age of 17 had a 24% greater risk of breast cancer, while those who started smoking between the ages of 17 and 19 had a 15% increased risk. Smoking for more than 10 years raised the chance of having breast cancer by 21%, whereas smoking for more than 30 years increased the risk somewhat (22%).

HERP II
HER2-positive breast cancer is a breast cancer that tests positive for a protein called human epidermal growth factor receptor 2 (HER2). This protein stimulates cancer cell proliferation. Experts advise testing for the presence of HER2 in all invasive breast cancers since the results have a substantial influence on treatment recommendations and decisions.

Classification
classification attribute is used to classify the type of tumour that is present in the breast i.e. Benign, Premalignant and malignant.

Data Pre-processing
Data preprocessing is a crucial stage in Breast Cancer methodology in which it involves handling missing values, noisy data and inconsistent data.

Table 2. Summary of the attribute with missing values.

Attribute	No of Missing Cells
Age	0
Sex	0
Marital Status	0
Occupation	0
Cancer Location	0
Smoothness	0
Histology	0
ER	0
PR	0
Family History	0
Smoking History	0
Herp2	0
Classification	0

Table 2 shows the summary of the dataset highlighting missing values indicating that there is no missing value, respectively, which will make it easier to train and test the dataset accordingly.

Selection Techniques Used for the Classification
Machine learning techniques used in this research work include Support Vector Machine (SVM), Decision Tree (DT), and K-nearest neighbors (KNN). These models were used for the classification and the results obtained will be compared for best model selection.

Support Vector Machine (SVM)
SVM is a supervised learning technique that has been widely used to analyze data for classification into their specific classes using a given label. In our case, we used SVM, DT, and KNN to classify breast cancer into 3 classes (Benign, Premalignant and Malignant). **Equation 1** depicts the SVM model [21].

$$f(x) = \beta_o + \sum_{i \in S} \alpha_i \binom{k}{2} (x, x_i) \quad \textbf{(Eqn. 1)}$$

Where
$f(x)$ = equation for predicting a new input using the dot product between the input (x) and each support vector x_i
β_o= Intercept of slope
$\sum_{i \in S}$ ≡Summation of all the observations in our sample space S
α_i= number of parameters/observations
$(k2)$= A kernel; function that quantifies the similarity between two observations
(x, x_i)= input (x) and each support vector x_i

K-nearest neighbours (KNN)
classifier works by estimating the conditional distribution of Y given X, and it classifies a given observation under consideration to the class with the maximum estimated probability. KNN then applies the Bayes rule and classifies the test observation x_o to the class with the largest probability. **Equation 2** depicts the KNN classifier, [21].

$$Pr\left(Y = \frac{j}{X} = x_o\right) = \frac{1}{K} \sum_{x_i \in N_o} I(y_i = j) \quad \textbf{(Eqn. 2)}$$

https://doi.org/10.54987/jobimb.v10i2.754

Where;
K = Points in the training data closest to x_o
x_o= Test observations
N_o= Training observations
y_i= Observations

Decision tree (DT)
There are two steps involved. First, possible classes for which the observations belong are generated. Then, every observation that falls into a particular class is labelled as either A, B or C depending on the number of classes generated. **Equation 3** depicts the DT classifier [21].

$$\sum_{j=1}^{j}\sum_{i=R_j}^{i}\left(y_i - y_{R_j}\right) \qquad \textbf{(Eqn. 3)}$$

Where,

J = Predictor space
R_j= J distinct and non-overlapping regions
y_i= Response variable for the i[th] observations
y_{R_j}= Mean response for the training observations within the j[th] region

RESULTS AND DISCUSSION

Data Visualisation and Statistics
Meanwhile, as mentioned before that our dataset consists of three classifications; benign (0), malign (1) and pre-malign (2), knowing the number of datasets for each class is significant. This statistic is significant as it helps to determine the next cause for action.

Fig. 2 shows the bar chart depicting the pictorial representation of our Classification attribute/feature, which is used as our target variable used in the model, the results shows that benign has 83 samples, malign has 145 and pre-malign has 188 samples respectively. This imply that Nigerians has a history of late presentation of breast cancer and needs a different approach for their classification.

```
Class 0: BENIGN:  85
Class 1: MELIGN:  145
Class 2: PRE-MELIGN:  188
```

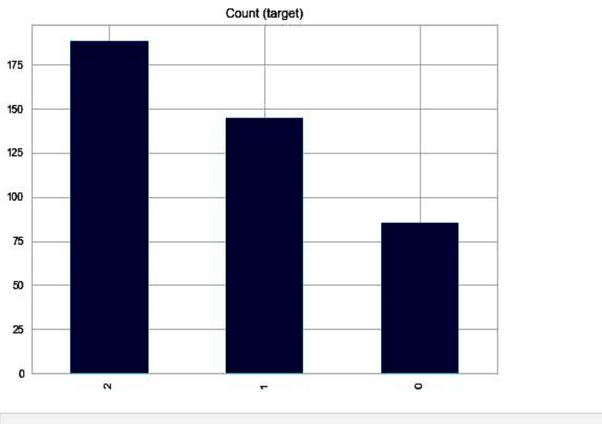


Fig. 2. Data Statistics of the dataset.

Fig. 3 shows the summary of our dataset containing all the 12 inputs variables and 1 target value and showing the total number of the samples used for the classification making 418 instances. No cells are having [NaN] indicating that there are no missing values in the variables.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Sex             418 non-null    object
 1   Age             418 non-null    int64
 2   OCCUPATION      418 non-null    object
 3   MARITAL STATUS  418 non-null    object
 4   LOCATION L/R    418 non-null    object
 5   SMTNES          418 non-null    int64
 6   HISTOLOGY       418 non-null    int64
 7   SMKIN HISTORY   418 non-null    object
 8   HERP 2          418 non-null    object
 9   P.R Y/N         418 non-null    object
 10  E.R Y/N         418 non-null    object
 11  STAGE           418 non-null    int64
 12  FMLY HISTRY     418 non-null    int64
 13  CLASS           418 non-null    object
dtypes: int64(5), object(9)
memory usage: 45.8+ KB
None
```

Fig. 3. The summary of the dataset.

Feature importance
Fig. 4 is a representation of our dataset based on the feature importance, this show that the parameter "Stage" will be given more consideration, this is because of the history of late presentation of breast cancer by the patient, followed by the progesterone, estrogen receptor and HERP2 respectively being the key featured that will be used for the classification. The parameters Age, Smoothness, Marital Status etc. will be used to support the prediction and suggest the best prognosis for the patient.

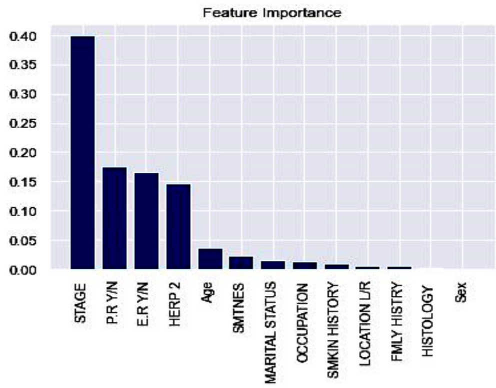


Fig. 4. Feature importance.

Support Vector Machine (SVM)
During experiments shown in **Fig. 5**, data used for training was 70%, while the dataset from testing was 30% which amounted to an accuracy of 0.99% with a precision of 1.00 for Benign, 0.98 for Malign and 1.00 for Pre-malign, recall of 0.97 for Benign, 1.00 for Malign and 1.00 for Pre-malign, f1 score of 0.98 for Benign, 0.99 for Malign and 1.00 for Pre-malign, support of 31 for Benign, 43 for Malign and 52 for Pre-malign.

```
print(classification_report(y_test, y_pred_1))

              precision    recall  f1-score   support

           0       1.00      0.97      0.98        31
           1       0.98      1.00      0.99        43
           2       1.00      1.00      1.00        52

    accuracy                           0.99       126
   macro avg       0.99      0.99      0.99       126
weighted avg       0.99      0.99      0.99       126
```

Fig. 5. Support Vector Machine (SVM) Results.

https://doi.org/10.54987/jobimb.v10i2.754

As shown in **Fig. 6**, SVM produced a good multi-class confusion matrix which shows the performance of the 3 classifiers used i.e Benign, Malignant and Pre-malign. Unlike binary classification, there are no positive or negative classes therefore, in this multi-class confusion matrix the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) of each individual class will be obtained. However, in this model, 30 sample were predicted correctly of be Benign (0), 43 samples were predicted correctly to be Malignant (1) only 1 sample was predicted to be Benign (0) instead Malignant (1) and 52 samples was correctly predicted to be Pre-malign (2).

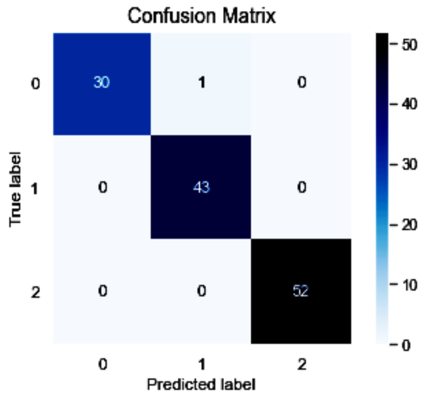


Fig. 6. SVM Confusion Matrix.

K nearest Neighbors (KNN)

During experiments shown in **Fig. 7**, data used for training was 70%, while the dataset from testing was 30% which amounted to an accuracy of 96.8% with a precision of 1.00 for Benign, 0.95 for Malign and 0.96 for Pre-malign, recall of 0.97 for Benign, 0.95 for Malign and 0.98 for Pre-malign, f1 score of 0.98 for Benign, 0.95 for Malign and 0.97 for Pre-malign, support of 31 for Benign, 43 for Malign and 52 for Pre-malign.

KNeighbors accuracy score : 96.82539682539682

	precision	recall	f1-score	support
0	1.00	0.97	0.98	31
1	0.95	0.95	0.95	43
2	0.96	0.98	0.97	52
accuracy			0.97	126
macro avg	0.97	0.97	0.97	126
weighted avg	0.97	0.97	0.97	126

\
Fig. 7. KNN Results.

As shown in **Fig. 8**, KNN produced a good multi-class confusion matrix which shows the performance of the 3 classifiers used i.e Benign, Malignant and Pre-malign. Unlike binary classification, there are no positive or negative classes therefore, in this multi-class confusion matrix the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) of each individual class will be obtained. However, in this model, 30 sample were predicted correctly of be Benign (0), 41 samples were predicted correctly to be Malignant (1) only 1 sample was predicted to be Benign (0) instead Malignant (1), and 1 sample was predicted to be Pre-malign (2) instead of Malignant (1) while 51 samples was correctly predicted to be Pre-malign (2) and 2 samples was predicted to be Malignant (1) instead of Pre-malign (2).

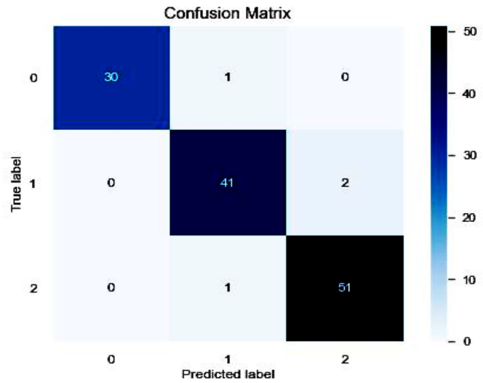


Fig. 8. KNN Confusion Matrix.

Decision Tree (DT)

During experiments shown in **Fig. 9**, data used for training was 70%, while the dataset from testing was 30% which amounted to an accuracy of 95.2% with a precision of 1.00 for Benign, 0.95 for Malign and 0.93 for Pre-malign, recall of 0.97 for Benign, 0.95 for Malign and 0.98 for Pre-malign, f1 score of 0.98 for Benign, 0.93 for Malign and 0.95 for Pre-malign, support of 31 for Benign, 43 for Malign and 52 for Pre-malign.

Decision Tree accuracy score : 95.23809523809523

	precision	recall	f1-score	support
0	1.00	0.97	0.98	31
1	0.95	0.91	0.93	43
2	0.93	0.98	0.95	52
accuracy			0.95	126
macro avg	0.96	0.95	0.96	126
weighted avg	0.95	0.95	0.95	126

Fig. 9. DT Results.

As shown in **Fig. 10**, DT produced a good multi-class confusion matrix which shows the performance of the 3 classifiers used i.e Benign, Malignant and Pre-malign. Unlike binary classification, there are no positive or negative classes therefore, in this multi-class confusion matrix the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) of each individual class will be obtained. However, in this model, 30 sample were predicted correctly of be Benign (0), 39 samples were predicted correctly to be Malignant (1) only 1 sample was predicted to be Benign (0) instead Malignant (1), and 1 sample was predicted to be Pre-malign (2) instead of Malignant (1) while 51 samples was correctly predicted to be Pre-malign (2) and 4 samples was predicted to be Malignant (1) instead of Pre-malign (2).

https://doi.org/10.54987/jobimb.v10i2.754

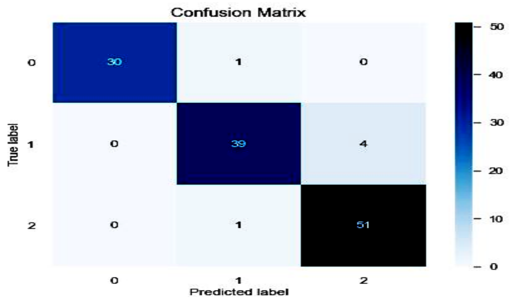


Fig. 10. DT Confusion Matrix.

Table 3 compare the result of all the models used in the Breast Cancer classification in three classes Benign, Malign and Pre-malign alongside their accuracy, precision, recall, F1 score and Support. The model SVM emerged the best model used in the classification with the highest accuracy of 99.2%.

Table 3. Comparison of breast cancer classification models in three classes benign, malign and pre-malign.

Model	Class of Breast Cancer (BC)	Accuracy	Precision	Recall	F1 Score	Support
Support Vector Machine (SVM)	Benign	99.2%	1.00	0.97	0.98	31
	Malign		0.98	1.00	0.99	43
	Pre-Malign		1.00	1.00	1.00	52
K Nearest Neighbors (KNN)	Benign	96.8%	1.00	0.97	0.98	31
	Malign		0.95	0.95	0.95	43
	Pre-Malign		0.96	0.98	0.97	52
Decision Tree (DT)	Benign	95.2%	1.00	0.97	0.98	31
	Malign		0.95	0.91	0.93	43
	Pre-Malign		0.93	0.98	0.95	52

CONCLUSION

SVM, DT and KNN was employed on our local datasets since our goal and challenge in breast cancer classification is to construct precise and trustworthy classifiers. After developing the model, the SVM has a greater efficiency of 99 percent. The research classifies breast cancer into three classes i.e Benign, Pre-malign and Malignant for better prognosis and medication in our locality.

ETHICAL CLEARANCE

Ethical approval was obtained from Ahmad Bello University Zaria with reference number ABUTHZ/HREC/W35/2021

CONFLICT OF INTERESTS

"The authors declare that there is no conflict of interests regarding the publication of this article."

REFERENCES

1. Łukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanisławek A. Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review. Cancers. 2021 Aug 25;13(17):4287.
2. Akram M, Iqbal M, Daniyal M, Khan AU. Awareness and current knowledge of breast cancer. Biol Res. 2017 Dec;50(1):33.
3. Prastyo PH, Paramartha IGY, Pakpahan MSM, Ardiyanto I. Predicting Breast Cancer: A Comparative Analysis of Machine Learning Algorithms. Proceeding Int Conf Sci Eng. 2020 Apr 30;3:455–9.
4. Momenimovahed Z, Salehiniya H. Epidemiological characteristics of and risk factors for breast cancer in the world. Breast Cancer Targets Ther. 2019 Apr;Volume 11:151–64.
5. Pruitt LCC, Odedina S, Anetor I, Mumuni T, Oduntan H, Ademola A, et al. Breast Cancer Knowledge Assessment of Health Workers in Ibadan, Southwest Nigeria. JCO Glob Oncol. 2020 Nov;(6):387–94.
6. Singh LK, Khanna M, Singh R. Artificial intelligence based medical decision support system for early and accurate breast cancer prediction. Adv Eng Softw. 2023 Jan;175:103338.
7. Zubair M, Wang S, Ali N. Advanced Approaches to Breast Cancer Classification and Diagnosis. Front Pharmacol. 2021 Feb 26;11:632079.
8. Ak MF. A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications. Healthcare. 2020 Apr 26;8(2):111.
9. Dubey AK, Gupta U, Jain S. Analysis of k-means clustering approach on the breast cancer Wisconsin dataset. Int J Comput Assist Radiol Surg. 2016 Nov;11(11):2033–47.
10. Sharma P, Klemp JR, Kimler BF, Mahnken JD, Geier LJ, Khan QJ, et al. Germline BRCA mutation evaluation in a prospective triple-negative breast cancer registry: implications for hereditary breast and/or ovarian cancer syndrome testing. Breast Cancer Res Treat. 2014 Jun;145(3):707–14.
11. Venkatesan E. Performance Analysis of Decision Tree Algorithms for Breast Cancer Classification. Indian J Sci Technol. 2015 Jan 20;8(1):1–8.
12. Ntekim A, Oluwasanu M, Odukoya O. Breast Cancer in Adolescents and Young Adults Less Than 40 Years of Age in Nigeria: A Retrospective Analysis. Singh A, editor. Int J Breast Cancer. 2022 Jul 29;2022:1–8.
13. Johnson O. Awareness and practice of breast self examination among women in different African countries: A 10-year review of literature. Niger Med J. 2019;60(5):219.
14. Olayide AS, Halimat AJ, Samuel OA, Ganiyu RA, Soliu OA. Level of awareness and knowledge of breast cancer in Nigeria. A systematic review. Ethiop J Health Sci. 2017 Mar 15;27(2):163.
15. Ojewusi AA, Obembe T, Arulogun OS, Olugbayela T. Breast cancer awareness, attitude and screening practices in Nigeria: A systematic review. Clin Rev Opin. 2016 Jul 31;7(2):11–25.
16. Oladimeji O, Bidemi YO, Olufisayo JAY, Sola AO. Prostate Cancer Awareness, Knowledge, and Screening Practices among Older Men in Oyo State, Nigeria. Int Q Community Health Educ. 2010 Oct;30(3):271–86.
17. Adeniji AA, Dawodu OO, Habeebu MY, Oyekan AO, Bashir MA, Martin MG, et al. Distribution of Breast Cancer Subtypes Among Nigerian Women and Correlation to the Risk Factors and Clinicopathological Characteristics. World J Oncol. 2020;11(4):165–72.
18. Irabor OC, Nwankwo KC, Adewuyi SA. The Stagnation and Decay of Radiation Oncology Resources: Lessons From Nigeria. Int J Radiat Oncol. 2016 Aug;95(5):1327–33.
19. Akinyemiju T, Jones K, Gupta A, Oyekunle T, Saraiya V, Deveaux A, et al. Association of body composition with odds of breast cancer by molecular subtype: analysis of the Mechanisms for Established and Novel Risk Factors for Breast Cancer in Nigerian Women (MEND) study. BMC Cancer. 2021 Dec;21(1):1051.
20. Adebayo IA, Arsad H, Samian MR. antiproliferative effect on breast cancer (MCF7) OF Moringa oleifera seed extracts. Afr J Tradit Complement Altern Med. 2017 Feb 13;14(2):282–7.
21. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. New York: springer; 2013 Jun.