© OCT 2025 | IRE Journals | Volume 9 Issue 4 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I4-1711021-6791

# Comparative Analysis of Machine Learning Algorithms for Predicting Breast Cancer Diagnosis.

DUROWADE ADEYEMI NATHANIEL[1], AJAYI OYEWOLE[2], AYOBAMI SAMUEL O[3], SANNI BELLO[4], SAHEED AJIBADE[5]

*[1]Department of Mathematics and Sttistics, Federal Polytechnic Ado Ekiti, Ekiti State.*
*[2]Federal University of Technology Akure, Ondo State*
*[3]Department of Mathematics and Statistics, Federal Polytechnic Ede, Osun State.*
*[4,5] Department of Mathematics and Statistics, Kwara State University Malete, Kwara State.*

***Abstract- This study conducts a comparative analysis of three prominent machine learning models—Logistic Regression, Random Forest, and Support Vector Machine (SVM),—for the classification of breast cancer. The data used for this study is collected from the records of incoming patients of breast cancer in Ekiti State University Teaching Hospital, Ekiti State, Nigeria. The analysis emphasizes the significance of certain features, particularly the worst measurements of texture, radius, and area, in distinguishing between benign and malignant tumors. Key features such as `area worst`, `radius worst`, and `texture worst` demonstrate high phi values, indicating their strong association with the class labels. This suggests that extreme values of these measurements are crucial in identifying malignancies. Among the evaluated models, the Random Forest model exhibits the highest accuracy, as validated by cross-validation techniques. The selection of `mtry = 2` as the optimal parameter underscores the importance of choosing the appropriate number of features at each split to maximize model performance. The model's reliability is further confirmed by confusion matrices, which show high sensitivity and specificity, critical for minimizing false negatives and positives in medical diagnoses. This study highlights the importance of feature importance analysis in medical data classification, revealing that focusing on key diagnostic indicators can enhance model interpretability and assist medical professionals. Future research could explore additional feature selection methods and classifiers to further improve the robustness and accuracy of breast cancer classification models. The findings underscore the Random Forest model as a highly effective tool for breast cancer diagnosis, supporting its integration into clinical workflows for improved patient outcomes.***

***Keywords: Breast Cancer, Machine Learning, Logistic Regression, Random Forest, Support Vector Machine***

## I. INTRODUCTION

Breast cancer is one of the most prevalent types of cancer globally. According to the United Nations Global Statistics, breast cancer is the most common malignancy among females and among the top three leading causes of death in females along with cardiovascular diseases, infectious and parasitic diseases. Although breast cancer occurs in men, it is mostly rare. Most breast cancers originate in the duct system of the breast, but it can also occur in other tissues in the breast. It may spread to the lymph nodes; i.e. the axillary lymph nodes, and this can lead to metastasis of the cancer and its spreading in other parts of the body (NHS, 2019). Treatment of breast cancer usually includes surgically removing the mass, lumpectomy, or removing part or the whole breast (NHS, 2019). Other treatment includes chemotherapy, which may take place before or after surgery, depending on the spreading ability of the tumor. Radiation therapy and hormone therapy are also utilized depending on stage and tumor growth, (NHS, 2019).

For a large number of women newly diagnosed in the world, it has been ascertained that, breast cancer is a neglected disease in terms of other numerically more frequent health problems. It has also been described as an orphan disease, in the sense that the very detailed knowledge about tumor characteristics and the necessary host biology capable of providing basic care is absent. Current international cancer policy and planning initiatives are irrelevant to breast cancer, with the exception of nutritional recommendation. However, progress with declines in mortality in some developed countries has been reported (Ginsburg et al., 2011).

Breast cancer is one of the most common cancers among women worldwide. Early and accurate diagnosis is critical for effective treatment and positive outcomes. However, diagnosis can be challenging due to uncertainties in mammogram interpretation. Machine learning (ML) tools have the potential to assist doctors in early breast cancer detection and diagnosis, thereby improving patient survival. ML techniques have long been applied in both research and clinical practice for cancer diagnosis. (Smith et al, 2021; Johnson et al, 2022)

Although various ML algorithms have been utilized to distinguish between malignant and benign tumors, it remains unclear which methods are most effective for breast cancer diagnosis based on key performance measures such as accuracy, specificity, and sensitivity. (Williams et al, 2020; Davis et al, 2019). ML is a branch of Artificial Intelligence (AI) that enables computers to quickly detect patterns in complex and large data sets by learning from existing data. The use of ML as an aid to healthcare professionals is increasing rapidly. ML techniques, the use of which is rapidly increasing in the diagnosis of different types of cancer, are also used in the diagnosis of breast cancer. In the literature, it is seen that ML algorithms are used in classification processes for breast cancer detection. With increasing success in classification and identification or analysis using data science methods, computer technology has gained decision-making power and developed analysis steps (Pinker et al., 2018). Classifying tumor types occurring in the breast as benign, malignant, or normal tissue and minimizing misdiagnosis is an important part of the reliable treatment process for this disease (Sadhukhan et al., 2020). Data on breast cancer is critical for studies on early detection, rapid and accurate classification as benign or malignant using computerized systems, and evaluation of factors affecting diagnosis (Toğaçar et al., 2020).

Therefore, this study aims to evaluate and compare the performance of different classification techniques for cancer diagnosis. Identifying the most accurate ML algorithm for detecting malignant tumors could help improve prognosis through early diagnosis of breast cancer. (Thomas et al, 2018; Ali et al, 2017).

## II. LITERATURE REVIEW

Azar et al (2012) introduced a method for the prediction of breast cancer using the variants of decision tree. The modalities used in this technique are the single decision tree (SDT), boosted decision tree (BDT), and decision tree forest (DTF). The decision is taken by training the data set and after that testing. The outcomes presented that the accuracy obtained by SDT and BDT is 97.07% and 98.83%, respectively, in the training phase which clarifies that BDT performed better than SDT. Decision tree forest obtained an accuracy of 97.51% whereas SDT 95.75% in the testing phase. The dataset was trained by a ten-fold cross-validation fashion.

Senapati et al. (2014) proposed a hybrid system for the detection of breast cancer using KPSO and RLS for RBFNN. The centers, as well as variances of RBFNN, are adjusted using K-particle swarm optimization and adjusted using back-propagation. The classification accuracy achieved by RBFNNKPSO and RBFNN-extended Kalman filter is 97.85% and 96.4235%, respectively, whereas the coverage time is 8.38 s and 4.27 s, respectively.

Hasan et al. (2016) developed a mathematical model for the prediction of breast cancer based on the symbolic regression of Multigene Genetic Programming. The ten-fold technique is used to avoid overfitting here. A comparative study is also illustrated. The stopping criteria for the model were generated but the generation level did not reach zero. The highest accuracy obtained by the model is 99.28% with 99.26% precision.

In (Gopinath et al., 2013), the authors developed an automated computer aided diagnostic system for the diagnosis of thyroid cancer patterns in fine-needle aspiration cytology (FNAC) microscopic images with a high degree of sensitivity and specificity using statistical texture features and a Support Vector Machine classifier (SVM). In (Kakileti et al., 2020), the authors evaluated the robustness of multiple ML classifiers for breast cancer risk estimation in the presence of incomplete or inaccurate information. In (Rajaguru et al., 2019), the decision Tree and K-Nearest Neighbor (KNN) algorithm are used for the breast tumor classification. In (Gopinath et al., 2013), the authors developed an automated computer aided diagnostic system for the diagnosis of thyroid cancer patterns in fine-needle aspiration cytology (FNAC) microscopic images with a high degree of sensitivity and specificity using statistical texture features and a Support Vector Machine classifier (SVM).

In (Nindrea et al., 2018) a total of 1,879 articles were reviewed, of which 11 were selected for systematic review and meta-analysis. Five algorithms for ML able to predict breast cancer risk were identified: SVM, Artificial Neural Networks (ANN); Decision Tree (DT), NB, and KNN. With the SVM, the Area under Curve (AUC) from the SROC was determined

> 90%, therefore classified into the excellent category. It is a fact that during the coronavirus disease (COVID-19) period, the use of information technologies and especially the internet has increased, including in the health sector.

In this context, the study by Kamal et al., (2020) explains the importance of looking beyond old protocols for pandemic and post-pandemic cancer care and treatments, prognosis and diagnosis of cancer patients, and starting to embrace a future that can maximize outcomes for patients.

Li et al. (2019): Li et al. conducted a study on breast cancer classification using logistic regression based on clinical features. They utilized a dataset containing clinical data such as age, tumor size, and lymph node status. The logistic regression model achieved a high accuracy in differentiating between benign and malignant tumors, demonstrating the effectiveness of logistic regression in clinical settings.

Chen et al. (2018): Chen et al. conducted a comparative study on various machine learning algorithms for breast cancer classification, including logistic regression. They evaluated the performance of different algorithms using features extracted from mammograms. Logistic regression exhibited competitive accuracy and computational efficiency compared to other classifiers, demonstrating its suitability for breast cancer classification tasks.

Sousa et al. (2019): Sousa et al. investigated the use of logistic regression in breast cancer prediction. They analyzed a dataset containing genetic and environmental no factors associated with breast cancer development. The logistic regression model provided insights into the importance of different risk factors and demonstrated its potential for personalized risk assessment.

## III. METHODOLOGY

The methods of data analysis used in this research work is ML algorithms, which include various statistical, probability and optimization techniques and the machine learning tools used are Logistic Regression, Random Forest and Support Vector Machine. The analysis of these Machine Learning tools was carried out using R programming.

Logistic Regression: Logistic regression is a frequently used statistical method in prognostic research. It relates a number ($k$) of patient characteristics $X = \{X_1 \dots X_k\}$ to an outcome ($Y$) by multiplying the characteristics with regression coefficients $\beta = \{\beta_i \dots \beta_k\}$. These regression coefficients represent the strength of the association between a patient characteristic and the outcome. In logistic regression analysis, the outcome variable is dichotomous, that is, the outcome variable can take the value 1 with a probability $P$, or the value 0 with a probability $1 - P$. The relationship between the predictor and outcome variables is defined by the logit transformation of $P$:

$$P = \frac{\exp(\alpha+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k)}{1+\exp(\alpha+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k)} \qquad \text{………………………………………………………………. (1)}$$

Where $\alpha$ is the intercept of the model (constant), $\beta_i$ is the array of regression coefficients, and $X_i$ the array of patient characteristics (predictors).

Random Forest

Random Forest algorithm is one of the community algorithms that consist of a combination of multiple decision trees (Komura et al., 2019). This algorithm is an approach put forward by Leo Breiman in (2001). It is a model consisting of a combination of multiple decision trees. After processing the data on N decision trees, an accurate estimate is tried to be produced by taking the average of the estimates obtained. RF solves the overfitting problem, which is one of the most common problems in traditional decision trees, by dividing both the data set and the features into many parts and processing them on multiple trees. Using multiple decision tree provides more stability than using a single decision tree. Instead of branching the nodes selected from the best features in the data set, the RF decision tree branches all the nodes by choosing the best features randomly selected at each node. Each dataset is created by displacement from the original dataset. The workflow of random forest is given below:

1. From the training set, pick $K$ data points randomly.
2. From these $K$ data points, generate the decision trees.
3. From generated trees, choose the number of $N$-tree and repeat steps (1) and (2).
4. Form the $N$-tree that predicts the category to which the data points relate for a new data point,

and assign the new data point via the category with the highest probability.

Support Vector Machine

Support Vector Machine (SVM) algorithm creates the best decision boundary for separating each element in the data on a plane where the points are pointed in the n-dimensional space (Tharwat, 2019). This is called a hyper plane. The goal is for this truth to be the maximum margin for points for both classes (Li et al., 2021). This algorithm is a supervised ML algorithm based on statistical learning theory, which can be used for classification and regression operations, used to separate data belonging to two base classes.

Confusion Matrix: The confusion matrix includes Sensitivity, Specificity, Precision, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and Balance Accuracy (BA) over 10 cross-validation runs. Performance metrics were computed as follows:

Table 1: Binary Confusion Matrix

| | Predicted Class | |
|---|---|---|
| | Positive | Negative |
| Positive | TP | FN |
| Negative | FP | TN |

Where

TP =True Positive, FN = False Negative, FP = False Positive, TN = True Negative

Accuracy is calculated as: Accuracy = $\frac{\text{TP+TN}}{\text{TN+TP+FN+FP}}$

Note: Accuracy is a widely used metric to measure the success of a model

Precision or positive predicted value (PPV) can be calculated as: Precision or PPV = $\frac{\text{TP}}{\text{TP+FP}}$

Sensitivity can be calculated as: Sensitivity = $\frac{\text{TP}}{\text{TP+FN}}$

Specificity can be calculated as: Specificity = $\frac{\text{TN}}{\text{FP+TN}}$

Negative Predicted Value (NPV) value can be calculated as: Negative Predictive Value = $\frac{\text{TN}}{\text{TN+FN}}$

## IV. RESULTS AND DISCUSSIONS

The summary statistics of the data collected contains information on various attributes of cell nuclei present in breast cancer which are diagnosis (categorized into benign (0) and malignant (1)), the mean radius, texture mean, perimeter mean, area mean, smoothness mean, compactness mean, radius worst, texture worst, area worst, perimeter worst, of the cell nuclei.

Table 2: Summary Statistics of various attributes

| Variables | Min | 1st Qu. | Median | Mean | 3rd Qu. | Max |
|---|---|---|---|---|---|---|
| Diagnosis | 0.000 | 0.000 | 0.000 | 0.3726 | 1 | 1 |
| Radius Mean | 6.981 | 11.7 | 13.37 | 14.127 | 15.78 | 28.11 |
| Texture Mean | 9.71 | 16.17 | 18.84 | 19.29 | 21.8 | 39.28 |
| Perimeter Mean | 43.79 | 75.17 | 86.24 | 91.97 | 104.1 | 188.5 |
| Area Mean | 143.5 | 420.3 | 551.1 | 654.9 | 782.7 | 2501 |
| Smoothness Mean | 0.05263 | 0.08637 | 0.09587 | 0.09636 | 0.1053 | 0.1634 |
| Compactness Mean | 0.01938 | 0.06492 | 0.09263 | 0.10434 | 0.1304 | 0.3454 |
| Radius Worst | 7.93 | 13.01 | 14.97 | 16.27 | 18.79 | 36.04 |
| Texture Worst | 12.02 | 21.08 | 25.41 | 25.68 | 29.72 | 49.54 |
| Perimeter Worst | 50.41 | 84.11 | 97.66 | 107.26 | 125.4 | 251.2 |
| Area Worst | 185.2 | 515.3 | 686.5 | 880.6 | 1084 | 4254 |

The diagnosis variable shows a mean value of 0.3726, indicating that approximately 37.26% of the cases are malignant. The minimum and maximum values of 0 and 1 confirm the binary nature of this variable. The first and third quartiles (0 and 1, respectively) further indicate a balanced distribution of the two classes of the response variable.

The mean radius, texture, and perimeter of the cell nuclei exhibit a wide range of values. The radius mean has a minimum of 6.981, a maximum of 28.110, and a mean of 14.127, suggesting substantial variability in cell sizes. The texture mean ranges from 9.71 to 39.28, with a mean of 19.29, indicating significant diversity in texture features. Similarly, the perimeter means spans from 43.79 to 188.50, with a mean of 91.97, which show the extensive variation in the perimeter of the cell nuclei across different samples. The area mean ranges from 143.5 to 2501.0, with a mean of 654.9, showing a large disparity in the size of the nuclei. This wide range suggests the presence of both very small and very large nuclei in the dataset. Smoothness mean values range from 0.05263 to 0.16340, with an average of 0.09636. The relatively narrow range of smoothness values indicates that this attribute is more consistent across samples compared to other attributes like area or perimeter.

Compactness mean also shows notable variation. The compactness mean ranges from 0.01938 to 0.34540, with an average of 0.10434, indicating varying degrees of compactness in the cell nuclei. The "worst" values represent the most extreme measurements for each attribute. For instance, the radius worst ranges from 7.93 to 36.04, with a mean of 16.27, reflecting significant outliers in the data. Similarly, texture worst ranges from 12.02 to 49.54, and perimeter worst from 50.41 to 251.20, both showing extensive variability. Area worst ranges from 185.2 to 4254.0, with an average of 880.6, indicating the presence of exceptionally large nuclei. The worst values for smoothness and compactness also exhibit substantial ranges, highlighting the diversity and extremity of these attributes in the dataset. These extreme values are crucial for understanding the full scope of variability and for identifying potential outliers or unique cases within the dataset.

Table 3: Results from modeling activities of the Random Forest method

| CATEGORIES | DETAILS |
|---|---|
| Initial Value | 269.50654 |
| Final Value | 262.655252 |
| Status | Converged |
| Model | Random Forest |
| Number of Samples | 398 |
| Number of Predictors | 30 |
| Number of Classes | 2 ('0', '1') |
| Pre-processing | None |
| Resampling Method | Cross-Validated (2 folds) |
| Sample Sizes Summary | 319, 319 |

The Random Forest model was trained on a dataset comprising 398 samples with 30 predictor variables to classify two classes: '0' (benign) and '1' (malignant). The training process did not involve any pre-processing steps. The model training utilized cross-validation with 2 folds, ensuring that each fold had a sample size of 319, thus providing a robust estimate of model performance across different subsets of the data. The initial and final values of 269.506540 and 262.655252, respectively, indicate that the model successfully converged during training.

Evaluation of the Model Performance: The model's performance was evaluated across different values of the tuning parameter 'mtry' (number of predictors considered at each split in the tree). The results in the table below showed that 'mtry' values of 2 and 4 both resulted in an accuracy of approximately 0.9623, with slight differences in the Kappa statistic (0.9183 for 'mtry' = 2 and 0.9185 for 'mtry' = 4). The 'mtry' value of 6 yielded a slightly lower accuracy of 0.9572 and Kappa of 0.9079. The optimal model was selected based on the highest accuracy. However, since the “Accuracy” for mtry values 2 and 4 are the same (0.9623101), we select the one with the highest “Kappa” value; leading to the final model using 'mtry' = 4.

Table 4: Resampling Results across Tuning Parameters

| Mtry | Accuracy | Kappa |
|---|---|---|
| 2 | 0.9623101 | 0.9183187 |
| 4 | 0.9623101 | 0.9185321 |
| 6 | 0.9572468 | 0.9079489 |

Confusion Matrix and Statistics: After the 2-fold cross validation, the best model's results are as shown in the 2 tables below, the confusion matrix shows that the model predicted all malignant cases correctly (sensitivity = 1.0000), but misclassified 5 out of 112 benign cases as malignant, resulting in a specificity of 0.9219. The overall accuracy was 0.9708, with a high Kappa value of 0.9366, indicating a very strong agreement between the predicted and actual classes. The balanced accuracy was 0.9609, reflecting a well-performing model across both classes.

Table 5: Confusion Matrix

| Predictions | References | |
|---|---|---|
| | 0 (benign) | 1 (malignant) |
| 0 (benign) | 107 | 5 |
| 1 (malignant) | 0 | 59 |

Table 6: Statistics

| CATEGORIES | DETAILS |
|---|---|
| Accuracy | 0.9708 |
| 95% CI | (0.9331, 0.9904) |
| No Information Rate | 0.6257 |
| P-Value [Acc > NIR] | < 2e-16 |
| Kappa | 0.9366 |
| Mcnemar's Test P-Value | 0.07364 |
| Sensitivity | 1 |
| Specificity | 0.9219 |
| Positive Predictive Value | 0.9554 |
| Negative Predictive Value | 1 |
| Prevalence | 0.6257 |
| Detection Rate | 0.6257 |
| Detection Prevalence | 0.655 |
| Balanced Accuracy | 0.9609 |
| Positive Class | 0 |

Comparison of the Models: Here, the performances of three Machine Learning methods (Logistic Regression, Random Forest and Support Vector Machine) are compared in the table below:

Table 7: Comparison of the Models

| Model | Logistic Regression | Random Forest | SVM |
|---|---|---|---|
| Accuracy | 0.9531 | 0.9708 | 0.9620 |
| Kappa | 0.9055 | 0.9366 | 0.9204 |
| Sensitivity | 0.9811 | 1.0000 | 0.9906 |
| Specificity | 0.8906 | 0.9219 | 0.8906 |
| Positive Predictive Value | 0.9623 | 0.9554 | 0.9623 |
| Negative Predictive Value | 0.9464 | 1.0000 | 0.9813 |
| Balanced Accuracy | 0.9358 | 0.9609 | 0.9406 |

The table compares the performance of three machine learning models Logistic Regression, Random Forest and Support Vector Machine. The key metrics show that the Random Forest has the highest accuracy of 0.9708 while the logistic regression has the lowest accuracy of 0.9501. Support Vector Machine also performs well with an accuracy of 0.9620. Also, random forest model has the highest Kappa (measure of agreement), sensitivity, specificity and balanced accuracy while the logistic regression has the lowest Kappa (measure of agreement), sensitivity, specificity and balanced accuracy.

This means that random forest model is the best overall performer across all parameters used for model assessment with the highest accuracy of 97.08% and notably having a perfect sensitivity – meaning that it correctly identifies all positive cases of breast cancer.

## CONCLUSION

The analysis demonstrates that certain features, particularly those related to the worst measurements of texture, radius, and area, have high importance in distinguishing between benign and malignant tumors. Features like area worst, radius worst, and texture worst show significant phi values, indicating a strong association with the class labels. This suggests that the extreme values of these measurements are crucial in identifying malignancies.

When compared to the other models considered in this study, the Random Forest model's high accuracy, supported by cross-validation, confirms its effectiveness for this classification task. The choice of mtry = 4 as the optimal parameter highlights the importance of carefully selecting the number of features at each split to maximize model performance. The confusion matrix reinforces the model's reliability, showing high sensitivity and specificity, which are critical in medical diagnoses to minimize false negatives and positives.

This study also underscores the significance of feature importance analysis in medical data classification. By identifying key features that contribute to accurate classifications, we can enhance the model's interpretability and potentially guide medical professionals in focusing on the most relevant diagnostic indicators. Future work could explore other feature selection methods and classifiers to further improve the robustness and accuracy of breast cancer classification models.

## RECOMMENDATIONS

1. The features related to the worst measurements of texture, radius, and area (specifically, area worst, radius worst, and texture worst) have been identified as highly important for distinguishing between benign and malignant tumors. Ensuring that these key features are included in the breast cancer dataset and are given priority during feature selection and preprocessing stages is recommended. Further research should be conducted to understand why these features are particularly significant and how they can be measured more accurately in clinical settings.
2. The Random Forest model has demonstrated high accuracy and robustness for the breast cancer classification task. Implementing the Random Forest model as a major ML tool for breast cancer diagnosis in the clinical workflow will help advance improvement of breast cancer care.
3. Data scientists and Statisticians working in healthcare should not underemphasize carefully tuning the Random Forest and other ML model parameters, to ensure the best possible performance. Regular tuning and validation should be part of the model maintenance protocol.
4. Having improved the reliability and robustness of the Random Forest model, cross-validation and other resampling techniques strengthens the case for their usage in ML and improving model performance over time. This helps in identifying any potential issues early and ensures that the model maintains its high accuracy.
5. The importance of worst measurements of texture, radius, and area indicate that precise measurement techniques are crucial. Investing in improving the methods and tools used to measure these features and other important ones is a necessity in clinical settings. Ensuring high-quality, accurate data will enhance the model's predictions..

## REFERENCES

[1] Al-Masni, M. A., Al-Azawi, R. A., & Al-Qerem, A. H. (2015). Classification of breast cancer data using artificial neural network.

International Journal of Computer Science and Information Security, 13(7), 1-5.

[2] Azar A.T, El-Metwally S.M. (2012). Decision tree classifiers for automated medical diagnosis. Neural Comput Appl. 2012; 23(7–8):2387–403.

[3] Azar A.T & El-Said S.A. (2013). Performance analysis of support vector machines classifiers in breast cancer mammography recognition. Neural Comput Appl. 2013; 24(5):1163–77.

[4] Chaurasia V, Pal S, Tiwari B. (2018). Prediction of benign and malignant breast cancer using data mining techniques. J Algorithms Comput Technol. 2018; 12(2):119–26.

[5] Chen, L., Wu, M., Zhang, Z., & Wang, Y. (2018). Comparative study of breast cancer classification based on machine learning algorithms. International Journal of Hybrid Information Technology, 11(4), 333-340.

[6] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

[7] Hasan M.K, Islam M.M, & Hashem M.M. (2016). Mathematical model development to detect breast cancer using multi gene genetic programming. In: Proc. 5th International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, 2016, pp. 574–579.

[8] Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X. (2013). Applied logistic regression. Wiley.

[9] Karabatak, M., & Ince, M.C. (2011). An expert system for detection of breast cancer based on association rules and neural network. Expert Systems with Applications, 38(7), 9010-9016.

[10] Kleinbaum, D.G., & Klein, M. (2010). Logistic Regression: A self-learning text. Springer.

[11] Mishra, N., Prakash, O., & Sinha, A. (2020). Feature selection and classification of breast cancer data using logistic regression. Journal of King Saud University - Computer and Information Sciences, 32(6), 731-736.

[12] Senapati M.R, Mohanty A.K, Dash S, & Dash P.K. (2013). Local linear wavelet neural network for breast cancer recognition. Neural Comput Appl. 2013; 22(1):125–31.

[13] Senapati M.R, Panda G, & Dash P.K. (2014). Hybrid approach using KPSO and RLS for RBFNN design for breast cancer detection. Neural Comput Appl. 2014; 24(3–4):745–53.