Transactions of NAMP 21, (2025) 65-78

Transactions of
The Nigerian Association of
Mathematical Physics
Journal homepage: https://nampjournals.org.ng

PERFORMANCE EVALUATION OF DATA MINING CLASSIFIERS FOR PREDICTING RECURRENCE AND SURVIVABILITY OF BREAST CANCER PATIENTS

*[1]Nurudeen A. A., [2]Umar U. M., [3]Asare B.K. and [4]Abdulkarim B.

[1]Department of Mathematics and Statistics, Kaduna Polytechnic, Kaduna.
[2,3]Department of Statistics, Usmanu Danfodiyo University, Sokoto, Sokoto.
[4]Department of Computer Science, Usmanu Danfodiyo University, Sokoto, Sokoto.

ARTICLE INFO

Article history:
Received xxxxx
Revised xxxxx
Accepted xxxxx
Available online xxxxx

Keywords:
Breast cancer, Machine learning, Ensemble classifier, Recurrence Survivability, Performance

ABSTRACT

Worldwide, breast cancer is currently the most common cancer, accounting for 12.5% of all new annual cancer cases and it is one of the leading causes of cancer-related death in women second only to lung cancer. Incorporating machine learning (ML) classifiers into predicting the recurrence and survival of patients with breast cancer has emerged as a promising approach to enhance performance metrics. This study analyzed dataset on breast cancer obtained from clinical studies- Barau Dikko Teaching Hospital, Kaduna and when the performance of employed Conventional ML classifiers- Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Proposed ensemble learning classifier (ANN-KNN) was evaluated, it was observed that both Conventional ML and Proposed ensemble learning classifiers could predict recurrence of breast cancer and survivability of breast cancer patients. However, the performance of these conventional ML classifiers and the proposed ensemble learning classifier were compared. The results showed that the proposed ensemble learning classifier outperformed ML classifiers with 97% and 91.04% accuracy on recurrence and survival prediction of breast cancer respectively. It remains the best classifier in predicting breast cancer patients' recurrence and survivability, followed by the ANN classifier with accuracy of 90.5% and 81.93% respectively on recurrence and survivability prediction of breast cancer patients. The findings demonstrate that ensemble learning can enhance the performance of weak classifiers like SVM and KNN. Further extensive evaluation of other ML classifiers like decision tree and random forest can be performed using some combinations that can predict the recurrence and survivability of breast cancer patients with greater accuracy.

1. Introduction

Breast cancer remains a critical issue among women globally, and is currently the most common cancer, accounting for 12.5% of all new annual cancer cases worldwide. It is one of the leading causes of cancer-related death in women (15% of all deaths among women) second only to lung cancer and it remains the world's leading type of cancer [1] [2]. The first breast cancer case was recorded in Egypt in 3000 BC [3]. A breast tumour is an abnormal growth of tissues in the breast, and it may be felt as a nipple or discharge or a change of skin texture around the nipple. However, early diagnosis of cancer is essential.

*Corresponding author: NURUDEEN, A. A.
E-mail address: nurkarmeds@gmail.com
https://doi.org/10.60787/tnamp.v21.474
1115-1307 © 2025 TNAMP. All rights reserved

In 2020, there were 2.3 million women diagnosed with breast cancer and 685,000 deaths globally, and as of the end of 2020, there were 7.8 million women alive who were diagnosed with breast cancer in the past 5 years, making it the world's most prevalent cancer [4]. In Nigeria, breast cancer remains the most prevalent and highest mortality among other types of cancer with 28,380 (22.7%) new cases and 14,274 (18.1%) respectively [5].

Breast cancer is a malignant disease that initiates in the breast cells. Patients with a family history of breast or ovarian cancer have the possibility of developing breast cancer [6]. Some of the risk factors for breast cancer are gender (more in females), hereditary, genetic mutation, smoking, alcohol, consumption, obesity (As in a sedentary lifestyle), canned foods, chemicals, and carcinogens used as preservatives and in cosmetics [6]. The high burden of breast cancer was attributed to the low or lack of cancer awareness among the population as well as early delay in cancer screening and detection. Increasing cancer cases in developing countries are also linked to the aging population, and changes in lifestyle such as unhealthy dietary practices and lack of physical activities [6].

1.1 Machine Learning (ML) Classifiers Used in this study

ML is a branch of Artificial Intelligence, that relates the problem of learning from data samples to the general concept of inference. Every learning process consists of two phases: (i) estimation of unknown dependencies in a system from a given dataset and (ii) use of estimated dependencies to predict new outputs of the system.

ML has also been proven an interesting area in biomedical research with many applications, where an acceptable generalization is obtained by searching through an n-dimensional space for a given set of biological samples, using different techniques and algorithms [7]. The incorporation of ML classifiers into the prediction of recurrence and survival of patients with breast cancer has emerged as a promising approach to enhance performance metrics. However, applying specific ML techniques could enhance the accuracy of cancer vulnerability, recurrence, and prognostication of survival, which could enhance detection before symptoms become severe [8].

2.0 RELATED STUDIES

Numerous studies have been conducted to predict the recurrence and survival of patients with breast cancer. However, the majority of these studies were carried out using statistical methods such as parametric, semi-parametric models or ML classifiers but very few of them used ensemble learning classifiers.

Many authors like [9]; [10]; [11]; [12]; [13]; [14]; [15]; [16]; [17]; [18]; [2]; [19]; [20]; [21]; [22]; [23]; and [24] had captured the performance of some machine learning classifiers like Artificial Neural Network, Decision tree, K-Nearest Neighborhood, Bayesian Networks, Support Vector Machine, Ensemble learning classifiers and so on in the prediction of recurrence or survival of breast cancer patients.

The ensemble learning classifier offers several advantages over conventional ML classifiers such as improved accuracy and performance, especially for complex and noisy problems. However, it is evident that the ensemble learning classifier outperforms single models and also turns the performance of multiple weak models into one strong model [25]; [26]; [27]; [22]; [23]; and [24]. A lot of studies have been reported about many deaths associated with breast cancer worldwide. According to [28], over 600,000 deaths were reported worldwide due to breast cancer. A major reason is that patients are not aware when to consult doctors. However, this study would be of help in assisting patients in knowing about the recurrence and survivability pattern. In this study, we developed an ensemble learning classifier and evaluated and compared the performance metrics of conventional ML classifiers with our developed ensemble classifier in predicting the recurrence and survival of patients with breast cancer.

3.0 METHODOLOGY

In the preparation of this manuscript, the researchers had carefully undertaken the following steps: Data collection, Data preprocessing, Feature selection, Data splitting (Training and Testing), Data mining classifiers, Data analysis

3.1 Method of Data Collection

The data used for this study was extracted from the records of the hospital's cancer registry department. The breast cancer data include variables like identification number, Age, Marital status, Menopausal status, Family history, Classification of breast cancer, Laterality, breast cancer stage classification, Estrogen receptor status, Progesterone receptor status, c-er-b2 status, Primary treatment type, Surgery type, Status, Tumour size (cm), Total axillary nodes removed, Number of positive lymph nodes and date of diagnosis, (date of clinical diagnosis).

3.2 Data preprocessing: The datasets in today’s real world are highly susceptible to noise, missing values, and inconsistency due to their typically huge size, as a result of this the dataset used for this study underwent thorough preprocessing to improve its quality and consequently improve the mining results.

3.3 Data Cleaning and Balancing: This involves routine work to “clean” the data by filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies. This step is very important because dirty data can cause confusion in the mining procedure and hence result in unreliable output.

3.4 Feature Selection: It involves reducing the number of attributes to improve the accuracy of the outcome. Here, the irrelevant and redundant features were removed. The random forest classifier was used to select important features: family history, age at diagnosis, method of diagnosis, time, and laterality were selected for the recurrence case, also twenty-four important features were selected for the survivability case. The importance of feature selection is to improve the performance of the classification techniques (Figures 2 and 3).

3.5 Data Splitting: Upon the completion of the dataset preprocess, the dataset was divided into training and testing. However, 80% of the dataset was for training, whilst the remaining proportion was for testing.

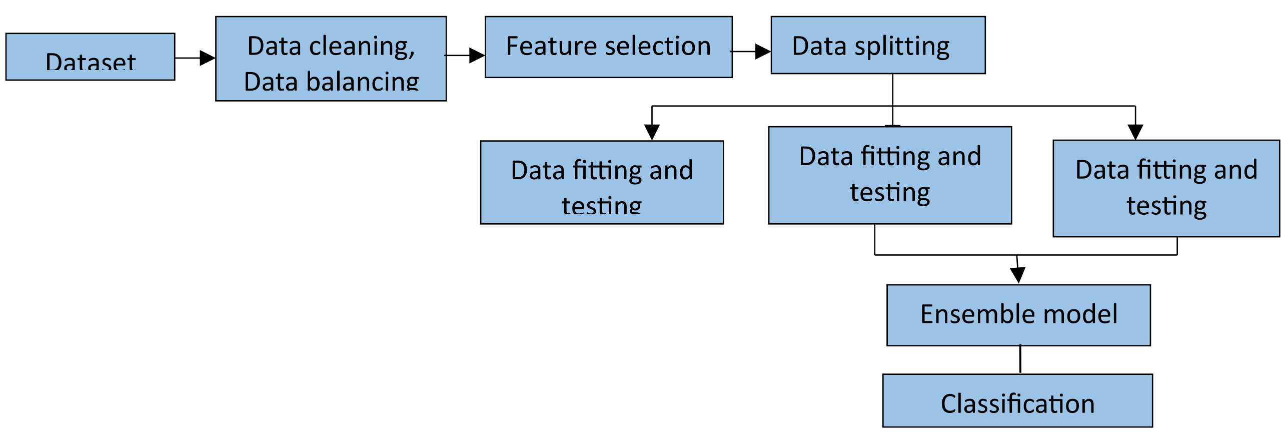

Figure 1: System Block Diagram

3.6 Data Mining Techniques Used in this Study

In this study, Artificial Neural Network (ANN), K-nearest Neighbors (KNN), Support Vector Machine (SVM) and the proposed ANN- SVM were employed as conventional machine learning classifiers and ensemble learning classifier (proposed classifier) respectively to predict the recurrence and survivability of women with breast cancer. However, the selection of these data mining classifiers met two criteria. Those that have shown the best performance in the related studies reviewed and the most frequently used in clinical datasets for classification problems. Let us provide a brief mathematical representation of each technique:

3.6.1 Artificial Neural Network (ANN) Classifier

Artificial Neural Networks are computational models inspired by the structure and function of biological neural networks. They consist of interconnected nodes (neurons) organized in layers (input layer, hidden layers and output layer). The output of a neuron is typically calculated using an activation function, such as the sigmoid function (often used in binary classification problems). The forward propagation in an ANN can be represented mathematically as follows:

$$a^{(l)} = g\left(z^{(l)}\right) \qquad (3.1)$$

$$z(l+1) = W^{(l)} a^{(l)} + b^{(l)} \qquad (3.2)$$

Where $a^{(l)}$ is the activation of layer l, $z^{(l+1)}$ is the weighted sum of activations of a layer $l, W^{(l)}$ $b^{(l)}$ are the weights and biases of layer l, and g is the activation function.

3.6.2 K-Nearest Neighbors (KNN) Classifier

K-Nearest Neighbors is a non-parametric classification algorithm that classifies an input by a majority vote of its neighbors, with the input being assigned to the class most common among its k-nearest neighbors (where k is a hyperparameter). Mathematically, the classification of a new data point x can be represented as:

$$c(x) = \textit{majority}\ \text{vote}\left(C(x_1), C(x_2), \dots, C(x_k)\right) \qquad (3.3)$$

Where $C(x_i)$ is the class label of the i nearest neighbor x, and $C(x)$ is the predicted class label of x.

3.6.3 Support Vector Machine (SVM) Classifier

Support Vector Machine is a supervised learning algorithm that separates classes by finding the hyperplane that maximizes the margin between classes. Mathematically, SVM aims to solve the optimization problem:

$$\text{minimize} \frac{1}{2} \| w \|^2 \qquad (3.4)$$

subject to:

$$y_i \left(w \,.\, x_i + b\right) \geq 1 \text{ for all i} \qquad (3.5)$$

Where w is the weight vector, b is the bias term, x_i is the training sample, and y_i is its corresponding class label.

3.7.0 Ensemble Learning Classifier

The Ensemble Learning (EL) method creates multiple instances of conventional ML methods and combines them to evolve a single optimal solution to a problem. This approach is capable of producing better predictive model compared to the conventional approach. Ensemble classifier is better than a single classifier because it turns weak models to one strong model (together we are stronger). The top reasons to employ the EL method include situations where there are uncertainties in data representation, solution objectives, modelling techniques, or the existence of random initial seeds in a model. The instances of candidate methods are called base learners. Each base learner works independently as a conventional ML method, and the eventual results are combined to produce a single robust output. The combination could be done using any of the averaging (simple or weighted) methods and voting (majority or weighted) for regression and classification methods, respectively.

EL classifiers are also known as "committee of machines" or "committee of experts" with the latter following the assumption that each base learner is an "expert" and its output is an "expert opinion. The mathematical representation provides a basic understanding of how the ensemble learning technique works in predicting the recurrence and survivability of women with breast cancer in the study. For a hybrid classifier combining Artificial Neural Network (ANN) with K-nearest Neighbors (KNN), the mathematical representations vary depending on the specific architecture and methodology used. However, one can provide a general idea of how this hybrid can be represented:

3.7.1 Proposed ANN-KNN (Ensemble learning Classifier)

In this hybrid approach, the ANN is used to extract features from the input data, which are then fed into the KNN algorithm for classification. The ANN can be trained to learn representations of the input data and the output of one of its hidden layers (or the output layer) can be used as the feature vector. The KNN algorithm then classifies the data points based on the distances between their feature vectors. Mathematically, the hybrid model can be represented as follows:

$$\left.\begin{aligned} ANN &: h = f_{ANN}(x) \\ KNN &: \hat{y} = KNN(h) \end{aligned}\right\} \quad (3.6)$$

Where x represents the input data, h represents the feature vector extracted by the ANN, f_{ANN} represents the function learned by the ANN, and $\hat{y}$ represents the predicted class label.

3.8 Evaluation of Model

The performance of these conventional ML classifiers and ensemble classifier employed on the testing dataset in predicting the recurrence and survivability of breast cancer was evaluated using performance metrics such as precision, recall, F1 score, and accuracy.
Accuracy: The ratio of correctly predicted instances to the total cases.

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

Precision: The ratio of correctly predicted positive observations to the total predicted positives.

$$\text{Precision} = \frac{TP}{FP + TP}$$

Recall: The ratio of correctly predicted positive observations to all observations in the actual class.

$$\text{Recall} = \frac{TP}{FN + TP}$$

F1Score: The harmonic mean of precision and recall.

$$\text{F1Score} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

Confusion Matrix: A table used to describe the performance of a classification algorithm, displaying the true positives, true negatives, false positives and false negatives.

3.9 Method of Data Analysis

The study employed four data mining classifiers to analyze clinical data of women with breast cancer. All analyses were performed using Python 3.7. Below are the primary libraries employed in this study.
Jupyter Notebook: Interactive coding and documentation
Pandas: For data preprocessing
Sklearn: For implementing and evaluation of classifiers
Numpy: For numerical calculations
Matplotlib: For data visualization and presentation of results in graphical form.

RESULTS AND DISCUSSION

4.1 Evaluation of Model Performance

The performance of the conventional ML classifiers- ANN, KNN, SVM, and proposed ensemble learning classifier- ANN-KNN was evaluated using the testing dataset. These classifiers were evaluated based on standard classification metrics stated above, which provided a deeper understanding of their capability to predict the recurrence and survivability of breast cancer patients.

4.1.1 Accuracy of the Model

4.1.1.1 ANN-KNN (Proposed ensemble classifier): The proposed classifier had the highest accuracy of 97.10% and 91.04% respectively for recurrence and survivability of breast cancer. This indicates that the proposed classifier correctly predicted 97% of cases of the recurrence of breast cancer. Similarly, it correctly predicted 91% of cases of the survivability of women with breast cancer. The confusion matrix shown in figure1 revealed that, in the prediction of recurrence of breast cancer, there are 603 correct predictions and 18 false predictions. However, this classifier predicted 317 data as 0 and 286 data as 1, this is its correct prediction. The classifier also predicted 18 data as 0 and zero data as 1, this is its wrong prediction. In the case of survivability, there are 630 correct forecasts and 62 incorrect predictions. However, this classifier predicted 301 data as 0 and 329 data as 1, this is its correct prediction. The classifier also predicted 16 data as 0 and 46 data as 1, this is an absolutely wrong prediction.

4.1.1.2 SVM Classifier: This classifier achieved 82.29% accuracy for recurrence prediction and 63.29% accuracy for survivability prediction of breast cancer patients. This means that the classifier correctly predicted 82.3% and 63.3% cases of the recurrence and survivability of breast cancer patients respectively. Looking at confusion matrix shown in figure 3. Here, for the recurrence of breast cancer, there are 511 correct predictions and 100 erroneous predictions. The classifier predicted 229 data as 0 and 282 data as 1. So, this is its correct prediction. This same classifier also predicted 22 and 88 data points to be 0 and 1 respectively. So, this is an absolutely wrong prediction. In the case of survivability, there are 438 correct guesses and 254 incorrect predictions. However, this classifier predicted 209 data as 0 and 229 data as 1, so this represents its correct prediction. The classifier also predicted 116 data as 0 and 138 data as 1, this is a wrong prediction. Here, the number of wrong predictions by SVM is greater than the number of wrong predictions by the proposed classifier. For this reason, its accuracy is less than that of ANN and proposed ensemble classifiers for both recurrence and survivability cases.

4.1.1.3 KNN classifier: This classifier correctly predicted 90.49% of breast cancer recurrence. Hence its accuracy is nearly 90.5% of the cases, which is better than that of SVM but lower than the proposed classifier. Similarly, the KNN classifier has achieved 81.93% accuracy, which indicates that it has correctly predicted the survivability of breast cancer in almost 82% of the cases, which is better than SVM but lower than ANN and the proposed classifiers. In the case of recurrence, there are 562 correct and 59 incorrect predictions, as shown by the confusion matrix in the figure. This classifier predicted 291 data as 0 and 271 data as 1, this is its correct prediction. However, it also predicted 33 and 26 data as 0 and 1, respectively, which is its wrong prediction. In the case of survivability, there are 567 correct guesses and 125 false predictions. However, this classifier predicted 243 data as 0 and 324 data as 1, representing its correct prediction. However, the classifier also predicted 21 data as 0 and 104 data as 1, this is a wrong prediction. It can be seen that the number of wrong predictions in the KNN classifier is higher than that of the proposed ensemble and ANN classifiers but lower than the SVM classifier, and this is the reason why its accuracy is less than that of the proposed ensemble and ANN classifiers but greater than the SVM classifier.

4.1.1.4 ANN classifier: The classifier correctly predicted 94.84% of the recurrence of breast cancer. Hence, its accuracy is approximately 95%, which is less than the proposed ensemble classifier but better than the SVM and KNN classifiers. Similarly, the ANN classifier correctly predicted 90.46% of cases of survivability prediction of breast cancer patients, so its accuracy is closely 90.5%. The confusion matrix shown in figure 4, in the case of breast cancer recurrence prediction, the number of correct and false predictions in this case is 589 and 32, respectively. This classifier predicted 317 as data 0 and 272 as data 1. This is a correct prediction. However, the classifier also predicted 32 data as 0 and zero data as 1. So, this is a wrong prediction. However, in the case of survivability of breast cancer patients, there are 626 correct predictions and 66 erroneous predictions. The classifier predicted 316 data points as 0 and 310 data points as 1, and this is the correct prediction. However, it also predicted 35 and 31 data as 0 and 1, respectively. This is its wrong prediction. Here, it can be seen that the number of wrong predictions for both recurrence and survivability of breast cancer is more than that of the proposed ensemble classifier but less than that of SVM and KNN classifiers, and this is the reason why its accuracy is less than that of the proposed ensemble classifier but greater than SVM and KNN classifiers.

4.1.2 Precision and Recall of the Models

A deeper understanding of classifiers' performance in predicting the recurrence and survivability of breast cancer patients was provided by precision and recall metrics. The proposed classifier achieved a precision of 100% and a recall of 94.63%, highlighting the model's effectiveness in accurately predicting both “yes” and “no” recurrence prediction. However, the same model achieved a precision of 86.74% and a recall of 94.95%, expressing the model's effectiveness in accurately predicting both “alive” and “dead” survivability prediction. The ANN classifier displayed a precision of 100% and a recall of 90.83%, the SVM classifier achieved a precision of 72.24% and a recall of 91.24%, and the KNN classifier demonstrated a precision of 91.80% and a recall of 89.81% for recurrence prediction. Similarly, the ANN classifier displayed a precision of 91.07 and a recall of 90.03%, the SVM classifier achieved a precision of 60.23% and a recall of 64.31%, and the KNN classifier demonstrated a precision of 70.03% and a recall of 90.05% for survivability prediction.

4.1.3 F1Score of the Model

This metric balances precision and recall. The proposed classifier had the highest F1score of 97.24% and 90.66% for recurrence and survivability respectively. However, this confirmed the superiority of the proposed ensemble classifier in the recurrence and survivability prediction of breast cancer patients over the conventional ML classifiers.

4.1.4 Confusion Matrix

The matrices and Table 1displayed the number of correct predictions and the number of incorrect predictions by the classifiers used. The proposed ensemble classifier had the lowest number of incorrect predictions when compared with the conventional ML classifiers, which validates its superiority over conventional ML classifiers which displayed a higher number of incorrect predictions.

Table 1: METRICS OF THE CONFUSION MATRIX OF THE DATASET

Classifier	Recurrence	Correct predictions	Incorrect predictions
SVM	0: Yes 1: No	229 282	22 88
ANN	0: Yes 1: No	317 272	32 0
KNN	0: Yes 1: No	291 271	33 26
Proposed ANN-KNN	0: Yes 1: No	317 286	18 0
Classifier	**Survivability**	**Correct guesses**	**False predictions**
SVM	0: Alive 1: Dead	209 229	116 138
ANN	0: Alive 1: Dead	316 310	35 31
KNN	0: Alive 1: Dead	243 324	21 104
Proposed ANN-KNN	0: Alive 1: Dead	301 329	16 46

Table 2: **PERFORMANCE METRICS OF CLASSIFIERS ON TEST DATA**

	Classifier	Precision (%)	Recall (%)	F1score (%)	Accuracy (%)
RECURRENCE	SVM	72.24	91.24	80.63	82.29
	ANN	100	90.83	95.19	94.84
	KNN	91.80	89.81	90.80	90.49
	Proposed ANN-KNN	**100**	**94.63**	**97.24**	**97.10**
SURVIVABILITY	**Classifier**	**Precision (%)**	**Recall (%)**	**F1score (%)**	**Accuracy (%)**
	SVM	60.23	64.31	62.39	63.29
	ANN	91.07	90.03	90.54	90.46
	KNN	70.03	90.05	79.54	81.93
	Proposed ANN-KNN	**86.74**	**94.95**	**90.66**	**91.04**

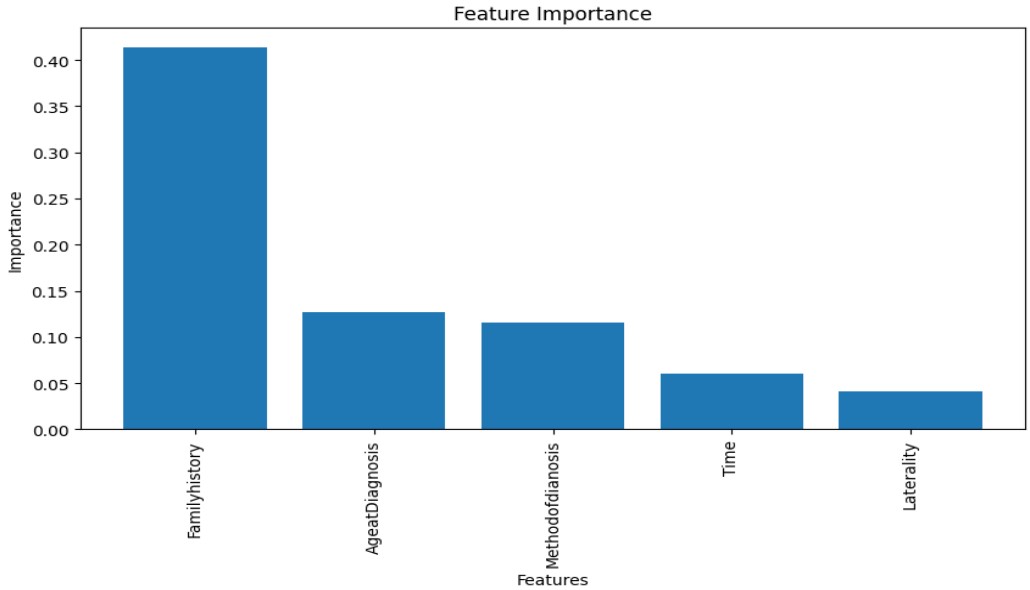

Figure 2: Feature importance in predicting recurrence of breast cancer.

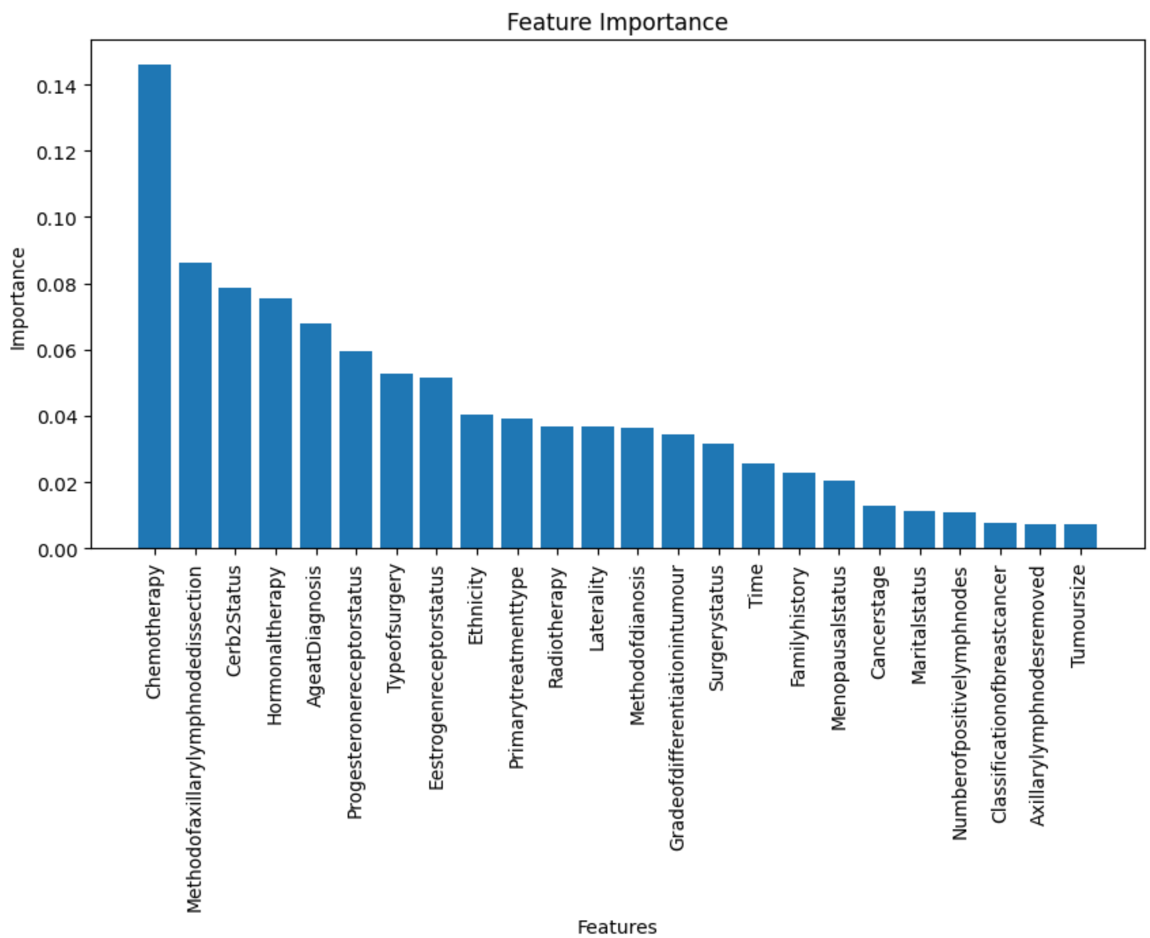

Figure 3: Feature importance in predicting Survivability of breast cancer patients.

Table 3: Model Comparison

Recurrence				
Current paper (Classifier name)	Accuracy (%)	Reference paper	(Classifier name)	Accuracy (%)
SVM	82.29	Ref. [9]	SVM	78.7
ANN	94.84	Ref. [16]	ANN	94.70
KNN	90.49	Ref. [29]	KNN	88.88
ANN-KNN (PROPOSED)	95.65	Ref. [24]	Ensemble	81.75
Survivability				
Current paper (Classifier name)	Accuracy (%)	Reference paper	(Classifier name)	Accuracy (%)
SVM	63.29	Ref. [14]	SVM	85.0
ANN	90.46	Ref. [31]	ANN	91.6
KNN	81.93	Ref. [30]	KNN	83.9
ANN-KNN (PROPOSED)	91.47	Ref. [3]	Ensemble	97.42

4.1.5 Model Comparison.

The classifiers used for this study are compared to those in existing studies on the prediction of the recurrence of breast cancer or predicting the survivability of breast cancer patients. However, it can be seen that all four classifiers we employed have a good accuracy level. The proposed ensemble classifier demonstrated outstanding performance over conventional ML classifiers and is consistent with findings in other studies where the ensemble learning classifiers have been shown to improve predictive accuracy. The findings in this current study align with existing pieces of literature on the prediction of the recurrence of breast cancer or predicting the survivability of breast cancer patients (Table 3).

CONCLUSION

This study has used only four data mining classifiers comprised of three conventional ML classifiers: ANN, SVM, KNN, and a proposed ensemble classifier (ANN-KNN) to predict breast cancer recurrence and survivability of breast cancer patients. The results obtained indicated that these classifiers could predict the recurrence of breast cancer and the survivability of breast cancer patients with varying degrees of accuracy. This study has demonstrated that the performance of the ensemble classifier (ANN-KNN) was the best among the classifiers in terms of accuracy, precision, recall, and F1 score in predicting recurrence and survivability of breast cancer patients, followed by the ANN classifier. It was observed that ensemble classifier can enhance the performance of weak classifiers like SVM and KNN. The key contribution of this study is that a precise literature review of the related works was carried out. Development of the hybrid ML (ensemble learning classifier) approach, employing feature selection, voting with the proposed ensemble classifier, and classification techniques for predicting the recurrence and survivability of breast cancer patients. Lastly, Comparing the performance metrics of the conventional ML classifiers with the proposed ensemble classifier indicates the novelty and significance of the study. However, this prediction can encourage patients to consult doctors in a timely manner, thereby saving their lives. Further extensive evaluation of other machine learning classifiers can be carried out using some combinations such as decision tree and random forest to predict the recurrence and survivability of breast cancer patients. Finally, it would be of interest to test the proposed model in a different real-world dataset.

4.1.4.1 CONFUSION MATRIX OF DATA MINING CLASSIFIERS FOR RECURRENCE OF BREAST CANCER.

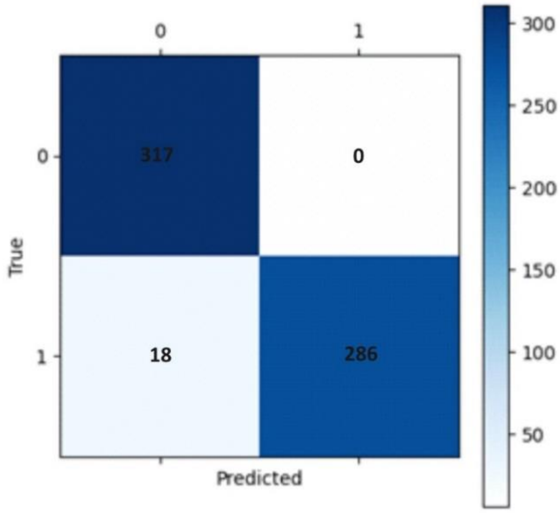

Figure 4: Confusion matrix of ANN-KNN classifier.

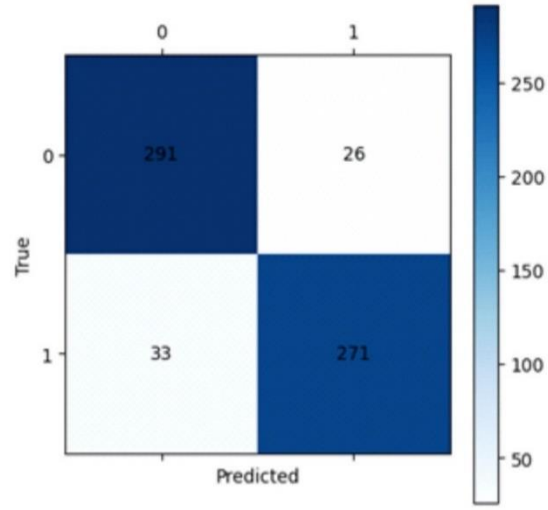

Figure 5: Confusion matrix of KNN classifier.

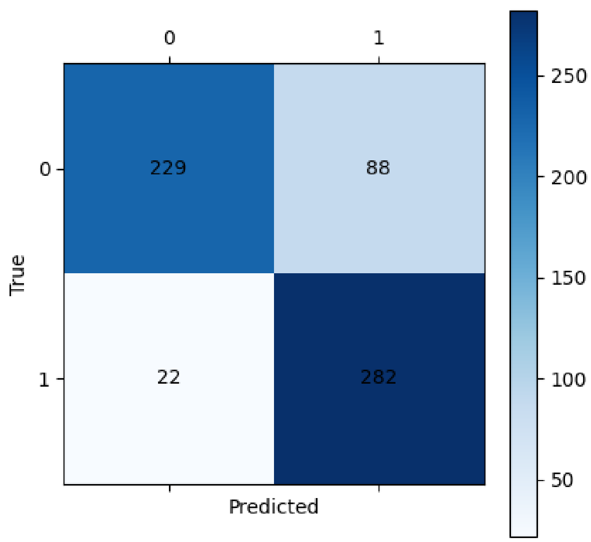

Figure 6: Confusion matrix of SVM classifier.

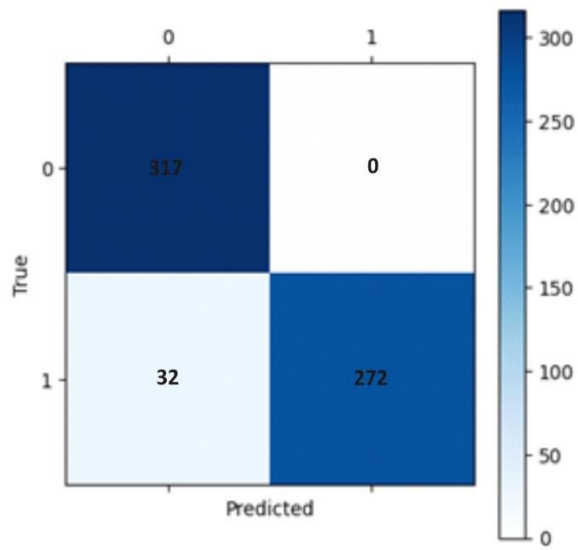

Figure 7: Confusion matrix of ANN classifier.

4.1.4.2 CONFUSION MATRIX OF DATA MINING CLASSIFIERS FOR SURVIVABILITY OF BREAST CANCER PATIENTS.

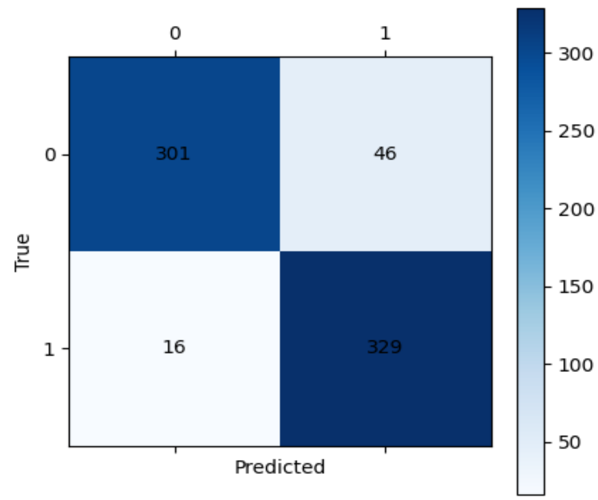

Figure 8: Confusion matrix of ANN-KNN classifier.

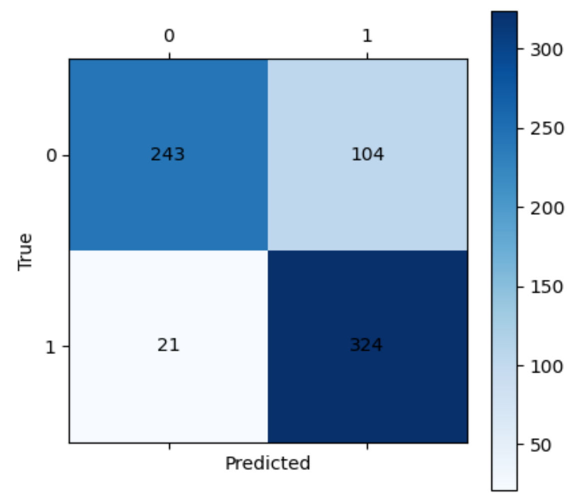

Figure 9: Confusion matrix of KNN classifier.

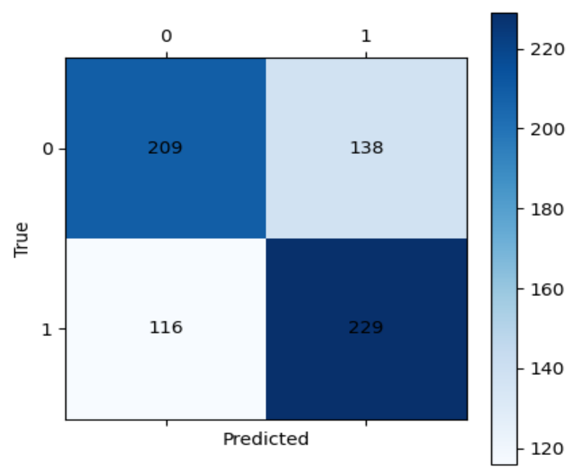

Figure 10: Confusion matrix of SVM classifier.

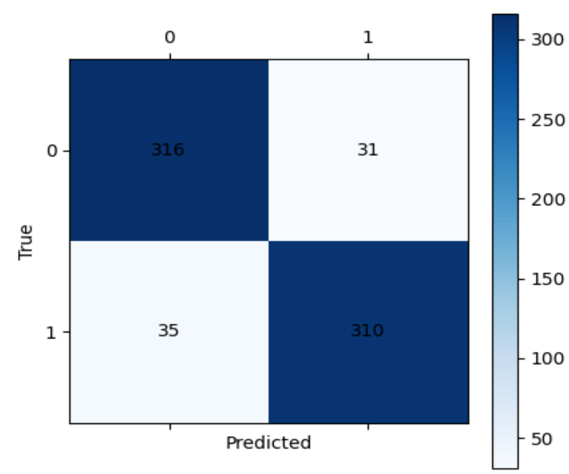

Figure 11: Confusion matrix of ANN classifier.

REFERENCES

[1] Breast Cancer.org (2023). Retrieved online on May 15, 2023.

[2] Ganggayah M.D., Taib N. A., Har Y. C., Lio P. and Dhillon S.K. (2019). Predicting factors for Survival of breast cancer patients using machine learning techniques. *BMC Medical Informatics and Decision making*, **http://doi.org/10.1186/s12911-019-0801-4 19:48**.

[3] abbar, Meerja Akhil. (2021). Breast cancer data classification using ensemble machine learning, *Engineering and Applied Science Research* 2021;48(1):65-72.

[4] WHO (2021). Bulletin on Breast Cancer. Retrieved on July 12, 2021.

[5] Global Cancer Observatory (2020). Information on Cancer. Retrieved online on August 15, 2020.

[6] Musa A. A., Aliyu U.M., (2020). Application of Machine Learning Techniques in Predicting Breast Cancer Metastases Using Decision Tree Algorithm, in Sokoto Northwestern Nigeria, *Journal of Data Mining Genomics Proteomics.* 11:220. DOI: 10.35248/2153- 0602.20.11.220.

[7] Jason Brownlee (2017). *Master Machine Learning Algorithms 'Discover how they work and implement them from scratch'* edition, V1.12, Jason Brownlee publications, USA.

[8] Rivera, D. L., Narvaez, R. A. (2023). Application of Machine Learning Approaches in CancerPrediction. *Canadian Journal of Nursing Informatics,* 18(2). https://cjni.net/jo urn al/ ?p=11581.

[9] Gupta S.R. (2022). Prediction time of breast cancer tumor recurrence using Machine Learning. *Cancer Treatment and Research Communications*. 32. 100602. 10.1016/j.ctarc.2022.100602.

[10] Izci, Hava & Macq, Gilles & Tambuyzer, Tim & De Schutter, Harlinde & Wildiers, Hans & Duhoux, Francois & Azambuja, Evandro & Taylor, Donatienne & Staelens, Gracienne & Orye, Guy & Hlavata, Zuzana & Hellemans, Helga & Rop, Carine & Neven, Patrick & Verdoodt, Freija. (2023). Machine Learning Algorithm to Estimate Distant Breast Cancer Recurrence at the Population Level with Administrative Data. Clinical Epidemiology. Volume 15. 559-568. 10. /CLEP.S400071. 2147.

[11] Pechprasarn, S. & Wattanapermpool, O. & Warunlawan, M. & Homsud, P. & Akarajarasroj, T. (2023). Identification of Important Factors in the Diagnosis of Breast Cancer Cells Using Machine Learning Models and Principal Component Analysis. Journal of Current Science and Technology.13. 642-656.10.59796/ jcst. V13N3.2023.700.

[12] Mosayebi A, Mojaradi B, Bonyadi Naeini A, Khodadad Hosseini SH. (2020). Modeling and comparing data mining algorithms for prediction of recurrence of breast cancer. PLoS One. Oct 15;15(10): e0237658. doi: 10.1371/journal.pone.0237658. PMID: 33057328; PMCID: PMC7561198.

[13] Shikha Agrawal and Jitendra Agrawal (2015). Neural Network Techniques for Cancer Prediction: A Survey, *Procedia Computer Science*, Volume 60, Pages 769-774, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.08.234.

[14] Haque MN, Tazin T, Khan MM, Faisal S, Ibraheem SM, Algethami H, Almalki FA. Predicting Characteristics Associated with Breast Cancer Survival Using Multiple Machine Learning Approaches. (2022). *Comput. Math. Methods Med.* doi: 10.1155/2022/1249692. PMID: 35509861; PMCID: PMC9060999.

[15] Carlos D. C. L., Deevyankar A., Isabel D T., and Lourdes M. R. (2023). Automatic detection of Breast Cancer by using Ensemble Learning, DOI: https://doi.org/10.21203/rs.3.rs-2934498/v1.

[16] Ahmed L.G., Eshiaghy A.T., Poorebrahim A. Ebrahimi M. and Razavi A.R. (2013). Using three Machine learning techniques for predicting breast cancer recurrence, *Journal of Health and Medical Informatics*, 4: 124. Doi:10.4172/2157-7420.100124.

[17] Kalafi EY, Nor NAM, Taib NA, Ganggayah MD, Town C, Dhillon SK. (2019). Machine Learning and Deep Learning Approaches in Breast Cancer Survival Prediction Using Clinical Data. Folia Biol (Praha). 2019;65(5-6):212-220. PMID: 32362304.

[18] Khaoula Chtouki, Maryem Rhanouni, Mounia Mikram, Siham Yousfi and Kamelia Amazian(2023). Supervised Machine Learning for Breast Cancer Risk Factors Analysis and Survival Prediction. *Journal of arXiv:2304.07299v1* [cs.LG] 13.

[19] Li J, Zhou Z, Dong J, Fu Y, Li Y, Luan Z, Peng X. (2021). Predicting breast cancer 5-year survival using machine learning: A systematic review. PLoS One. 2021 Apr 16;16(4):e0250370. doi: 10.1371/*journal.pone.0250370. PMID: 33861809; PMCID:* PMC8051758.

[20] Maria Nisheva, Iliyan Mihaylov, and Dimitar Vassilev (2019). Application of machine Learning models for survival prognosis in breast cancer studies, 93; doi:10.3390/info10030093. *www.mdpi.com/journal/information*10.

[21] Hugo Aerts J.W.L., Yiwen Xu, Ahmed Hosny, Roman Zeleznik, Chintan Parmar, Thibaud Coroller, Idalid Franco, Raymond H. Mak, (2019). Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging. *American Association for Cancer, Research Clin Cancer Res* 25:3266–75 doi: 10.1158/1078-0432.CCR-18-2495.

[22] Dawngliani M.S, Chandrasekaran, N. & Samuel Lalmuanawma (2019). A Comparative Study between Data Mining Classification and Ensemble Techniques for Predicting Survivability of Breast Cancer Patients. *International Journal of Computer Scienceand Mobile Computing*, Vol.8 Issue.9, September- 2019, pg. 01-10.

[23] Dawngliani, M.S., Chandrasekaran, N, Lalmuanawma, Samuel & Thangkhanhau, H (2020). Prediction of Breast Cancer Recurrence Using Ensemble Machine Learning Classifiers.10.1007/978-3-030-46828-6_20.

[24] Dawngliani, M.S, Chandrasekaran, N., Lalmawipuii, R. & Thangkhanhau, H. (2021). Breast Cancer Recurrence Prediction Model Using Voting Technique. 10.1007/978-3-030-49795-8_2.

[25] Anwar Hina, Qamar Usman, Muzaffar Qureshi Abdul Wahab, (2014). "Global Optimization Ensemble Model for Classification Methods", *The Scientific World Journal*, vol. 2014, Article ID 313164, 9 pages https://doi.org/10.1155/2014/313164.

[26] Shahzad, R. K., and Lavesson, N. (2013). Comparative analysis of voting schemes for ensemble-based malware detection. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *4*(1), 98-117.

[27] Prusa, J., Khoshgoftaar, T. M., and Dittman, D. J. (2015). Using ensemble learners to improve classifier performance on tweet sentiment data. In *2015 IEEE international conference on information reuse and integration* (pp. 252-257). IEEE.

[28] WHO (2020). Bulletin on Breast Cancer. Retrieved on September, 2024.

[29] Wang. H., LI, Khan, S.A., and Lou, Y. (2020). Prediction of breast cancer distant recurrence using natural language preprocessing and knowledge-guided convolutional neural network, *Artif. Intell. Med.* 110,101977.

[30] Pilaftis A. and Rubio J. (2015). The Higgs machine learning challenge, Journal of Physics: Conference Series, Vol. 664, no. 7, article 072015.

[31] Maabreh R.S.A., Malik B.A. and Ahmed S.A. (2021). Machine learning algorithms for prediction of survival curves in breast cancer patients. Journal of Applied Bionics and Biomechanics, Vol. 2021, Article ID 9338091, http://doi.org/10.1155/2021.9338091