Hybrid Ensemble Model with Bagging (HEMBAGG) Leveraging Machine Learning for Breast Carcinoma Diagnosis

Vandana Rawat[1*], Devesh Pratap Singh[1], Neelam Singh[1], Rajkumar Singh Rathore[2*], Saad Aldosary[3], Walid El-Shafai[4]

[1] Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun 248002, India
[2] Cardiff School of Technologies, Cardiff Metropolitan University, Cardiff CF5 2YB, United Kingdom
[3] Department of Computer Science, Community College, King Saud University, Riyadh 11437, Saudi Arabia
[4] Department Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt

Corresponding Author Email: vandanarawat2405@gmail.com

ABSTRACT

The sort of cancer that most frequently strikes women is breast cancer. The death rate is significant in contrast to the survival rate, with the vast majority of Indian women who receive a diagnosis of breast cancer having an equal probability of surviving. Breast cancer has surpassed other cancers such as lung, liver, and cervix regarding its frequent diagnosis worldwide. The present research introduces a hybrid ensemble model with a bagging strategy for screening breast cancer utilizing ML techniques. The draught model evaluates the frequently employed ML algorithms and their approaches, which are periodically employed in predicting breast cancer, which include Random Forest, KNN, SVM, decision tree, and logistic regression. The WBCD data collection serves as an experiment set for evaluating the staging of dissimilar algorithms using ML algorithms in terms of key metrics like accuracy, recall, precision, and F-measure. To analyze the identification of cancer of the breast data for forecasting and identification of Breast Cancer, a hybrid ensemble model with bagging (HEMBAGG) is presented in this work. This model comprises an ensemble model with bagging that employs multiple supervised techniques for learning and integrates classifiers to raise the accuracy scores of breast cancer sufferers. The accuracy rate for identifying cancerous tumors in patients in the early stages is increasing with this new proposed model (HEMBAGG), which could potentially save lives.

1. INTRODUCTION

It is among the most widespread kinds of cancer and a substantial global impact on female mortality worldwide [1]. It is a general issue instead of a regional one. It is usually considered to be higher than fifty percent in most countries. The type with the greatest incidence is breast cancer in comparison to other cancer types like liver cancer, cervical cancer, lung cancer, etc. More than three million females are estimated to be affected annually in India. Although breast cancer has no proven medication then in this scenario early detection and diagnosis may significantly enhance survival rates of this among females. The lack of clear symptoms at the beginning of the illness delays diagnosis [2].

Women older than 40 are encouraged by the National Breast Cancer Foundation (NBCF) to schedule mammography every twelve months. According to predictions, more than 2.2 lakh women in India were projected to receive the news of confirmation of breast cancer in 2023, and more than eighty thousand fatalities were anticipated [3]. The 2020 National Cancer Registry Programme Report estimates that by 2025, there will be more than 2.9 lakh new cases [4].

Consequently, according to the WHO, some behaviors and interventions such as continued infant feeding, consistent physical activity, keeping a healthy weight, avoiding drinking too much alcohol, and avoiding exposure to tobacco smoke could potentially reduce the chance of developing cancer.

Mammography, which involves taking a digital image of the breast, is employed to identify breast cancer. Since the technique necessitates a medical strategy which discovers tumors of the breast in female patients without causing any negative consequences, it is safe. The percentage of women who survive is higher for those who regularly undergo mammograms [5].

Over six million individuals died in 2018 as a result of breast carcinoma. Breast carcinoma excision screening and therapy including mammograms, ultrasounds, and biopsies consume quite a bit of time, hence a computerized diagnosis system utilizing a machine learning approach used [6, 7]. Machine learning algorithms and techniques help in early detection and screening. Unsupervised learning and Supervised learning are the two main categories into which the learning process in ML approaches may be differentiated. In supervised learning, a set of data instances that have been

marked with the correct output are used to train the model [8-10]. Because there are no prepared data sets or expertise regarding the intended outcome in the situation of unsupervised education the objective is more challenging to accomplish. Unsupervised machine learning itself draws insights and hidden patterns from data [11]. One of the technologies that is frequently employed in supervised learning is 'Classification' [12]. It develops a model that uses information labeled to generate forecasts about the future [13].

In Figure 1, the survival rate has been shown based on a survey [14, 15]. Clinics and institutions in the medical industry keep substantial databases with patient medical histories, symptoms, and diagnoses [16]. The researchers use this information to design classification models by refining them with algorithms to generate predictions according to patient's prior medical records [17]. Therefore, using the enormous amount of medical data that currently exists, medical inference has become much simpler with machine-based support [18].

All the methods used in this study fall under the category of binary classification models which diagnose whether the tumor is present or not based on certain features [19].

Patients with breast cancer have grown during the past ten years. Breast cancer may be the most common malignancy in women, with almost 2.3 million cases reported annually [20]. Breast carcinoma is the primary or secondary reason for death worldwide among women.
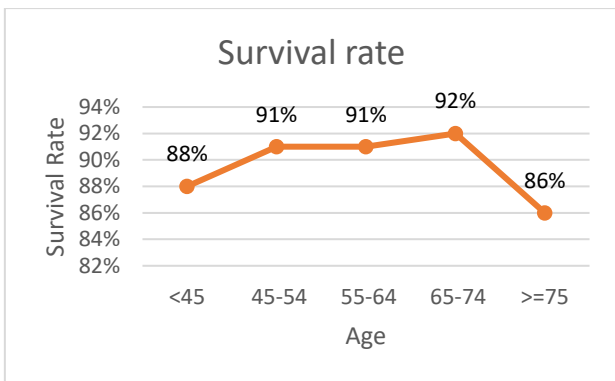
Figure 1. Breast cancer survivors' mortality rate (age-wise)
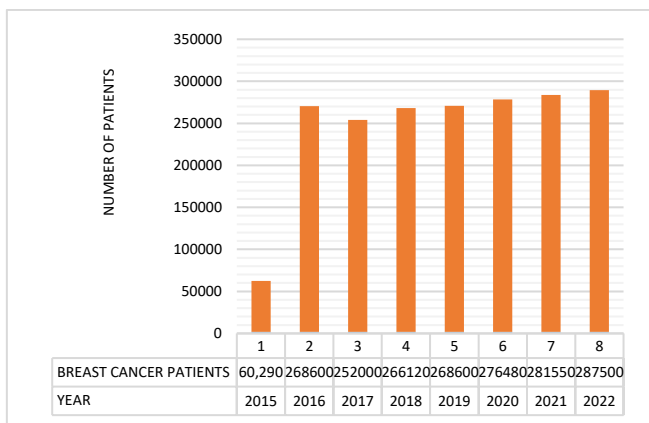
Figure 2. Breast cancer patients' data (year-wise)

Female breast carcinoma has now passed carcinoma of the lungs as the illness that is most frequently diagnosed cancer [21]. There were 2,290,619 new cases of carcinoma of the breast globally during the year 2020 [22] and the graph getting wider every single day. The rise in the number of cancerous breast cancer patients from 2015 to 2023 is depicted in Figure 2 [23].

The above Figure 2 depicts that as per increasing year of age, the proportion of breast carcinoma patient's also increased year-wise [24]. Survival rate was lowest for people >=75 years of age. In every year, increase in age, the chance of survival decreases, and the younger the age, the higher the chances of survival exist [25, 26].

2. LITERATURE REVIEW

Incorporating various machine-learning techniques, researchers have conducted a few associated studies on finding and diagnosing cases of carcinoma of the breast in the past, which is covered in this section [27]. For breast cancer detection, relevant material from various sources is cited. Using the Wisconsin original dataset, SVM was able to diagnose breast cancer with a 97% accuracy rate [28]. Traditional procedures for determining illness and surveillance depend significantly on a human spectator spotting the existence of signal features. Because thousands of patients in hospitals with Intensive Care Units required ongoing monitoring throughout the previous ten years, an assortment of computer-assisted diagnostic (CAD) techniques were created [29]. To address the problem of classification in these systems, the mostly qualitative criteria for diagnosis are transformed into more specific statistical attributes [30-32].

The classification accuracy for various classifiers has been investigated on three different datasets. Sequential minimum optimization and multi-layer perceptron (MLP) classifiers were additionally included in this research in addition to Naive Bayesian, Decision Tree classifiers, and SVM [33-35]. The WBCD original dataset and other Wisconsin datasets, including WBCD (Diagnostic), and WDBC prognostic, were used to test the information in the dataset. For the WPBC dataset, the combination of every single one of the machine learning (ML) approaches mentioned above showed superior cancer recognition performance over [36] conventional methods But with the WDBC dataset, the recommended SMO proved to be a more effective and accurate technique [37].

Additionally, a multi-classifier is suggested in this study by analyzing which hybrid classifier will perform best for the breast cancer dataset [38]. Numerous researchers have used the following machine learning techniques like Bi-clustering, Adaboost Techniques, CNN Classifier, and Bidirectional Recurrent Neural Networks (HA-BIRNN), among additional methods [39-41].

In Table 1, different types of ML techniques are discussed with the accuracy rate. Several algorithms are available that can assist in handling and analyzing even enormous amounts of breast cancer data [39, 40]. These kinds of algorithms belong to the category of machine learning, and two of them are focused on logistic regression and Random Forest that is being utilized in this study of a hybrid model to analyze Wisconsin Diagnosis Breast Cancer data to increase accuracy and precision.

In hybrid Machine Learning, multiple simple algorithms work together to complement and augment each other. So, with numerous studies of the research mentioned, it is confined that by creating any hybrid model using some of the supervised learning algorithms, accuracy could be increased to find the breast cancer patients in the early stages so that the

lives of people could be saved.

Table 1. Evaluation of the results of the most frequently utilized breast carcinoma screening techniques

Authors	Algorithm	Dataset	Accuracy
Sharma et al. [17]	Decision Tree J48 algorithm [18]	Wisconsin	94.56%
Anklesaria et al. [1]	Support Vector Machine	WBCD	95.8%
Al Bataineh [2]	MLP (Multilayer Perceptron)	Wisconsin Data Set	99.12%
Nahid and Kong [3]	Deep neural network	M. G Cancer Hospital Visakhapatnam	97.21%
Tabrizchi et al. [7]	CART with feature selection (Chisquare)	WBCD	94.56%
Mihaylov et al. [9]	Optimized Genetic algorithm	University of Bari Aldo Moro	89.77%
Maajani et al. [14]	Gaussian Mixture Model (GMM)	DDSM	86%
Mihaylov et al. [9]	Optimized Genetic algorithm	University of Bari Aldo Moro	89.77%
Anklesaria et al. [1]	LR (Logistic Regression)	WBCD	95%
Salama et al. [7]	SVM-RBF kernel	WBCD	96.84%
Anklesaria et al. [1]	RF (Random Forest)	WBCD	94.3%
Al Bataineh [2]	NB (Naive Bayes)	Wisconsin Data Set	94.73%
Tabrizchi et al. [7]	Support Vector Machine (SVM)	Wisconsin	94.3%
Anklesaria et al. [1]	K- Nearest Neighbour	WBCD	95.3%
Wang et al. [15]	CNN-LSTM	Kaggle	81.16%
Al Bataineh [2]	MLP	Wisconsin Data Set	99.2%
Al Bataineh [2]	CART	Wisconsin Data Set	93.85%
Liu et al. [19]	SVM	WBCD	95.80%
Anklesaria et al. [1]	Decision Tree	WBCD	92.9%

3. METHODOLOGY

In this section, proposed methodology is working with ensemble model with bagging on multiple machine learning models for analyzing the breast cancer patients in early phases so that the lifespan of these patients can be increased.

3.1 Dataset

The dataset for this breast cancer research has been taken from WBCD (Wisconsin breast cancer dataset), which has 569 instances with 30 attributes, where 357 are benign and 212 patients are malignant. This dataset was included with missing values in it and some of the data was unstructured.

Every parameter value in the WBCD data has a distinct numerical measurement value. With the exception of ID and diagnosis results, all other numerical measurement values for this investigation were normalized using the following calculation:

$$Y' = Y - min(Y)/max(Y) - min(Y) \qquad (1)$$

Here Y is an actual value and Y' is the normalized value.

3.2 Performance evaluation metrics

In this research, ensemble model is used with bagging so after preprocessing of the dataset, following performance evaluation metrics [42, 43] is calculated:

$$Accuracy = (True\ Positive + True\ Negative)/ \\ (True\ Positive + True\ Negative \\ + False\ Positive \\ + Flase\ Negative) \qquad (2)$$

$$Precision = True\ Positives/(True\ Positives \\ + False\ Positives) \qquad (3)$$

$$Recall = True\ Positive/(True\ Positive \\ + Flase\ Negative) \qquad (4)$$

$$F1\ Score = 2 * ((precision * Recall)/(Precision \\ + Recall)) \qquad (5)$$

Here in the above equations, True Positive, False Negative, True Negative and False Positive, are indicated for correct prediction and false prediction.

3.3 Bagging

This procedure, which is also known as an ensemble method, is bootstrap aggregating. Several models are trained individually on arbitrary portions of the data, and their predictions are then combined by averaging the outcome.

Following this, the bagging classifier trains several machine learning models separately on arbitrary data subsets before considering the median of their forecasts.

3.4 Model description

Machine learning models act as weak learners when applied individually on the dataset and the ensemble model works in a better way once applied on a dataset with machine learning algorithms so here in this proposed work, HEMBAGG is applied to machine learning algorithms like on KNN, decision tree, logistic regression, Random Forest, XG Boost, support vector machine etc to enhance the result.

Firstly, data extraction is performed, then data preparation, data analysis, implementation of the model, and the proposed hybrid model is defined. The Wisconsin Diagnosis Breast Cancer data has been employed in the present research to address the issue of prompt patient detection of breast cancer. The intention is to categorize and detect malignant individuals at the earliest stages. Implementing the suggested HEMBAGG entails the following measures.

4. PROPOSED FRAMEWORK

In this research, the major focus is on cancerous and noncancerous breast cancer patients. So, in this section, HEMBAGG is employed as follows:

4.1 Data extraction

To examine how effectively different machine learning algorithms are implemented, the training set was the Wisconsin data. The code is implemented using the Jupyter with Pandas package, a Python Integrated Development Environment (IDE) featuring advancements in testing, editing, and numerical computing environments. In this dataset firstly, data extraction techniques are applied for the mentioned data set.

4.2 Data preparation

This is a Data Mining technique to transform the dataset into an understandable format. It is the process of preparing a dataset so that it can be further processed and analyzed as done in various research. As data often contain missing values, inaccuracies, and other errors, the missing values are found. The blank fields are replaced with null, and duplicate columns are removed so that the dataset becomes optimized and more useful.

4.3 Exploratory data analysis

In this, the data set is analyzed in all its possible modes. In this fitting or training the model is the process of identifying patterns in the data. The most common fields are to be identified to fit into this model. This step is the important step in the analysis.

4.4 Selection and implementation of a model

Several Machine Learning Algorithms have already been implemented to discover the optimum accuracy. Then, the Ensemble method is employed, which is a technique that integrates numerous models rather than utilizing just one to improve the accuracy of outcomes in models. In the Ensemble technique, the Bagging method is utilized to improve model accuracy through decision trees, which significantly reduces variance.

4.5 Hybrid ensemble model

Whether the tumor is benign or malignant it can be predicted using a hybrid model called the HEMBAGG and applied to various ML algorithms like logistic regression, naive bayes and Random Forest, etc.

Algorithm: Ensemble model with Bagging classifier

Step 1: Start
Step 2: Breast cancer dataset acquired as input.
Step 3: Read the dataset file.
Step 4: Establish the dataset's dimensions and columns.
Step 5: Provide a summary of the data frame.
Step 6: Make a summary statistics calculation for the data.
Step 7: Store the file in a CSV file.
Step 8: Based on the "diagnosis" column, create a binary target variable.
Step 9: Check missing data, if yes, go to STEP 10.
Step 10: Split the data into the target variable (y) and features (X).
Step 11: Check all fields of the dataset for positive and Negative.
Perform test and train split.
Step 12: Create an ensemble model with a bagging classifier with entropy for a maximum depth of 4.
Step 13: Make predictions of test data by splitting the model into the training set.
Step 14: Create an ensemble model with a bagging classifier for confusion matrix and accuracy score.
Step 15: Make predictions of test data after fitting the ensemble model to the training set.
Step 16: Determine and print the ensemble model's accuracy score.
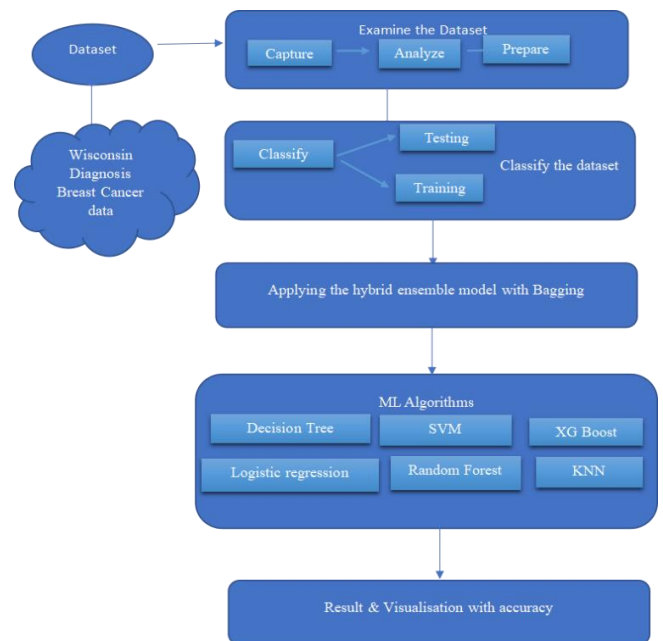Step 17: End

Figure 3. Systematic architecture of the proposed HEMBAGG

In above Figure 3, the flow of representation of the proposed hybrid ensemble model is depicted.

5. RESULT AND DISCUSSION

In this research, a breast cancer dataset has been taken of 569 patients with 30 attributes from WBCD as defined in the methodology. Some of the machine learning algorithms like DT, KNN, LR, SVM, RF, XGB, etc. applied individually to this dataset. By using these algorithms, with the help of confusion matrix, it has been calculated that these algorithms are giving better accuracy than other machine learning algorithms as shown in Table 2 but by enhancing the algorithm with an ensemble model with bagging, it is improved.

According to the above described and proposed hybrid ensemble model, to forecast the labels for the test data (X_test), the model must first be trained through the initial training data (X_train and y_train) using an ensemble model with bagging. It is computed using the anticipated labels, and its results are

compared to the actual labels (y_test). In the below figures, the confusion matrix score is displayed using an ensemble model with bagging on the decision tree for the WBCD dataset.

Table 2. ML algorithms accuracy on WBCD dataset

ML Algorithms	Accuracy
Logistic regression	62.940%
XG Boost	95.102%
Support vector machine	62.940%
KNN	76.921%
Decision Tree	92.312%
Random Forest	95.801%
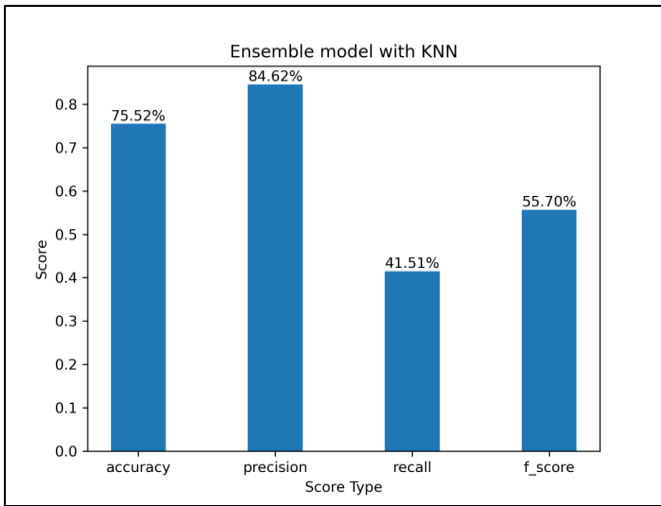
Figure 4. Metrics score for ensemble model using bagging with KNN
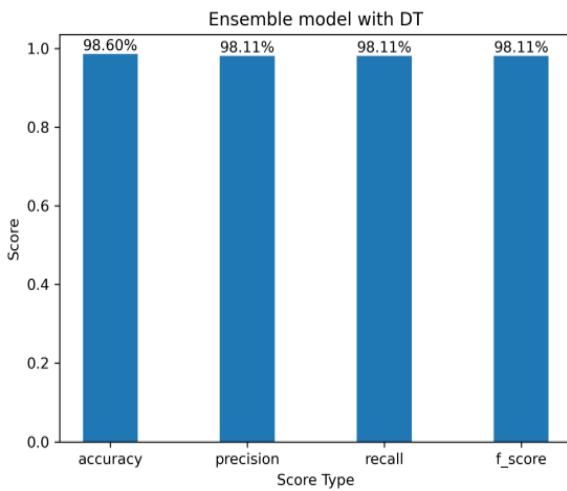
Figure 5. Metrics score for ensemble model with decision tree

Figure 4 shows the metric scores for the ensemble model with KNN. Figure 4 shows the ensemble model's performance scores with KNN serving as the base estimator. Each score is shown as a labeled bar with the height of the bar corresponding to the score value. The graph makes it simple to evaluate and analyze the result of the ensemble model with the performance measures visually.

The results indicate that the overall derived precision rating is only 84.62% and accuracy 75.52% when the ensemble model with KNN is used to assess the accuracy, precision,

recall, f-score, and support of patients classified as benign and malignant.

Figure 5 shows the ensemble model's performance scores with the Decision tree serving as the base estimator. Each score is shown as a labeled bar with the height of the bar corresponding to the score value. The graph makes it simple to finalize and analyze the result of the ensemble model with the performance measures visually.

The results indicate that the overall derived precision rating is 98.11% and accuracy is 98.60%, when the ensemble model with the Decision tree is used to assess the accuracy, precision, recall, f-score, and support of patients classified as benign and malignant.

Figure 6 shows the ensemble model's performance scores with Random Forest serving as the base estimator. Each score is shown as a labeled bar with the height of the bar corresponding to the score value. This graph shows the score types for the ensemble model with Random Forest.

The results indicate that the overall derived precision rating is 91.23% and accuracy is 95.80%, when the ensemble model with Random Forest is used to assess the accuracy, precision, recall, f-score, and support of patients classified as benign and malignant. Using an ensemble model with a Random Forest gives an accuracy of 95.80%.
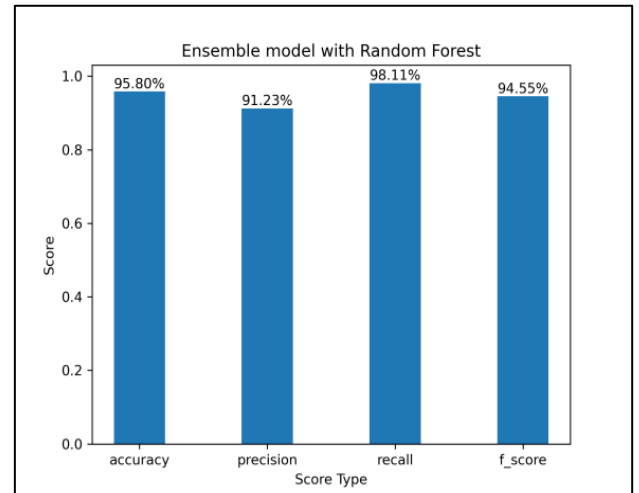
Figure 6. Metrics score for ensemble model with Random Forest
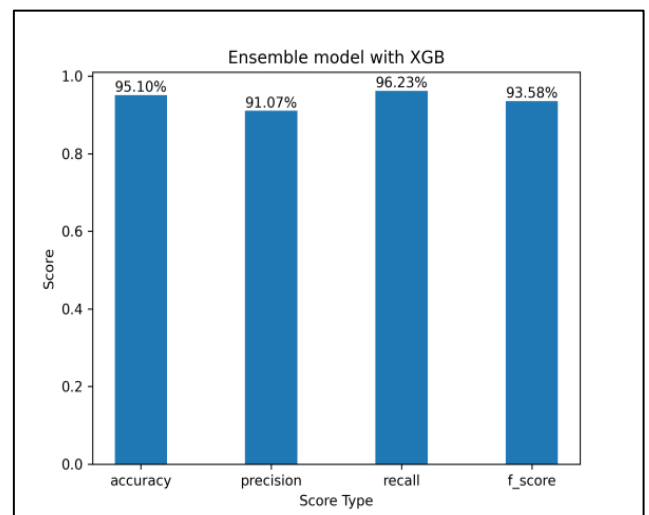
Figure 7. Metrics score for ensemble model with XGB boost

**Table 3.** HEMBAGG (Ensemble model with bagging) accuracy score

HEMBAGG Model	Accuracy Score
Logistic regression (LR)	62.940%
XG Boost (XGB)	95.102%
Support vector machine (SVM)	62.940%
K-nearest neighbour (KNN)	75.522%
Decision Tree (DT)	98.601%
Random Forest (RF)	95.801%

Table 4. Comparative accuracy score with ML algorithms and HEMBAGG model

ML Algorithms	ML Algorithms Accuracy Score	HEMBAGG Model Accuracy Score
Decision Tree (DT)	92.312%	98.601%
K-nearest neighbour (KNN)	76.921%	75.522%
Logistic regression (LR)	62.940%	62.940%
Support vector machine (SVM)	62.940%	62.940%
Random Forest (RF)	95.801%	95.801%
XG Boost (XGB)	95.102%	95.102%

Table 5. Accuracy score of decision tree with different parameters

Method	Accuracy Score
Decision Tree	92.31%
Proposed Model (HEMBAGG)	98.6%
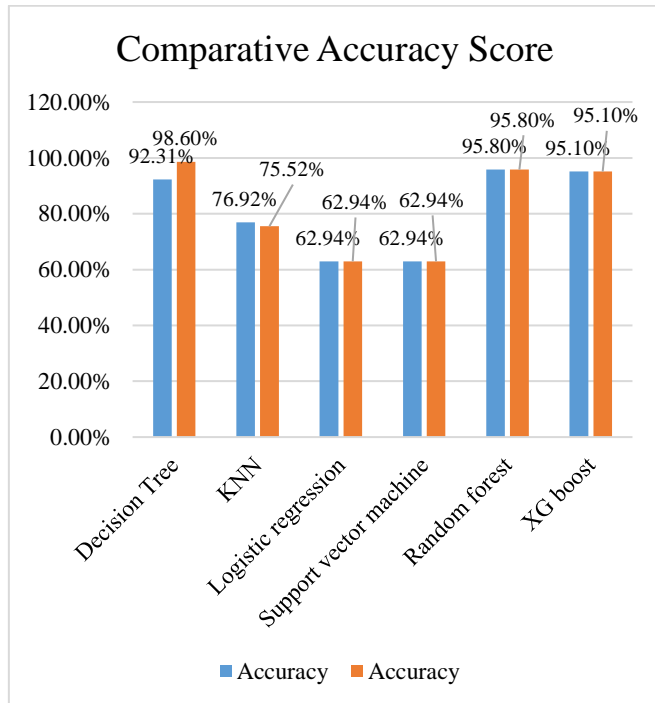
Figure 8. Accuracy comparison of ml algorithms vs. HEMBAGG model

Figure 7 shows the ensemble model's performance scores with XGB boost serving as the base estimator. Each score is shown as a labelled bar with the height of the bar

corresponding to the score value. The results indicate the accuracy i.e. 95.10% and the overall derived precision rating is only 91.07%, when the ensemble model with XGB boost is used to assess the accuracy, precision, recall, f-score, and support of patients classified as benign and malignant.

To increase the prediction score percentage for breast cancer patients, in the same algorithms, a HEMBAGG is used. Based on the given model, we can conclude that the decision tree is performing more efficiently and giving higher accuracy as compared to the rest of the algorithms shown in Table 3 and Figure 8.

In above Table 4, the accuracy score of ML algorithms is compared with the HEMBAGG model, and comparatively high accuracy is obtained with the new hybrid model (HEMBAGG) that is shown in Table 4, and the graph is also displayed below with the accuracy score of both of the algorithms.

HEMBAGG model overcomes the drawbacks of the poor generalization of the traditional machine learning models by overcoming the problem of over-fitting and providing better accuracy, robustness, and generalization.

Table 4 demonstrates the accuracy comparison of ML Algorithms like DT, KNN, LR, SVM, RF, and XG Boost with the Accuracy of the HEMBAGG Model. In most of the algorithm cases the accuracy is not being changed but and Decision Tree and K-nearest neighbor it is changed and the HEMBAGG model with the Decision Tree gives the highest prediction.

The above Table 5 represents the accuracy score of the decision tree which is a base estimator with different parameters. So, for this confusion matrix is calculated first and it is found that the ensemble method with bagging on decision tree classifier is giving a very high accuracy.

Figure 9 shows the graph of labels of accuracy scores of the Decision tree with different parameters. Each score is shown as a labeled bar, with the height of the bar corresponding to the score value. The graph provides a visual comparison of the performance metrics, allowing for easy interpretation and analysis of the ensemble model's effectiveness with the proposed ensemble method with bagging which enhances the result.

The actualization of the code for the hybrid ensemble model with a bagging classifier is employed to forecast breast carcinoma patients which uses a development framework with Python, also known as Jupyter with Pandas library. This environment has advanced in the numerical environment. The ensemble model is an enhanced version of the decision tree and with bagging, it determines the accuracy, precision, recall, and f1-score. The result shows that the overall derived precision rating of classifiers is 92.31% and 98.6%, respectively, as shown in Table 5 and Figure 9. So, we can conclude that the HEMBAGG model has increased the accuracy using the ensemble method so that people's lives could be saved in the early stage of cancer.

In the end, it is concluded that in various research independent machine learning algorithms have been applied but, in those researches, the result is not as efficient as it is predicted with the proposed model i.e., ensemble model with bagging because the ensemble model is a group of multiple algorithms. It generates better results compared with already existing research.
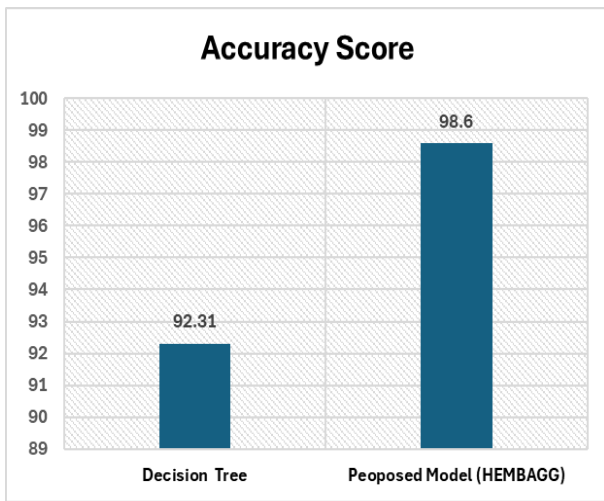
Figure 9. Accuracy with HEMBAGG with decision tree

6. CONCLUSIONS

In this study, a HEMBAGG is proposed for early-stage detection of breast carcinoma. The ensemble model is incorporated with bagging on the breast carcinoma data set. This is applied with bagging on Decision Tree, KNN, DT, LR, and SVM, and with decision tree, it gives the highest accuracy. Comparing and analyzing the success rate and accuracy of prominent ML models including KNN, DT, SVM, and LR, firstly accuracy of any model is obtained using a Confusion Matrix which is a table of actual values and predicted values with accuracy Score. The performance of the new proposed HEMBAGG model with bagging has been applied to the decision tree with a great accuracy of 98.6%, which is the highest accuracy in comparison to other algorithms. Additional dimensional reduction methods, particle swarm optimization, etc. could be applied to this data set to enhance the performance of specific models.

REFERENCES

[1] Anklesaria, S., Maheshwari, U., Lele, R., Verma, P. (2022). Breast cancer prediction using optimized machine learning classifiers and data balancing techniques. In 2022 6th International Conference on Computing, Communication, Control and Automation (ICCUBEA, Pune, India, pp. 1-7. https://doi.org/10.1109/ICCUBEA54992.2022.10010783

[2] Al Bataineh, A. (2019). A comparative analysis of nonlinear machine learning algorithms for breast cancer detection. International Journal of Machine Learning and Computing, 9(3): 248-254. https://doi.org/10.18178/ijmlc.2019.9.3.794

[3] Nahid, A.A., Kong, Y. (2017). Involvement of machine learning for breast cancer image classification: A survey. Computational and Mathematical Methods in Medicine, 2017(1): 3781951. https://doi.org/10.1155/2017/3781951

[4] Bhattacharjee, A., Patil, S., Talole, M.S., Singh, A., Chaturvedi, P., Dikshit, R. (2020). An impact of reduction in point prevalence of tobacco use on cancer incidence-A challenge for global policy makers. Clinical Epidemiology and Global Health, 8(4): 1287-1296. https://doi.org/10.1016/j.cegh.2020.04.029

[5] Naderi, M.A., Mahdavi-Nasab, H. (2010). Analysis and classification of EEG signals using spectral analysis and recurrent neural networks. In 2010 17th Iranian Conference of Biomedical Engineering (ICBME), Isfahan, Iran, pp. 1-4. https://doi.org/10.1109/ICBME.2010.5704931

[6] Bashivan, P., Rish, I., Yeasin, M., Codella, N. (2015). Learning representations from EEG with deep recurrent-convolutional neural networks. arXiv Preprint arXiv: 1511.06448. https://doi.org/10.48550/arXiv.1511.06448

[7] Tabrizchi, H., Tabrizchi, M., Tabrizchi, H. (2020). Breast cancer diagnosis using a multi-verse optimizer-based gradient boosting decision tree. SN Applied Sciences, 2(4): 752. https://doi.org/10.1007/s42452-020-2575-9

[8] Ghasemzadeh, A., Sarbazi Azad, S., Esmaeili, E. (2019). Breast cancer detection based on Gabor-wavelet transform and machine learning methods. International Journal of Machine Learning and Cybernetics, 10: 1603-1612. https://doi.org/10.1007/s13042-018-0837-2

[9] Mihaylov, I., Nisheva, M., Vassilev, D. (2018). Machine learning techniques for survival time prediction in breast cancer. In Artificial Intelligence: Methodology, Systems, and Applications: 18th International Conference, AIMSA 2018, Varna, Bulgaria, Proceedings. Springer International Publishing. Springer, Cham, 18: 186-194. https://doi.org/10.1007/978-3-319-99344-7_17

[10] Mohammed, S.A., Darrab, S., Noaman, S.A., Saake, G. (2020). Analysis of breast cancer detection using different machine learning techniques. In Data Mining and Big Data: 5th International Conference, DMBD 2020, Belgrade, Serbia, pp. 108-117. https://doi.org/10.1007/978-981-15-7205-0_10

[11] Benbrahim, H., Hachimi, H., Amine, A. (2020). Comparative study of machine learning algorithms using the breast cancer dataset. In Advanced Intelligent Systems for Sustainable Development (AI2SD'2019) Volume 2-Advanced Intelligent Systems for Sustainable Development Applied to Agriculture and Health. Springer International Publishing. Springer, Cham, pp. 83-91. https://doi.org/10.1007/978-3-030-36664-3_10

[12] Chandra, P., Tan, Y.N., Singh, S.P. (Eds.). (2017). Next generation point-of-care biomedical sensors technologies for cancer diagnosis. Singapore: Springer Singapore, 10: 978-981. https://doi.org/10.1007/978-981-10-4726-8

[13] Mojrian, S., Pinter, G., Joloudari, J.H., Felde, I., Szabo-Gali, A., Nadai, L., Mosavi, A. (2020). Hybrid machine learning model of extreme learning machine radial basis function for breast cancer detection and diagnosis; a multilayer fuzzy expert system. In 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), Ho Chi Minh City, Vietnam, pp. 1-7. https://doi.org/10.1109/RIVF48685.2020.9140744

[14] Maajani, K., Jalali, A., Alipour, S., Khodadost, M., Tohidinik, H.R., Yazdani, K. (2019). The global and

regional survival rate of women with breast cancer: A systematic review and meta-analysis. Clinical Breast Cancer, 19(3): 165-177. https://doi.org/10.1016/j.clbc.2019.01.006

[15] Wang, X., Ahmad, I., Javeed, D., Zaidi, S.A., Alotaibi, F.M., Ghoneim, M.E., Daradkeh, Y.I., Asghar, J., Eldin, E.T. (2022). Intelligent hybrid deep learning model for breast cancer detection. Electronics, 11(17): 2767. https://doi.org/10.3390/electronics11172767

[16] Dewangan, K.K., Dewangan, D.K., Sahu, S.P., Janghel, R. (2022). Breast cancer diagnosis in an early stage using novel deep learning with hybrid optimization technique. Multimedia Tools and Applications, 81(10): 13935-13960. https://doi.org/10.1007/s11042-022-12385-2

[17] Sharma, S., Aggarwal, A., Choudhury, T. (2018). Breast cancer detection using machine learning algorithms. In 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Belgaum, India, pp. 114-118. https://doi.org/10.1109/CTEMS.2018.8769187

[18] Abdollahi, J., Keshandehghan, A., Gardaneh, M., Panahi, Y., Gardaneh, M. (2020). Accurate detection of breast cancer metastasis using a hybrid model of artificial intelligence algorithm. Archives of Breast Cancer, 22-28. https://doi.org/10.32768/abc.20207122-28

[19] Liu, N., Qi, E.S., Xu, M., Gao, B., Liu, G.Q. (2019). A novel intelligent classification model for breast cancer diagnosis. Information Processing & Management, 56(3): 609-623. https://doi.org/10.1016/j.ipm.2018.10.014

[20] Ferro, S., Bottigliengo, D., Gregori, D., Fabricio, A.S., Gion, M., Baldi, I. (2021). Phenomapping of patients with primary breast cancer using machine learning-based unsupervised cluster analysis. Journal of personalized medicine, 11(4): 272. https://doi.org/10.3390/jpm11040272

[21] Altaf, M.M. (2021). A hybrid deep learning model for breast cancer diagnosis based on transfer learning and pulse-coupled neural networks. Mathematical Biosciences and Engineering, 18(5): 5029-5046. https://doi.org/10.3934/mbe.2021256

[22] Sivakami, K., Saraswathi, N. (2015). Mining big data: Breast cancer prediction using DT-SVM hybrid model. International Journal of Scientific Engineering and Applied Science (IJSEAS), 1(5): 418-429.

[23] Carvalho, D., Pinheiro, P.R., Pinheiro, M.C.D. (2016). A hybrid model to support the early diagnosis of breast cancer. Procedia Computer Science, 91: 927-934. https://doi.org/10.1016/j.procs.2016.07.112

[24] Burçak, K.C., Uğuz, H. (2022). A new hybrid breast cancer diagnosis model using deep learning model and ReliefF. Traitement du Signal, 39(2): 521-529. https://doi.org/10.18280/ts.390214

[25] Naji, M.A., El Filali, S., Aarika, K., Benlahmar, E.H., Abdelouhahid, R.A., Debauche, O. (2021). Machine learning algorithms for breast cancer prediction and diagnosis. Procedia Computer Science, 191: 487-492. https://doi.org/10.1016/j.procs.2021.07.062

[26] Wang, L., Qian, Q., Zhang, Q., Wang, J., Cheng, W., Yan, W. (2022). Classification model on big data in medical diagnosis based on semi-supervised learning. The Computer Journal, 65(2): 177-191. https://doi.org/10.1093/comjnl/bxaa006

[27] Peng, L., Chen, W., Zhou, W., Li, F., Yang, J., Zhang, J. (2016). An immune-inspired semi-supervised algorithm for breast cancer diagnosis. Computer Methods and Programs in Biomedicine, 134: 259-265. https://doi.org/10.1016/j.cmpb.2016.07.020

[28] Zeng, Y., Xu, X. (2023). Label diffusion graph learning network for semi-supervised breast histological image recognition. Biomedical Signal Processing and Control, 80: 104306. https://doi.org/10.1016/j.bspc.2022.104306

[29] Mohanasundaram, R., Malhotra, A.S., Arun, R., Periasamy, P.S. (2019). Deep learning and semi-supervised and transfer learning algorithms for medical imaging. In Deep Learning and Parallel Computing Environment for Bioengineering Systems. Academic Press, pp. 139-151. https://doi.org/10.1016/B978-0-12-816718-2.00015-4

[30] Azmi, R., Norozi, N., Anbiaee, R., Salehi, L., Amirzadi, A. (2011). IMPST: A new interactive self-training approach to segmentation suspicious lesions in breast MRI. Journal of Medical Signals and Sensors, 1(2): 138-148.

[31] Wodzinski, M., Ciepiela, I., Kuszewski, T., Kedzierawski, P., Skalski, A. (2021). Semi-supervised deep learning-based image registration method with volume penalty for real-time breast tumor bed localization. Sensors, 21(12): 4085. https://doi.org/10.3390/s21124085

[32] Podgorsak, A.R., Kumaraswamy, L.K. (2021). Semi-supervised planning method for breast electronic tissue compensation treatments based on breast radius and separation. Radiology and Oncology, 55(1): 106-115. https://doi.org/10.2478/raon-2020-0073

[33] Yin, X.X., Yin, L., Hadjiloucas, S. (2020). Pattern classification approaches for breast cancer identification via MRI: State-of-the-art and vision for the future. Applied Sciences, 10(20): 7201. https://doi.org/10.3390/app10207201

[34] Benabdeslem, K., Mansouri, D.E.K., Makkhongkaew, R. (2020). SCOs: Semi-supervised co-selection by a similarity preserving approach. IEEE Transactions on Knowledge and Data Engineering, 34(6): 2899-2911. https://doi.org/10.1109/TKDE.2020.3014262

[35] Liang, H., Zou, J. (2020). Rock image segmentation of improved semi-supervised SVM-FCM algorithm based on chaos. Circuits, Systems, and Signal Processing, 39: 571-585. https://doi.org/10.1007/s00034-019-01088-z

[36] Thakur, B., Kumar, N. (2022). Prediction, detection and recurrence of breast cancer using machine learning based on image and gene datasets. Recent Innovations in Computing: Proceedings of ICRIC 2021, 1: 263-273. https://doi.org/10.1007/978-981-16-8248-3_21

[37] Lin, H., Ji, Z. (2020). Breast cancer prediction based on K-Means and SOM Hybrid Algorithm. In Journal of Physics: Conference Series, 1624(4): 042012. https://doi.org/10.1088/1742-6596/1624/4/042012

[38] Jahangeer, G.S.B., Rajkumar, T.D. (2021). Early detection of breast cancer using hybrid of series network and VGG-16. Multimedia Tools and Applications, 80: 7853-7886. https://doi.org/10.1007/s11042-020-09914-2

[39] Şahan, S., Polat, K., Kodaz, H., Güneş, S. (2007). A new hybrid method based on fuzzy-artificial immune system and k-NN algorithm for breast cancer diagnosis. Computers in Biology and Medicine, 37(3): 415-423. https://doi.org/10.1016/j.compbiomed.2006.05.003

[40] Jebarani, P.E., Umadevi, N., Dang, H., Pomplun, M. (2021). A novel hybrid K-means and GMM machine

learning model for breast cancer detection. IEEE Access, 9: 146153-146162.

[41] Padmanaban, K., Senthil Kumar, A.M., Azath, H., Velmurugan, A.K., Subbiah, M. (2023). Hybrid data mining technique based breast cancer prediction. In AIP Conference Proceedings. AIP Publishing, 2523(1): 020047. https://doi.org/10.1063/5.0110216

[42] Mittal, D., Gaurav, D., Roy, S.S. (2015). An effective hybridized classifier for breast cancer diagnosis. In 2015 IEEE International Conference on Advanced Intelligent Mechatronics (AIM), Busan, Korea (South), pp. 1026-1031. https://doi.org/10.1109/AIM.2015.7222674

[43] Folorunso, S.O., Awotunde, J.B., Adigun, A.A., Prasad, L.N., Lalitha, V.L. (2023). A hybrid model for post-treatment mortality rate classification of patients with breast cancer. Healthcare Analytics, 4: 100254. https://doi.org/10.1016/j.health.2023.100254