Pertanika J. Sci. & Technol. 33 (S3): 157 - 180 (2025)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Breast Cancer Prediction: A Random Forest-based System with Expert Validation

Sharifah Nurulhikmah Syed Yasin*, Aiman Azhar and Rajeswari Raju

College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Cawangan Terengganu, Kampus Kuala Terengganu, 21080 Kuala Terengganu, Terengganu, Malaysia

ABSTRACT

Breast cancer (BC) is a fatal invasive disease among women that impacts women globally. It is listed as a significant disease among Malaysian women. Early detection and accurate diagnosis are important to improve the treatment outcome of a patient, as advanced stages of BC can increase fatality rates. The conventional methods of diagnosis are effective, but they face challenges such as high cost, radiation exposure, and the need for specialized operators. Therefore, this study focuses on developing a BC prediction system using a Random Forest (RF) algorithm. It is trained using the "BC Wisconsin (Diagnostic) Data Set" from Kaggle, consisting of 570 records with eight critical attributes selected for prediction. The algorithm and system are developed using Python and evaluated on accuracy, precision, recall, and F1-score, achieving 91.23%, 90.70%, 86.67%, and 88.89%, respectively. The algorithm was integrated with AdaBoost and XGBoost to add the experimental value, resulting in a better result than a single RF. Expert validation by a specialist confirmed the reliability of the dataset and accuracy of the prediction system, highlighting its potential to be a valuable tool for early BC detection. The study concludes that the RF-based system provides robust predictions, making it a promising approach for enhancing BC diagnostic processes.

Keywords: Algorithm, breast cancer, machine learning, prediction, random forest

ARTICLE INFO

Article history:
Received: 22 August 2024
Accepted: 24 February 2025
Published: 24 April 2025

DOI: https://doi.org/10.47836/pjst.33.S3.10

E-mail addresses:
nurulhikmah@uitm.edu.my (Sharifah Nurulhikmah Syed Yasin)
aimanazhar377@gmail.com (Aiman Azhar)
rajes332@uitm.edu.my (Rajeswari Raju)
*Corresponding author

INTRODUCTION

Breast cancer (BC) has increasingly become a common invasive disease in women, while it remains rare in men. BC forms the malignant cells within the breast tissue (Kinra, 2019; Minnoor & Baths, 2022). Patients may experience symptoms such as a breast lump, bloody nipple discharge, and alterations in the shape of the nipple or

e-ISSN: 2231-8526

breast (Kinra, 2019; Minnoor & Baths, 2022). In Malaysia, BC was the most common cancer from the year 2007–2016. The percentage increased from 17.7% in 2007–2011 to 19% in 2012–2016. This type of cancer has also been the most common cancer among Malaysian women, where 34.1% of the cancers reported was BC (National Cancer Registry, 2019). While the number of BC patients has increased in Malaysia, the disease has also impacted women globally. According to the World Health Organization (WHO), 2.3 million women were diagnosed with BC in 2022, with 670,000 of them succumbing to the disease.

Early detection and diagnosis of BC are vital for enhancing patient outcomes. Identifying BC at an early, localized stage greatly increases the likelihood of successful treatment and cure (Breast cancer, 2024). Moreover, an early and accurate diagnosis can increase survival rates, offering a better cure result and reducing the need for aggressive treatments (Li et al., 2024). On the contrary, advanced BC, particularly in stage four, often involves circulating tumor cells that drastically lower survival rates to as low as 40% (Zuo et al., 2017).

The most practical methods to carry out the diagnosis are performing clinical breast tests, mammograms, ultrasound tests, molecular breast imaging (MBI), magnetic resonance imaging (MRI), blood tests, and breast biopsy (He et al., 2020). Nevertheless, these methods have encountered several challenges, such as high cost, radiation exposure, requiring professional operators, long imaging times, and restrictions for patients with metal implants (He et al., 2020; Park et al., 2024). Therefore, there are potential solutions to these problems where a machine learning-based clinical prediction system can fill this gap and assist in the early identification of BC (Duan et al., 2024; Macaulay et al., 2021; Minnoor & Baths, 2022).

Machine learning (ML) can help the BC diagnosis process by predicting and classifying it based on the previous diagnosis data. ML analyses huge amounts of data containing the factors or symptoms of previously diagnosed, labeled data (Duan et al., 2024; Macaulay et al., 2021; Minnoor & Baths, 2022). The popular ML algorithms used in the previous study to predict disease are the decision tree, Support Vector Machine (SVM), K-nearest neighbor, multilayer perceptron, and random forest (RF). Among this algorithm, RF is found to produce the highest accuracy prediction (Mohamed et al., 2023; Rashid et al., 2024; Sumwiza et al., 2023).

The previous studies on RF for BC prediction conclude their research after accuracy testing and rarely extend the research to include expert validation. Therefore, this study seeks to develop and verify a prediction system for BC by adapting RF algorithms. The development is physically carried out using Python language, followed by accuracy testing using accuracy, precision, and recall metrics and result validation by a BC specialist consultant. In addition, the dataset utilized in this study is collected from the Kaggle website the dataset's name is "BC Wisconsin (Diagnostic) Data Set", which has 570 data person which has the BC attributes.

RELATED WORKS

Macaulay et al. (2021) study develops a predictive model for BC risk using an RF Classifier in African women. This study compares the results with the previous work that adopted the Gail model. The data involved in the prediction model are self-reported risk factor data and BMI values. Eleven significant risk factors were identified, including benign breast disease, a history of cancer, pesticide use, age at first child, exercise, and fruit intake. The study emphasizes the importance of these factors in predicting BC risk. The dataset was divided into training (70%) and testing (30%) sets during development. The RF classifier has undergone training using the selected features in the dataset. The output of the developed system shows high accuracy (98.33%) and sensitivity (100%) in predicting BC. This result has shown that the developed algorithm outperformed the previous Gail model. The main contribution of this study is that the algorithm proposed specifically addresses the unique risk profile of African women and has a high accuracy score.

Another study by Minnoor and Baths (2022) focuses on developing an automated system for BC diagnosis using an RF algorithm. This study emphasizes the importance of early diagnosis and aims to create a model to effectively classify malignant and benign tumors. This study utilizes the Wisconsin BC Diagnostic dataset from the UCI Machine Learning Repository, which contains 569 labeled instances of tumors (212 malignant and 357 benign) to train the RF engine. The dataset is imbalanced and needs to be fixed using upscaling techniques. Initially, the dataset contained seventeen key factors; however, this study chose eleven factors to use. The factors are diagnosis, symmetry, concavity, area, texture, compactness, radius, smoothness, concave points, perimeter, and fractal dimension. The reduction of the key factors is done to enhance computational efficiency. Consequently, the RF model surpasses the other machine learning algorithms evaluated, attaining a high accuracy rate of 99.3% in diagnosing malignant tumors.

Duan et al. (2024) conducted a study that created a machine learning-based prediction model for distant metastasis in BC. Distant metastasis refers to the spread of cancer cells from the primary tumor to other body parts. This study aims to identify the potential of biomarkers related to distant metastasis by using various bioinformatics techniques like weighted gene co-expression network analysis (WGCNA), differential expression analysis, and LASSO regression analysis. Therefore, 21 biomarkers related to distant metastasis were labeled and derived from a dataset analysis. Some machine learning models were trained using the recognized biomarkers, including logistic regression, RF, gradient boosting decision trees (GBDT), support vector machines (SVM), and XGBoost. The result shows that the RF model was the best-performing model for predicting distant metastasis, with a 93.6% accuracy score.

Yifan et al. (2021) recommended a method to improve the accuracy of BC diagnosis by combining two machine learning algorithms: RF and AdaBoost. This study aims to create

a classification model able to differentiate between benign and malignant breast tumors by using ML algorithms. This study applies to the Wisconsin Diagnostic BC Database, involving 569 samples (212 malignant and 357 benign) with 32 attributes relating to tumor characteristics. StandardScaler is employed to confirm a stable and standardized dataset. RF and AdaBoost are integrated to improve accuracy and effectively convert the classifier. As a result, the integrated model showed impressive results, with an accuracy of 98.6%. Therefore, this study concludes that integrating two ML algorithms (RF and AdaBoost) enhances the accuracy of prediction for BC diagnosis. Similarly, integrating RF with XGBoost has enhanced the imbalance dataset handling, as reported by Natras et al. (2022). XGBoost has shown an improvement of 6% in the overall performance of RF, where XGBoost leverages gradient boosting and regularization techniques. The purpose is to improve predictive accuracy while mitigating overfitting. Overall, the hybrid approach enhances model performance and addresses some limitations of RF, like vulnerability to class imbalance and challenges in interpretability.

The reviewed studies have revealed that RF has the capacity to predict BC in a patient very well. Nevertheless, none of the studies have shown expert validation of the results. Hence, this study works on developing a prediction system for BC, followed by validation of the results by a medical specialist. Table 1 provides a comparative analysis of the reviewed studies, detailing the datasets used and their respective percentages of accuracy.

Table 1
Comparative analysis of the related works

Author	Year & Publication	Dataset	Accuracy
Macaulay et al. (2021)	2021 Cancer Treatment and Research Communications	180 subjects of African women in Lagos State, Nigeria, with 90 confirmed as BC cases and 90 benign cases	Accuracy: 91.67% Sensitivity: 87.10% Specificity: 96.55% Area Under Curve (AUC): 92%
Minnoor & Baths (2022)	2022 International Conference on Machine Learning and Data Engineering	Wisconsin BC Diagnostic dataset of UCI Repository. 569 instances (samples) of tumors, with 212 classified as malignant and 357 classified as benign.	Initial dataset with 16 features: 100% Minimal dataset with eight features: 99.3%
Duan et al. (2024)	2024 Computers in Biology and Medicine	Gene Expression Omnibus (GEO) GSE9893 Dataset: 155 samples, 48 developed distant metastasis, while 107 did not. GSE43837 Dataset: 38 samples with 19 patients had developed distant metastasis, and 19 had not.	Accuracy: 93.6% F1-score: 88.9% Area Under Curve (AUC): 91.3%.

Table 1 *(continue)*

Author	Year & Publication	Dataset	Accuracy
Yifan et al. (2021)	2021 2021 IEEE 3rd International Conference on Communications, Information System and Computer Engineering (CISCE 2021)	Wisconsin BC Diagnostic dataset, UCI Machine Learning Repository. 569 samples of tumors, with 212 labeled as malignant and 357 labeled as benign.	Accuracy: 98.6%

Bootstrapping is widely used in random forests, particularly for breast cancer prediction. It creates multiple subsets of the original dataset through resampling, which helps reduce variance and improve overall prediction stability. Multiple studies show that this approach can also lessen the risk of overfitting by allowing more accurate uncertainty estimates (Ishwaran & Lu, 2019; Mentch & Zhou, 2020).

At the same time, bootstrapping can increase computational requirements, especially in high-dimensional settings or when data is limited (Ishwaran & Lu, 2019). Despite these challenges, it remains a valuable method for building random forest models to detect subtle patterns in breast cancer data and provide more reliable predictions.

Comparative Analysis of Machine Learning Models in Breast Cancer Prediction

This study further explores RF and compares it with other machine learning algorithms to deepen understanding. The main models explored in this study include Support Vector Machine (SVM), Logistic Regression (LR) and Random Forest (RF). Each model has its unique strengths and limitations, which must be understood in relation to each other to choose the optimal model for a given task. Rashidi et al. (2019) and Shehab et al. (2022) explain Artificial Intelligence (AI) and Machine Learning (ML) in their study, outlining the basic concepts, types, and applications across healthcare, finance, and transportation. The explanation covers key ML techniques, including supervised, unsupervised, and reinforcement learning, focusing on their use in predictive modeling. The study also highlights challenges such as data quality, model interpretability, and bias and looks at the future potential of AI/ML. The summary of this study is sorted in Table 2.

Based on Table 2, each machine learning model—Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LG)—offers unique advantages and limitations in breast cancer prediction. Random Forest excels in performance and feature interpretation but is limited by its lower interpretability and computational intensity. SVM is effective for high-dimensional and non-linear data but requires significant hyperparameter tuning and may struggle with large datasets. Logistic Regression is simple and computationally efficient but often underperforms in complex scenarios and is sensitive to outliers. Therefore, model selection should align with the dataset's characteristics, interpretability needs, and available computational resources (Rashidi et al., 2019; Shehab et al., 2022).

Table 2
Comparative table of strengths and limitations of ML models in breast cancer prediction

Model	Strength	Limitation
Support Vector Machine (SVM)	High accuracy, good for non-linear data, effective in high-dimensional spaces, robust to overfitting.	Requires significant computational power, expensive, sensitive to kernel choice and hyperparameter tuning, hard to interpret, limited scalability.
Logistic Regression (LR)	Simple, interpretable, fast to train, effective for linearly separable data.	Struggles with non-linear relationships and data, prone to underfitting with complex data, and sensitive to outliers.
Random Forest (RF)	It is robust against overfitting, handles large and complex datasets well, provides feature importance, high accuracy and robustness, and there is no need for feature scaling.	Computationally intensive, slow to predict, limited interpretability compared to LR and high memory usage.

MATERIALS AND METHODS

This study utilizes the RF algorithm to predict BC in a patient. The approach is designed to process user-inputted data via an interface that receives eight significant attributes to predict the disease's existence. The RF model starts with bootstrapping subsets of the dataset to construct multiple decision trees. Each tree uses information gain entropy to determine the best attribute for splitting the data at each node. This approach guarantees that the model captures complex interactions between tumor characteristics. Predictions are made across all individual trees within the group, and the outcome is determined through a majority voting mechanism. This method enhances the robustness and generalizability of predictions, providing reliable assessments of whether a tumor is malignant or benign. The system framework of this study is shown in Figure 1.

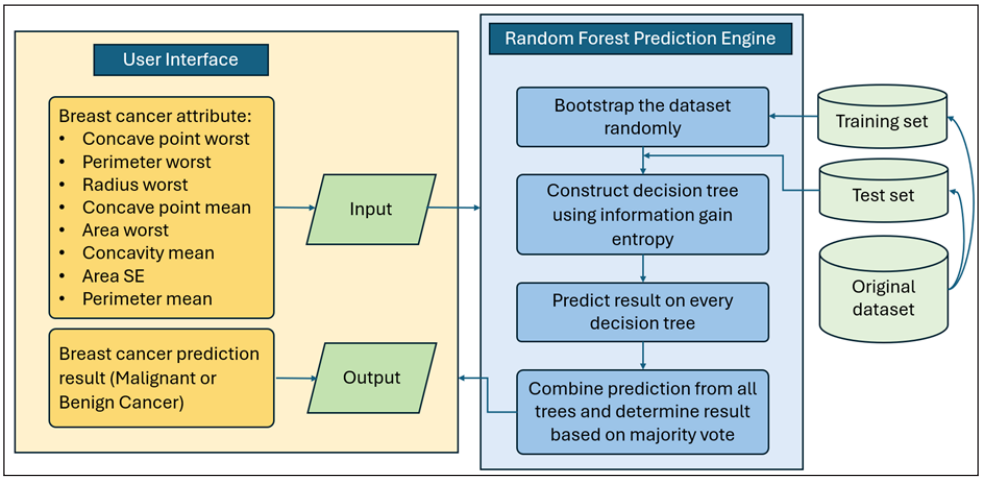

Figure 1. System framework

Dataset Selection and Preparation

Dataset Selection

The first step in the prediction system process is to gather the dataset. The dataset for this study is employed from the Kaggle platform: the BC Wisconsin (Diagnostic) Data. The dataset encompasses 570 medical records, each derived using the Fine Needle Aspiration (FNA) technique. FNA is a quick and straightforward procedure that involves extracting fluid or cells from a breast lesion or cyst using a thin needle, like those used for blood draws. The dataset was compiled by a physician at the University of Wisconsin Hospital, Dr William H. Wolberg. The dataset records BC diagnoses in 30 dimensions. This dataset is already cleaned up in the aspect of missing data by the creator of the dataset. The initial attributes are 30 and are reduced to 8, the top important attributes contributing to the algorithm learning. This study discovered Principal Component Analysis (PCA) as an extra feature selection method. PCA converts the dataset into a set of orthogonal components ranked by the amount of variance they capture, thus lowering dimensionality while retaining the most critical information. This method enhanced the RF model's computational effectiveness and predictive ability. It can recognize and prioritize key patterns within the 30-dimensional dataset. Using PCA, the system could effectively recognize the most informative features and eliminate redundancies. However, for this study, the manual selection of eight features was believed to be optimal based on their domain relevance and statistical importance. Figure 2 shows the importance of each attribute among many contributing factors between all 30 attributes (Dai et al., 2018).

Variable	Score
CONCAVE_POINTS_WORST	100.00
PERIMETER_WORST	79.32
RADIUS_WORST	71.99
CONCAVE_POINTS_MEAN	63.48
AREA_WORST	56.40
CONCAVITY_MEAN	30.88
AREA_SE	28.75
PERIMETER_MEAN	28.23
AREA_MEAN	25.19
CONCAVITY_WORST	24.13
RADIUS_MEAN	18.60
PERIMETER_SE	7.63
RADIUS_SE	7.30
COMPACTNESS_MEAN	4.85
COMPACTNESS_WORST	3.55
TEXTURE_WORST	3.36
SYMMETRY_WORST	2.78
CONCAVITY_SE	2.56
SMOOTHNESS_WORST	2.04
TEXTURE_MEAN	1.57
FRACTAL_DIMENSION_WORST	1.31
TEXTURE_SE	.56
SMOOTHNESS_MEAN	.50
SYMMETRY_SE	.48
FRACTAL_DIMENSION_MEAN	.46
CONCAVE_POINTS_SE	.42
SMOOTHNESS_SE	.41
FRACTAL_DIMENSION_SE	.39
SYMMETRY_MEAN	.34
COMPACTNESS_SE	.19

Figure 2. Level of importance for each attribute

The blue line in Figure 2 represents the significance of auxiliary diagnosis, implying the attributes that impact the prediction. For instance, the "concave point worst" has numerous blue lines, indicating a greater impact on the prediction result than the "concavity mean," which has fewer blue lines. After knowing the top 8 importance attributes, the original dataset is copied to create a new set encompassing 570 sets of data with only eight important

attributes, which are concave point worst, perimeter worst, radius worst, concave point mean, area worst, concavity mean, area standard error (SE), perimeter mean and target. The dataset has also been labeled with 1 for malignant cancer and 0 for benign cancer.

Import Dataset to Python

Several predefined Python libraries are employed to preprocess the dataset. Pandas and NumPy are the libraries used for data preprocessing. Panda is a library used to import and manage datasets. NumPy is a library used for arrays, matrices, and various tools for working with arrays, which is very useful for machine learning. After these predefined libraries are installed, the dataset is brought in through the read_csv() function of the Pandas library from the dataset file data.csv (Table 3).

Table 3
Imported dataset

No	Concave Point Worst	Perimeter Worst	Radius Worst	Concave Point Mean	Area Worst	Concavity Mean	Area Standard Error	Perimeter Mean	Target
1	0.2654	184.6	25.38	0.1471	2019	0.3001	153.4	122.8	1
2	0.186	158.8	24.99	0.07017	1956	0.0869	74.08	132.9	1
3	0.243	152.5	23.57	0.1279	1709	0.1974	94.03	130	1
:	:	:	:	:	:	:	:	:	:
568	0.1418	126.7	18.98	0.05302	1124	0.09251	48.55	108.3	1
569	0.265	184.6	25.74	0.152	1821	0.3514	86.22	140.1	1
570	0	59.16	9.456	0	268.6	0	19.15	47.92	0

It is crucial in machine learning to differentiate the feature matrix, consisting of independent and dependent variables in the dataset. In the *data.csv* dataset, the independent variables are the eight features: the concave point worst, perimeter worst, radius worst, concave point mean, area worst, concavity mean, area standard error (SE), and perimeter mean. The dependent variable is the target, referred to in the last column in Table 3. The *iloc[]* method of the Pandas library will be used to extract an independent variable. The method's function is to extract the first eight columns in the dataset, which are the independent variables, and store them into a NumPy array variable named 'X' as shown in Figure 3(a). A NumPy array variable 'Y' is applied for dependent variables (diagnosis results in the dataset), as shown in Figure 3(b).

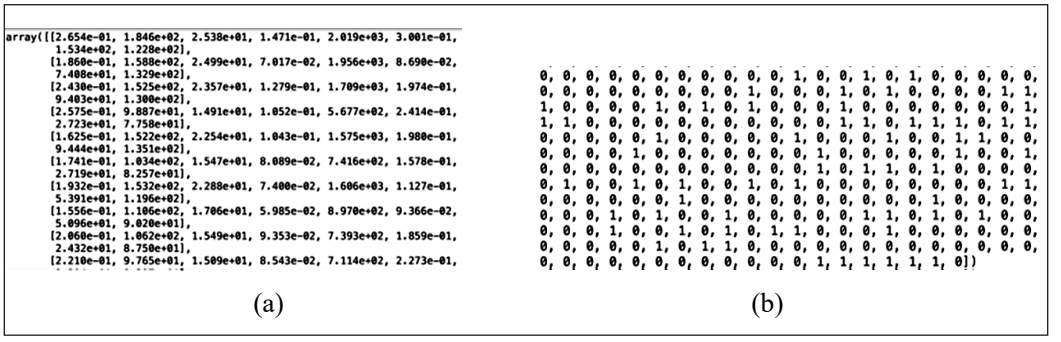

Figure 3. Extracted independent and dependent variables from the dataset

Splitting the Dataset into the Training and Test Set

Splitting the dataset into training and test sets is a crucial step in data preprocessing, as it improves the machine learning model's performance. The training set consists of data with known expected outputs used to train the model. Conversely, the test set assesses the model's accuracy by comparing predicted outputs with actual results. Figure 4 displays the code for splitting the dataset.

```
Initialize the random seed using random_state.
Shuffle the rows of X and y in unison to ensure matching feature/label pairs.
Calculate the test set size as test_size × total number of samples.
Split X into X_train (first portion) and X_test (remaining portion).
Split y into y_train (first portion) and y_test (remaining portion).
Return X_train, X_test, y_train, y_test.
```

Figure 4. Pseudocode to split the dataset

The code specifies that *X_train* represents the independent variables from the previous step used for training data, while *X_test* is for testing data. Similarly, *y_train* denotes the dependent variables from the previous step used for training, and *y_test* is for testing. A test size of 0.2 means that 80% of the data is allocated for training and 20% for testing. The random state value of 1234 ensures the reproducibility of the results.

Random Forest Prediction Algorithm

The RF prediction engine consists of four steps: bootstrapping the dataset randomly, constructing decision trees using information gain entropy, predicting results on every decision tree, combining predictions from all trees, and determining a result based on a majority vote.

Random Bootstrap Dataset

In an RF, bootstrapping means randomly picking data points (with the chance of picking the same one multiple times) from the original dataset to form smaller training sets for each tree. This way, each tree learns from a slightly different set of examples, which helps the whole forest make more accurate and stable predictions.

Based on Table 4, the dataset has been bootstrapped to allow each decision tree in the forest to be trained on a different data set and reduce overfitting. Moreover, Table 4 shows only one bootstrap dataset. The number of bootstrapped datasets depends on the number of trees constructed in the RF. If there are 100 trees, there will be 100 bootstrap datasets created from the original dataset. Figure 5 shows the code for bootstrapping the dataset.

```
Determine the total number of samples in X (n_samples).
Generate n_samples random indices (idxs) from the range [0..n_samples - 1] with replacement
Extract the rows in X corresponding to idxs, store in X_boot.
Extract the rows in y corresponding to idxs, store in y_boot.
Return X_boot, y_boot.
```

Figure 5. Bootstrap pseudocode

This function is called *bootstrap_samples*, which creates a bootstrapped sample of the input data represented by *X* and *y*. This bootstrapped sample is created by randomly selecting elements from *X* and *y* from the dataset with replacement, forming a smaller dataset. This function calculates the number of samples in *X* and then generates an array of indices using the NumPy function *np.random.choice* randomly. Finally, the function returns the *X* and *y* elements corresponding to the selected indices forming the bootstrapped sample.

Table 4
Bootstrap dataset

No	Concave Point Worst	Perimeter Worst	Radius Worst	Concave Point Mean	Area Worst	Concavity Mean	Area Standard Error	Perimeter Mean	Target
1	0.2654	184.6	25.38	0.1471	2019	0.3001	153.4	122.8	1
3	0.243	152.5	23.57	0.1279	1709	0.1974	94.03	130	1
202	0.108	92.15	14.44	0.04107	638.4	0.04187	27.24	78.54	0
548	0.02381	71.12	11.25	0.005495	384.9	0.01012	12.97	65.31	0
:	:	:	:	:	:	:	:	:	:
1	0.2654	184.6	25.38	0.1471	2019	0.3001	153.4	122.8	1

Construct Decision Trees Using Information Gain Entropy

The decision tree in the RF is constructed using entropy information gain as the impurity measure. Entropy measures the impurity or randomness within a dataset. It serves as a criterion for constructing decision trees to divide the data into homogeneous groups. Figure 6 shows the code for the decision tree construct using entropy information gain.

```
Calculate the parent entropy using the function entropy(y).
Split the data using split(X_column, threshold), which returns:
- left_idxs  : indices for the left subset
- right_idxs : indices for the right subset
If either left_idxs or right_idxs is empty, return 0.
(No information gain can be obtained if a split results in an empty set.)
Determine the total number of samples, n = length(y).
Compute the size of each split:
- n_l = length(left_idxs)
- n_r = length(right_idxs)
Calculate the entropy of each child set:
- e_l = entropy(y[left_idxs])
- e_r = entropy(y[right_idxs])
Compute the weighted average of these child entropies:
- child_entropy = (n_l / n) * e_l + (n_r / n) * e_r
Subtract the child entropy from the parent entropy to get the information gain:
- information_gain = parent_entropy - child_entropy
Return information_gain.
```

Figure 6. Decision tree construct using entropy information gain pseudocode

In this code, the function created is named *_information_gain*, which calculates the information gain to create the decision tree. The first line of the method assigns the parent entropy, which is the entropy of the target variable, before splitting. Next, the data is split based on the feature *X_column* and threshold value using the *_split* method, which returns two arrays of indices *left_idxs* and *right_idxs* corresponding to the decision tree's left and right branches. Then, the method calculates the entropy of each group and takes a weighted average to obtain the child entropy. The weight of each group is related to the number of samples in each group. Finally, the method calculates the information gain by subtracting the weighted average entropy of the children from the parent entropy.

Make a Prediction on Every Decision Tree

Many decision trees are created in the RF algorithm, each producing a prediction output regarding whether a given sample is malignant or benign. Each tree produces a prediction output indicating whether a given case is malignant or benign. This prediction process occurs for every tree in the RF algorithm. Figure 7 shows the code for prediction on every decision tree created.

```
Initialize an empty list, predictions.
For each instance x in X:
a. Use the function traverse_tree(x, root_of_decision_tree) to obtain a prediction.
b. Append this prediction to the predictions list.
Convert the predictions list to an array.
Return the array of predictions.
```

Figure 7. Prediction on every decision tree pseudocode

This code predicts every decision tree created based on the bootstrap dataset. The prediction's result is stored in an array of *X*.

Combine Predictions from All Trees and Determine the Result Based on the Majority Vote

Lastly, this RF engine determines the prediction based on the majority result, calculated by tallying the results from all decision trees. The prediction is obtained by determining which category (malignant or benign) has more occurrences, which involves summing up all the 1s and 0s and selecting the category with the greater count. Figure 8 shows the code used to combine tree predictions and determine the result.

This code will take a set of inputs (*X*) and run the prediction function for each decision tree in the set of trees stored in the model. It then collects the predictions for each sample, switches the axis to group the predictions for each sample, and finally chooses the one with majority votes from the tree predictions for each sample as the final prediction.

```
Initialize an empty list called tree_predictions.
For each tree in trees:
a. Obtain the tree's predictions for all instances in X, call this partial_preds.
b. Append partial_preds to tree_predictions.
Convert tree_predictions to a 2D array with shape (#trees, #instances).
Swap its axes to have shape (#instances, #trees), so each row now corresponds to a
single instance and contains predictions from each tree.
For each row (set of predictions for one instance):
a. Determine the majority-vote label (e.g., using a function most_common_label).
b. Add this label to a new list of final_predictions.
Convert final_predictions to an array.
Return final_predictions.
```

Figure 8. Combine prediction from all trees pseudocode

Hybrid Model Development

To further evaluate the robustness and performance of the prediction system, hybrid models were developed by integrating Random Forest (RF) with boosting algorithms such as AdaBoost and XGBoost (Refer Table 5). For the RF-AdaBoost hybrid, misclassified occurrences from RF were iteratively reweighted to enhance prediction accuracy. At the same time, XGBoost was used for its gradient-boosting capabilities and effective handling of imbalanced datasets. Both models were trained using the same dataset, with hyperparameters optimized through grid search to ensure fair comparisons. Each hybrid model's performance metrics, including accuracy, precision, recall, and F1-score, were calculated to assess their improvements over standalone RF.

Table 5
Configurations for hybrid models

Model	Algorithm Description	Key Parameters	Optimization Technique
RF (Standalone)	Ensemble of decision trees using majority voting.	Number of trees: 100, Max depth: 10	Random Search
RF + AdaBoost	Boosting misclassified instances iteratively with RF as a base.	Learning rate: 0.1, Number of estimators: 50	Grid Search
RF + XGBoost	Gradient boosting with RF as a base.	Learning rate: 0.1, Max depth: 6, Subsampling: 0.8	Grid Search

RF + AdaBoost

AdaBoost refines the Random Forest (RF) base estimator in this hybrid approach by iteratively adjusting the weights of misclassified instances. This process enables the model to concentrate on more challenging samples in subsequent iterations. The aggregated predictions are derived through a weighted majority voting mechanism, where greater weight is assigned to trees with higher accuracy, enhancing the final output. Key hyperparameters, including the number of estimators and the learning rate, were optimized through grid search to balance computational efficiency and predictive accuracy.

RF + XGBoost

The integration with XGBoost leverages its advanced gradient-boosting capabilities to further refine Random Forest (RF) predictions. XGBoost is particularly effective in handling imbalanced datasets, utilizing regularization techniques and tree pruning to reduce overfitting while maintaining robust predictive performance. RF serves as the base learner in this framework, with XGBoost applied iteratively to enhance its predictions. Key hyperparameters, such as the learning rate, maximum tree depth, and subsampling ratio,

were optimized using grid search to achieve superior model performance. Both hybrid models were trained and tested on the same dataset as the standalone Random Forest (RF) model to ensure direct comparability. Their effectiveness was evaluated using standard performance metrics, including accuracy, precision, recall, and F1-score.

User Interface

The user interface is created using *Streamlit*, an open-source framework mainly used to build data science or machine learning web apps. In this user interface, the user needs to input eight BC attributes: the concave point worst, perimeter worst, radius worst, concave point mean, area worst, concavity mean, area standard error (SE), and perimeter mean. Then, the user can click the Predict button to get the cancer result. Figure 9 shows the user interface.

Accuracy Evaluation

Accuracy assessment is vital in determining machine learning algorithms' efficiency. The process entails comparing the predicted outputs to the actual results and calculating the proportion of correct predictions. The outcome of this evaluation provides valuable information on the model's strengths and weaknesses and enables identifying areas for improvement. In this study, a comprehensive analysis of the model's accuracy is performed to determine its suitability for its intended purpose and make any necessary modifications to enhance its performance. There are four important things for accuracy testing: accuracy, precision, recall, and f1 score. The formulas for each of the tests are as follows.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad [1]$$

$$Precision = \frac{TP}{(TP+FP)} \quad [2]$$

$$Recall = \frac{TP}{(TP+FN)} \quad [3]$$

$$F1 - Score = 2 \times \frac{PRECISION \times RECALL}{PRECISION + RECALL} \quad [4]$$

Accuracy measures the proportion of correct predictions made by the model out of the total predictions. Precision is the ratio of true positive predictions to all positive predictions made by the model. Recall, or sensitivity, is the proportion of true positive predictions relative to the total number of actual positive samples. The F1 score, the harmonic mean

of precision and recall, provides a balanced metric that considers both. A high F1 score indicates a well-balanced model in terms of precision and recall. In the formulas, TP stands for True Positive, TN for True Negative, FP for False Positive, and FN for False Negative. These values are all represented in the confusion matrix, a table used in machine learning to assess a classifier's performance by showing the counts of true positive, true negative, false positive, and false negative predictions.

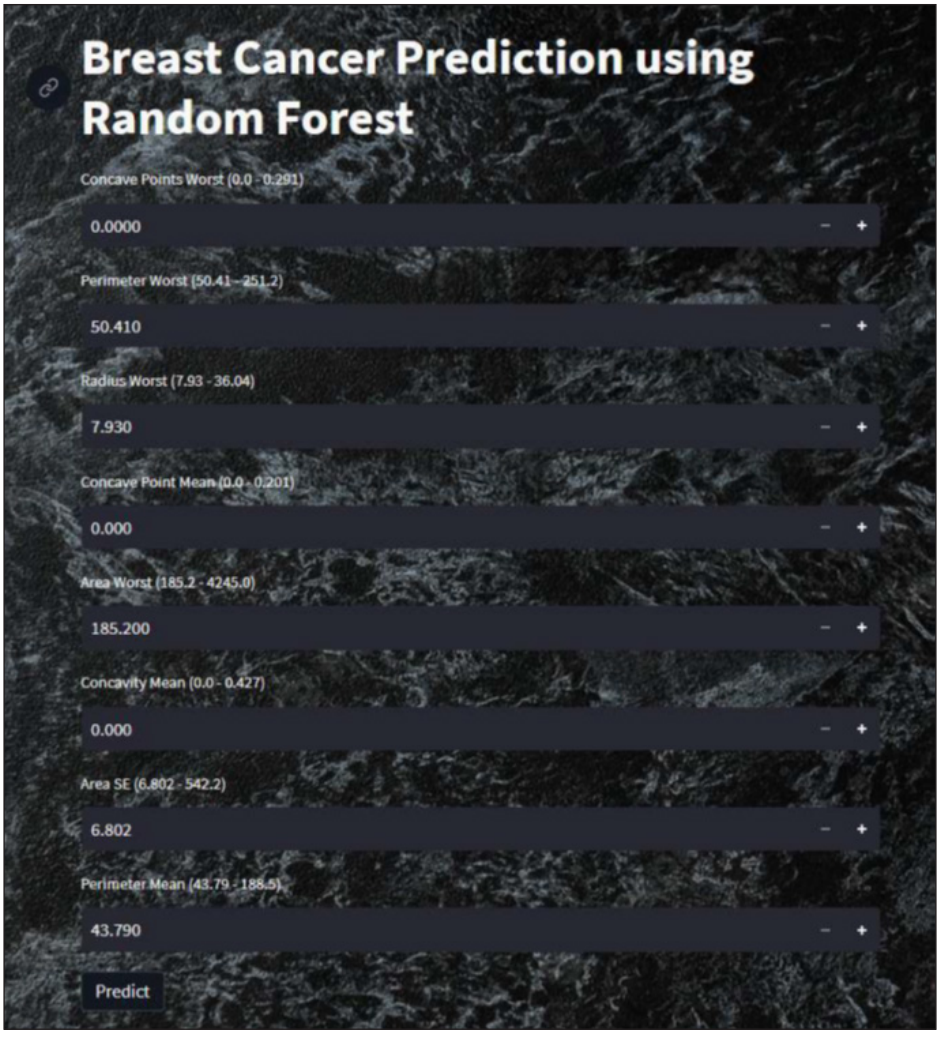

Figure 9. User interface of the prediction system

Expert Evaluation

Selecting a qualified medical expert was a critical step to ensure accurate evaluation. (Moosavi et al., 2024; Vazquez-Zapien et al., 2022) and validation of the research objectives.

The following criteria guided the identification and recruitment of the medical professional for this study, adopted and modified from (Moosavi et al., 2024):

a. Clinical Expertise in Breast Oncology

The medical expert was required to have specialized training in breast oncology and demonstrate extensive experience in diagnosing and treating breast cancer. This encompassed expertise typically found among oncologists, breast cancer surgeons, or radiologists focusing on breast imaging. These qualifications ensured the evaluator's capability to provide detailed and clinically relevant insights.

b. Familiarity with Diagnostic Standards

The expert needed to be well-versed in clinical guidelines, diagnostic criteria, and screening protocols for breast cancer to ensure alignment with current medical practices. This knowledge was necessary to ensure that the feedback and assessments were grounded in present standards.

c. Comfort with Data Analysis and Technology

Although advanced technical expertise was not a requirement, the expert was expected to have a basic understanding of computer predictive algorithms and associated performance metrics. This foundational knowledge enabled the evaluator to engage meaningfully with the study's model validation and effectively interpret its outputs.

d. Objective and Independent Perspective

The expert was selected with careful consideration of potential conflicts of interest to maintain the integrity and neutrality of the evaluation process. Preference was given to candidates without involvement in developing the predictive model, ensuring unbiased assessments and recommendations.

This systematic selection process was critical in identifying a qualified medical expert who could provide high-quality, evidence-based evaluations and contribute to the robustness of the study outcomes.

Figure 10 illustrates a streamlined Expert Validation Process for Clinical Model delineated into six sequential phases: Preliminary Briefing, Dataset Review, Case-Based Evaluation, Model Explainability Tools, Performance Metrics Discussion, Feedback Session, and Iterative Consultation.

The expert validation workflow begins with the Preliminary Briefing, which summarizes the clinical model's objectives, scope, and intended applications. This phase

ensures that stakeholders have a shared understanding of key concepts and expectations. The second phase, Dataset Review, involves a detailed evaluation of the data used for training and validating the model. Experts examine critical aspects such as data quality, representativeness, and potential biases to assess the reliability of the dataset.

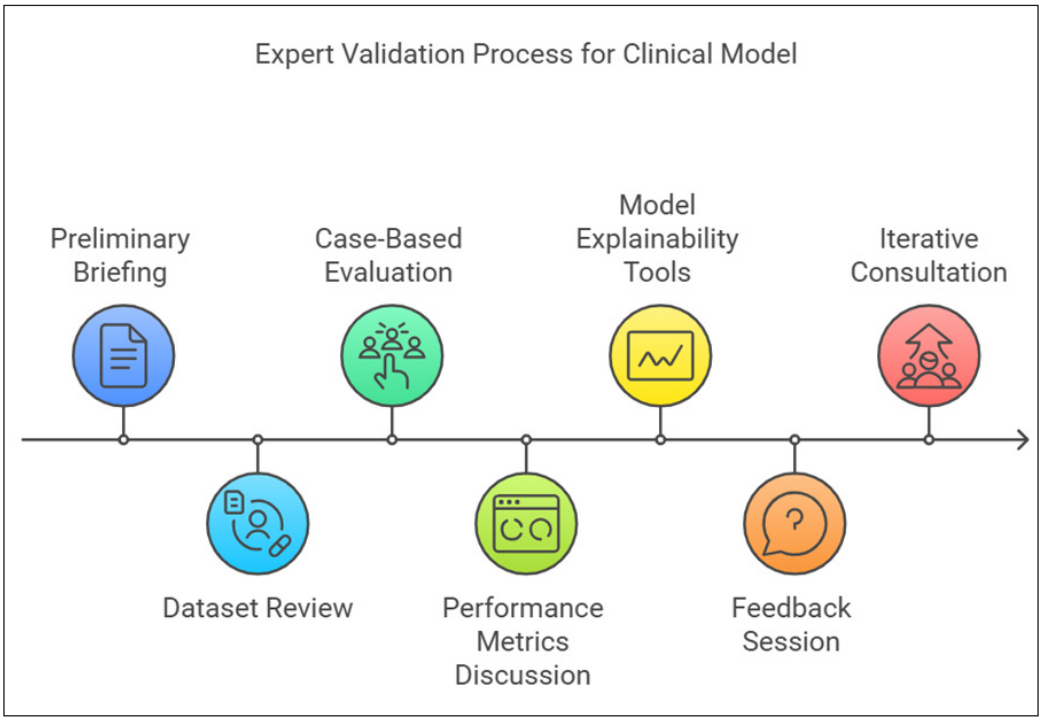

Figure 10. Expert evaluation process

The workflow then progresses to Case-Based Evaluation, where domain-specific test cases are analyzed to assess the model's performance in practical, real-world scenarios. This phase relies on clinical expertise to explore the complex situation that generic metrics may overlook. Next, the Model Explainability Tools phase focuses on interpreting the model's decision-making processes through visualization or analytical techniques. This step enhances transparency and allows experts to confirm that the model's logic aligns with clinical reasoning.

In the Performance Metrics Discussion, the expert reviews quantitative performance indicators tailored to the clinical context, such as accuracy. This phase ensures that the model meets the necessary performance standards for clinical implementation. Additionally, after the initial models are developed, the expert shares her evaluation in the Rating and Sharing stage. This step facilitates reflection on different approaches and quality assessments of the random forest model for breast cancer prediction.

The final stages, Feedback Session and Iterative Consultation, integrate insights from previous phases and refine the model based on expert input. This iterative approach promotes continuous improvement and alignment with clinical needs. The workflow integrates technical analysis with domain expertise to validate clinical models, ensuring they are reliable, interpretable, and suitable for their intended purpose.

RESULTS

This study aims to measure the accuracy of the RF algorithm in predicting BC and get the expert's validation on the dataset and the results produced by the developed system. The list of inputs in as shown in Table 6 illustrates the prediction. After going through all the procedures, the output for this input is "Benign Cancer," which means non-cancerous growth cells.

Table 6
System input samples

Attributes	Values
Concave Points Worst (0.0–0.291)	0.0000
Perimeter Worst (50.41–251.2)	50.410
Radius Worst (7.93–36.04)	7.930
Concave Point Mean (0.0–0.201)	0.000
Area Worst (185.2–4245.0)	185.200
Concavity Mean (0.0–0.427)	0.000
Area SE (6.802–542.2)	6.802
Perimeter Mean (43.79–188.5)	43.790

Next, the accuracy testing for this BC prediction system is calculated using confusion matrix accuracy. The formulas involved in this calculation are the accuracy formula [1], precision formula [2], recall formula [3], and F1 score formula [4]. Table 7 shows the confusion matrix output and table, and Table 8 shows the calculation for accuracy, precision, recall and F1 Score.

Table 7
Confusion matrix scores

	Predicted No	Predicted Yes
Actual No	TN = 65	FP = 4
Actual Yes	FN = 6	TP = 39

As indicated by the calculations in Table 8, the algorithm's accuracy is 91.23%, indicating that this percentage of predictions made by the model is correct. The algorithm's

precision is 90.70%, meaning that this proportion of positive predictions is accurate. Additionally, the model's recall is 86.67%, showing that the algorithm correctly identifies this percentage of actual positive samples. Finally, the algorithm's F1 score is 88.89%, reflecting a well-balanced performance between precision and recall.

Table 8
Calculation and scores for accuracy, precision, recall and F1

	Calculation	Answer	Percentage
Accuracy	(39 + 65) / (39 + 65 + 4 + 6)	0.9123	91.23%
Precision	39 / (39 + 4)	0.9070	90.70%
Recall	39 / (39 + 6)	0.8667	86.67%
F1 Score	2 * (0.9123*0.8667) / (0.9123+0.8667)	0.8889	88.89%

Additional experiments that combined RF with AdaBoost and XGBoost were conducted to assess the performance of hybrid models. The RF-AdaBoost model achieved an accuracy of 95.4%, a precision of 92.8%, and an F1-score of 91.7%. In comparison, the RF-XGBoost model demonstrated better results, with an accuracy of 96.8%, precision of 94.2%, and an F1-score of 93.5%. These findings highlight that hybrid models performed better than standalone RF by reducing false positives and improving overall robustness. The detailed comparative results are summarized in Table 9.

Table 9
Performance comparison of standalone and hybrid RF models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
RF (Standalone)	91.23	90.70	86.67	88.89
RF + AdaBoost	95.40	92.80	90.50	91.70
RF + XGBoost	96.80	94.20	92.70	93.50

Expert Evaluation

To ensure the reliability of the validation process, a medical expert who has specialized in breast cancer diagnosis for more than 20 years was selected. The expert selection was based on predefined standards, including clinical expertise in public health, the highest academic qualifications of PhD in public health, and knowledge of machine learning applications.

The first evaluation was on the dataset utilized in this study, derived from Fine Needle Aspiration (FNA) procedures. It is found to be significant and clinically relevant for breast cancer prediction. FNAs are a widely accepted diagnostic tool for assessing suspicious breast lesions, offering minimally invasive means to gather cytological data. Utilizing this dataset for machine learning model development aligns with contemporary diagnostic approaches, where predictive algorithms enhance the accuracy and efficiency of clinical

workflows. An expert reviewed the algorithm results to ensure the credibility of the RF algorithm and its real-world applicability.

While recall, precision, and F1-score indicate how well the machine learning model performs from a computational standpoint, expert review ensures these results are meaningful in actual clinical settings. A seasoned medical professional can interpret whether the model's high recall reduces missed diagnoses without overburdening the healthcare system or whether strong precision realistically minimizes unnecessary follow-ups. By aligning the model's metrics with real-world workflow constraints and patient needs, expert input helps confirm that the system excels on paper and holds tangible benefits for clinical decision-making, patient safety, and healthcare efficiency.

The expert's second evaluation assessed the prediction produced by the RF system. The expert was appointed to review the results based on the criteria and indicators in Table 10.

Table 10
Expert rating scale

Criterion	**Excellent (5/5)**	**Good (4/5)**	**Acceptable (3/5)**	**Not Acceptable (<3/5)**
Model Sensitivity (Recall)	Recall ≥ 90%	Recall ≥ 85%	Recall ≥ 80%	Recall < 80%
Model Precision	Precision ≥ 90% (Minimizes false positives, aligns with efficient resource use and patient confidence)	Precision ≥ 85%	Precision ≥ 80%	Precision < 80%
Overall Balance (F1-Score)	F1 ≥ 90%	F1 ≥ 85%	F1 ≥ 80%	F1 < 80%
Clinical Interpretability and Relevance	Key features and reasoning steps align strongly with medical knowledge; easy integration into workflows	Mostly aligns with standard clinical factors; minor gaps in transparency or complexity	Some clinical alignment may require additional explanation or data refinement	Not clinically interpretable or relies heavily on non-clinical features
Population and Generalizability	The dataset will represent typical patient populations and disease variability	Mostly representative, with minor known biases	Some representativeness concerns that may limit generalizability	Significant concerns about bias or lack of generalizability

Table 11 shows the medical expert's evaluation results. The expert's evaluation was based on standalone RF and hybrid RF performance. The expert also thoroughly reviewed all the inputs given to the system and the predictions produced.

Table 11
Expert evaluation results

Criterion	Standalone RF	RF + XGBoost	Rationale
Model Sensitivity (Recall)	Good (4/5) – (86.67%)	Excellent (5/5) – (92.7%)	Standalone RF misses about 13% of cases, which is acceptable but not ideal. The hybrid model's higher recall significantly reduces missed cancers.
Model Precision	Excellent (5/5) – (90.7%)	Excellent (5/5) – (94.2%)	Both models exhibit high precision, minimizing false positives and unnecessary interventions.
Overall Balance (F1-Score)	Good (4/5) – (88.89%)	Excellent (5/5) – (93.5%)	The hybrid model shows a well-balanced performance, indicating a strong synergy between precision and recall.
Clinical Interpretability and Relevance	Good (4/5)	Good (4/5)	Features used (e.g., FNA-related attributes) are clinically meaningful. Additional explainability details would be helpful for a higher rating.
Population and Generalizability	Acceptable (3/5)	Acceptable (3/5)	Although the dataset is relevant, more information on diversity, sample size, and representativeness is needed to confidently rate higher.

DISCUSSION

This study demonstrates the potential of Random Forest (RF)-based systems for improving breast cancer prediction. The standalone RF model performed well, achieving 91.23% accuracy and an F1-score of 88.89%. Enhancements to the RF algorithm, such as integrating boosting techniques like AdaBoost and XGBoost, further improve prediction accuracy and robustness. These hybrid approaches address the limitations of standalone RF models, such as susceptibility to overfitting and challenges with class imbalances. AdaBoost prioritizes misclassified instances to refine predictions, while XGBoost employs efficient parallel processing to handle large and complex datasets. However, these enhancements require greater computational resources and extensive hyperparameter tuning. These findings highlight the importance of combining algorithms to enhance prediction accuracy and reduce false positives. Expert validation was vital in confirming that the model's predictions were reliable and aligned with real-world clinical needs.

Incorporating expert validation has become a uniqueness of this study, introducing critical qualitative insights. Beyond statistical measures, expert input provides clinical credibility, ensuring the model's predictions align with real-world diagnostic practices. Experts also enhance contextual accuracy, interpreting patterns and assessing their clinical significance. Importantly, their feedback identifies hidden biases, such as systematic underperformance in specific subgroups, thereby promoting fairness in medical AI applications. This is aligned with the findings of a scoping review study by Moosavi et

al. (2024) that highlights the importance of executing an expert review of AI or machine learning clinical algorithms to build robust evidence of their applications.

However, there are still challenges to address. While effective, the dataset used in this study may not reflect the full diversity of breast cancer cases, which could limit its reliability for certain groups. Additionally, the advanced computing needs of hybrid models like RF + XGBoost might make it hard to use in places with fewer resources. Clinical experiments are also needed to see how well the system works in real healthcare environments.

Future work should focus on using more diverse datasets and making hybrid models more efficient and easier to use in actual settings. Expert feedback will continue to play an important role, ensuring the system is practical and fair and meets the needs of healthcare providers.

CONCLUSION

This study developed and validated a breast cancer prediction system based on the Random Forest (RF) algorithm, further enhanced with hybrid models integrating AdaBoost and XGBoost. The standalone RF achieved 91.23% accuracy, while the RF + XGBoost hybrid improved performance to 96.8% accuracy and an F1-score of 93.5%. This hybrid work highlights the system's strength in reducing false positives and enhancing diagnostic reliability. A key novelty of this work is the integration of expert validation in RF for breast cancer prediction, ensuring clinical relevance and alignment with real-world practices. Using clinically significant attributes and robust evaluation metrics underscores its potential as a practical early detection tool. However, limitations such as dataset diversity and computational demands remain. Future research should focus on diverse datasets and efficiency optimization to enhance usability in medical settings.

ACKNOWLEDGEMENTS

This work was supported by Universiti Teknologi MARA, Malaysia.

REFERENCES

Breast cancer. (2024). *World Health Organization.* https://www.who.int/news-room/fact-sheets/detail/breast-cancer

Dai, B., Chen, R.-C., Zhu, S.-Z., & Zhang, W.-W. (2018). Using Random Forest Algorithm for breast cancer diagnosis. In *International Symposium on Computer, Consumer and Control (IS3C)* (pp. 449-452). IEEE. https://doi.org/10.1109/IS3C.2018.00119

Duan, H., Zhang, Y., Qiu, H., Fu, X., Liu, C., Zang, X., Xu, A., Wu, Z., Li, X., Zhang, Q., Zhang, Z., & Cui, F. (2024). Machine learning-based prediction model for distant metastasis of breast cancer. *Computers in Biology and Medicine*, *169*(January), 107943. https://doi.org/10.1016/j.compbiomed.2024.107943

He, Z., Chen, Z., Tan, M., Elingarami, S., Liu, Y., Li, T., Deng, Y., He, N., Li, S., Fu, J., & Li, W. (2020). A review on methods for diagnosis of breast cancer cells and tissues. *Cell Proliferation*, *53*(7), 1-16. https://doi.org/10.1111/cpr.12822

Ishwaran, H., & Lu, M. (2019). Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Statistics in Medicine*, *38*(4), 558-582. https://doi.org/10.1002/sim.7803

Kinra, P. (2019). Market analysis of breast cancer. *Oncology & Cancer Case Reports*, *6*(1), 1-5. https://www.iomcworld.org/open-access/market-analysis-of-breast-cancer.pdf

Li, X., Li, X., Zhang, K., Guan, Y., Fan, M., Wu, Q., Li, Y., Holmdahl, R., Lu, S., Zhu, W., Wang, X., & Meng, L. (2024). Autoantibodies against Endophilin A2 as a novel biomarker are beneficial to early diagnosis of breast cancer. *Clinica Chimica Acta*, *560*(March), 119748. https://doi.org/10.1016/j.cca.2024.119748

Macaulay, B. O., Aribisala, B. S., Akande, S. A., Akinnuwesi, B. A., & Olabanjo, O. A. (2021). Breast cancer risk prediction in African women using Random Forest Classifier. *Cancer Treatment and Research Communications*, *28,* 100396. https://doi.org/10.1016/j.ctarc.2021.100396

Mentch, L., & Zhou, S. (2020). Randomization as regularization: A degrees of freedom explanation for random forest success. *Journal of Machine Learning Research*, *21*, 1-36.

Minnoor, M., & Baths, V. (2022). Diagnosis of Breast cancer using random forests. *Procedia Computer Science*, *218*(2022), 429-437. https://doi.org/10.1016/j.procs.2023.01.025

Mohamed, E. S., Naqishbandi, T. A., Bukhari, S. A. C., Rauf, I., Sawrikar, V., & Hussain, A. (2023). A hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms. *Healthcare Analytics*, *3*(March), 100185. https://doi.org/10.1016/j.health.2023.100185

Moosavi, A., Huang, S., Vahabi, M., Motamedivafa, B., Tian, N., Mahmood, R., Liu, P., & Sun, C. L. F. (2024). Prospective human validation of artificial intelligence interventions in cardiology: A scoping review. *JACC: Advances*, *3*(9), 101202. https://doi.org/10.1016/j.jacadv.2024.101202

Natras, R., Soja, B., & Schmidt, M. (2022). Ensemble machine learning of Random Forest, AdaBoost and XGBoost for vertical total electron content forecasting. *Remote Sensing*, *14*(15), 1-34. https://doi.org/10.3390/rs14153547

National Cancel Registry. (2019). *National Cancer Registry Report 2012-2016*. https://www.moh.gov.my/moh/resources/Penerbitan/Laporan/Umum/2012-2016 (MNCRR)/Summary_MNCR_2012-2016_-_06112020.pdf

Park, K. H., Loibl, S., Sohn, J., Park, Y. H., Jiang, Z., Tadjoedin, H., Nag, S., Saji, S., Md. Yusof, M., Villegas, E. M. B., Lim, E. H., Lu, Y. S., Ithimakin, S., Tseng, L. M., Dejthevaporn, T., Chen, T. W. W., Lee, S. C., Galvez, C., Malwinder, S., … Harbeck, N. (2024). Pan-Asian adapted ESMO clinical practice guidelines for the diagnosis, treatment and follow-up of patients with early breast cancer. *ESMO Open*, *9*(5), 102974. https://doi.org/10.1016/j.esmoop.2024.102974

Rashid, M. M., Yaseen, O. M., Saeed, R. R., & Alasaady, M. T. (2024). An improved ensemble machine learning approach for diabetes diagnosis. *Pertanika Journal of Science and Technology*, *32*(3), 1335-1350. https://doi.org/10.47836/pjst.32.3.19

Rashidi, H. H., Tran, N. K., Betts, E. V., Howell, L. P., & Green, R. (2019). Artificial intelligence and machine learning in pathology: The present landscape of supervised methods. *Academic Pathology*, *6*, 2374289519873088. https://doi.org/10.1177/2374289519873088

Shehab, M., Abualigah, L., Shambour, Q., Abu-Hashem, M. A., Shambour, M. K. Y., Alsalibi, A. I., & Gandomi, A. H. (2022). Machine learning in medical applications: A review of state-of-the-art methods. *Computers in Biology and Medicine*, *145*(November 2021), 105458. https://doi.org/10.1016/j.compbiomed.2022.105458

Sumwiza, K., Twizere, C., Rushingabigwi, G., Bakunzibake, P., & Bamurigire, P. (2023). Enhanced cardiovascular disease prediction model using random forest algorithm. *Informatics in Medicine Unlocked*, *41*(March), 101316. https://doi.org/10.1016/j.imu.2023.101316

Vazquez-Zapien, G. J., Mata-Miranda, M. M., Garibay-Gonzalez, F., & Sanchez-Brito, M. (2022). Artificial intelligence model validation before its application in clinical diagnosis assistance. *World Journal of Gastroenterology, 28*(5), 602-604. https://doi.org/10.3748/wjg.v28.i5.602

Yifan, D., Jialin, L., & Boxi, F. (2021). Forecast model of breast cancer diagnosis based on RF-AdaBoost. In *2021 IEEE 3rd International Conference on Communications, Information System and Computer Engineering (CISCE)* (pp. 716-719). IEEE. https://doi.org/10.1109/CISCE52179.2021.9445847

Zuo, T., Zeng, H., Li, H., Liu, S., Yang, L., Xia, C., Zheng, R., Ma, F., Liu, L., Wang, N., Xuan, L., & Chen, W. (2017). The influence of stage at diagnosis and molecular subtype on breast cancer patient survival: A hospital-based multi-center study. *Chinese Journal of Cancer*, *36*(1), 1-10. https://doi.org/10.1186/s40880-017-0250-3