Towards an Accurate Breast Cancer Classification Model based on Ensemble Learning

Aya Hesham[1], Nora El-Rashidy[2], Amira Rezk[3], Noha A. Hikal[4]
Information Technology Department-Faculty of Computers and Information, Mansoura University, Mansoura, 13518, Egypt[1, 4]
Machine Learning and Information Retrieval Department-Faculty of Artificial Intelligence,
Kafrelsheiksh University, Kafrelsheiksh,13518, Egypt[2]
Information Systems Department-Faculty of Computers and Information, Mansoura University, Mansoura, 13518, Egypt[3]

***Abstract*—Breast cancer (BC) is considered the most common cancer among women and the major reason for the increased death rate. This condition begins in breast cells and may spread to the rest of the body tissues. The early detection and prediction of BC can help in saving a patient's life. In the last decades, machine learning (ML) has played a significant role in the development of models that can be used to detect and predict various diseases at an early stage, which can greatly increase the survival rate of patients. The importance of ML Classification is attributed to its capability to learn from previous datasets, detects patterns that are difficult to comprehend in massive datasets, predicts a categorical variable within a predefined example and provide accurate results within a short amount of time. Feature selection (FS) method was used to reduce the data dimensionality and choose the optimal feature set. In this paper, we proposed a stacking ensemble model that can differentiate between malignant and benign BC cells. A total of 25 different experiments have been conducted using several classifiers, including logistic regression (LR), decision tree (DT), linear discriminant analysis (LDA), K-nearest neighbor (KNN), naive Bayes (NB), and support vector machine (SVM). In addition to several ensembles, the classifiers included random forest (RF), bagging, AdaBoost, voting, and stacking. The results indicate that our ensemble model outperformed other state-of-the-art models in terms of accuracy (98.6%), precision (89.7%), recall, and F1 score (93.33%). The result shows that the ensemble methods with FS have a high improvement of classification accuracy rather than a single method in detecting BC accurately.**

Keywords*—Breast cancer; feature selection; classification; machine learning

I. Introduction

Breast cancer (BC) is the major cause of cancer death among women worldwide. Most recent reports showed a significant increase in patients with BC [1]. Early detection of BC is crucial to increasing the survival rates, lowering the risks of treatments, and decreasing mortality. Cancers are uncontrolled cell divisions that can infect other tissues. BC usually starts in ducts that lead milk to the nipple. Symptoms of BC include changes in the size of one or both breasts, dimpling of breast skin, and a lump in one of the armpits. Breast tumors are classified into two types: benign and malignant. Benign tumors are non-cancerous because they are not diffused into the tissue of the breast surrounding the duct and do not cause harm on one's life. Malignant tumors are cancerous, and their cells spread to the surrounding breast tissue through the duct lining, posing a hazard to one's life.

The traditional approach for cancer diagnosis mainly depends on the experience of medical experts and their visual inspections. However, this type of diagnosis consumes long hours and may be susceptible to human errors[2]. Therefore, an automatic diagnosis system must be developed for the early detection of cancer. In this regard, several studies used machine learning (ML) techniques to develop classification and prediction models for medical diagnosis. ML is a type of artificial intelligence that allows a machine to learn by supplying a collection of facts and gaining knowledge through experience rather than through extensive programming [3].

Classification and prediction are the main techniques used to analyze data in ML. Classification refers to the predictive process of class objects whose label is unknown based on the training samples and predicts a categorical variable within a predefined example. Prediction is the process of determining numerical data that are missing or unavailable for a new observation based on previous information.

Different algorithms for classification, such as (KNN), (SVM), artificial neural networks (ANNs), and (RF) have been applied in several studies to classify BC.

Harikumar et al. [4] used DT and KNN algorithms in the classification of BC. Principal component analysis (PCA) technique was used to select the optimal feature set.

Hiba et al. [5]presented a comparison between the performance of four classifiers: SVM, NB, C4.5, and KNN. These studies mainly aimed to determine whether the classification of data is proper in regard to the efficiency and effectiveness of each algorithm.

Marie et al. [6] reviewed ML models that may predict BC in women and compared their performances. A total of 116 participants were measured for various parameters in the Coimbra dataset: insulin, glucose, resistin, adiponectin, homeostasis model assessment (HOMA), leptin, and monocyte chemoattractant protein-1. They implemented classification algorithms that included logistic regression LR, KNN, SVM, DT, RF, gradient boosting method (GBM), and NB.

Tsehay [7] proposed the KNN model for BC prediction using a grid search approach to find the best hyper-parameter. Then, a comparison was made between the default and tuned hyper-parameters. The results demonstrated that the performance significantly increased when the best parameter

or K value was utilized to train the KNN using the Wisconsin Breast Cancer dataset.

Salehi et al. [8] proposed an ensemble machine for survivability prediction. The main goal was to lower the feature set's size by deleting the more complex features. This machine depended on three basic learners, namely, multi-layer perceptron (MLP), SVM, and DT, based on 25 simple features. The ensemble multi-stage machine was based on a performance investigation using a 10-fold cross-validation (CV) method.

In this paper, we utilized a stacking ensemble model to detect BC. Ensemble learning refers to the training of a set of models instead of a single model and the combination of results using weighted or unweighted techniques [9]. The aim of combining several models is to achieve an improved predictive performance than any single-component model. The solidity of the ensemble limits the dispersion of single-model prediction. Thus, this type of learning contributes to reduction of variance and the generalization error of these models.

The following problems were considered in this article:

Various studies focused on the use of single ML algorithms in BC classification. These studies did not concentrate on the difference in the algorithms' performance in terms of the various data categories in the dataset (i.e., text and numeric). Several algorithms, such as SVM, perform better with text data than numeric data. Other algorithms, such as DT, perform better with text data. Thus, we suggested the ensemble model to aggregate the benefits of different algorithms.

The main contributions of this dataset are summarized in the following points:

We proposed medically intuitive and accurate ML models for BC prediction based on a medical text dataset, and they can be detailed in the following points:

- Survey of the most recent studies in BC classification and prediction;
- We explore the shortcoming of the previous literature and draw the guidelines for future improvements that could overcome these challenges;
- Proposal of a framework for BC classification based on textural features relating to tumor description;
- Framework model is created to solve a binary classification (0 = benign and 1 = malignant);
- Framework included the study of the effect of using FS and ensemble ML models on improving the quality and efficient for Bc classification;
- Comprehensive analysis of popular ML models (single models), such as NB, KNN, linear discriminant analysis (LDA), LR, DT, and support vector classifier (SVC), and ensemble classifiers, such as RF, bagging, AdaBoost, stacking, and voting classifiers;
- Use of feature selection (FS) to reduce data dimensionality and select the optimal feature set;
- Use of various metrics, including accuracy (Acc), specificity (Sp), precision (P), recall (R), F1 score (F1), and area under the receiver operating characteristic (ROC) (AUC) curve, to evaluate our model performance;
- Investigation of the capability of ensemble learning to perform binary classification tasks after the application of data to pipeline preprocessing steps;
- The results showing that the best Acc was achieved by the ensemble model with FS of 98.6%), which can adapt to several types of classification of medical data.

The rest of the paper is organized as follows. Section II discusses the literature review. Section III elaborates the methods and materials. Section IV explains the proposed framework. Section V presents the results. Section VI presents the discussion. Comparing with other works presented in Section VII. Finally, Section VIII concludes the paper.

II. Related Work

Numerous researchers are interested in BC prediction using ML. The related previous works can be divided into two groups: those that used single ML and those that applied ensemble learning. In the following, the two groups will be discussed.

A. The First Group Included Studies that Utilized Single ML

Abdulrahman et al. [10] compared six ML techniques. Their experiment concluded that SVM and RF were the most effective ML classifiers, with an Acc of 97.3% for BC prediction.

R. Karthiga1 et al. [11] proposed a novel strategy of BC diagnosis, which was collected from a visual laboratory, using image analysis and ML techniques. First, they applied image quality enhancement and a region of interest (ROI) selection using the morphological transform process. Then, hot regions, including armpits and neck, were segmented. The image's structural flaws were removed using morphological procedures. The statistical, intensity, and geometric features were extracted from the images of enhanced ROI. Curvelet transform was implemented using the wrapping process. The matrix of gray-level co-occurrence was used to extract textural features from the curvelet coefficients. Various ML techniques were investigated, and the cubic SVM showed the best Acc (93.3%) instead of the limited Acc for SVM.

Shiny et al. [12] proposed a new discriminate ratio for SVM. They compared three of the most common ML algorithms. RF, SVM, and NB have been applied in the original WDBC. The results have shown that SVM exhibited the best performance and an Acc of 97.2%.

Jiande et al. [1] also used the SVM algorithm in addition to KNN, NB, and DT. They utilized gene expression data to classify triple- and non-triple negative BC patients using a ML approach. The implementation of different algorithms

indicated that SVM had the highest performance with an Acc of 90%.

Mücahid et al. [13] suggested a model for the detection of BC using ANN and NB based on age, biomarkers, glucose, body mass index, HOMA, resistin, leptin, adiponectin, and insulin. According to the results, the markers can be used to obtain desirable findings from the classification algorithms; BC was classified with high accuracies of 86.95% and 83.54% with the use of ANN and NB algorithms, respectively.

Milon et al. [14] increased the Acc of the ANN classifier and compared five ML techniques, namely, SVM, KNN, RF, ANN, and LR, for BC prediction. The Acc, sensitivity, Sp, P, negative predictive value, false negative (FN) rate, false positive (FP) rate, F1, and Matthews correlation coefficient were all used to evaluate the study's performance. The best Acc achieved was by the ANN (98.57%).

Akizur et al.[15] introduced FS as a promising solution for maximizing Acc. FS was focused on to reduce the dimensionality of data and obtain a high level of BC classification Accs The proposed system was based on two stages. First, FS techniques selected 22 features from the WDBC dataset. For BC classification, the ANN with a 15-neuron (15-neuron ANN) classifier obtained an Acc of 96.4%, 99.9% sensitivity, 98.4% P, 1.6% FP rate, and 0.42 s processing time. The proposed 15-neuron ANN classifier achieved a 98.8% classification Acc in the initial experiment, which was performed without FS. The Acc of classification was improved with a median of 0.6% after the application of FS.

Sapiah et al. [16] also focused on improving the Acc. ML algorithms were used along with particle swarm optimization (PSO) for FS. The results indicated that using PSO achieved better performance than their counterparts in terms of Acc. With the use of PSO, NB, KNN, and reduced error pruning (REP) tree obtained accuracies of 81.3%, 66.3%, and 80%, respectively. Without PSO, the accuracies reached 70%, 75.5%, and 76.3%. Thus, the use of PSO increased the Acc of ML algorithms.

B. The Second Group Included Studies that Concerned Ensemble ML

Amal et al. [17] proposed a BC detection method that relies on a bagging ensemble classifier. They used One Rule (OneR) to extract the beneficial features of BC. The results of this method showed that the bagging classifier achieved the best Acc of 92.25% over other single ML classifiers, such as KNN, MLP, and SVM.

Royida et al. [18] proposed a new strategy for enhancing the Acc of BC dataset classification. Hoeffding trees were used for normal classification, whereas NB was applied to reduce data dimensionality. The results of the proposed model have shown an increase in the Acc to 95.9943%.

Meerja et al. [19] proposed an ensemble model for data BC classification using Bayesian network and radial basis (RB) function, which provided a good classification Acc. Various measurements were used to evaluate the model's performance. The results have shown that the proposed ensemble method recorded an Acc of 97.42%.

Tina et al. [20] proposed two methods: stacking -voting model and stacking ensemble models with Bayesian methods as base classifiers. In stacking, the model was used with three Bayesian methods, such as NB, hidden NB, and Bayesian network. Six models were created, and their performances were compared. In voting and stacking models, the classifiers were combined by the voting classifier, and three Bayesian methods were used individually with sequential minimal optimization (SMO) as the base classifier and a voting ensemble of REP Tree and RF as meta classifiers. The results revealed that the method using stacking alone and an ensemble of Bayesian methods exhibited the best performance, and NB combinations with LR and SMO gave the best prediction Acc of 97.8%.

Hajar et al. [21] improved the performance of ML techniques to test their classification performances in large BC databases, such as the Surveillance, Epidemiology, and End Results (SEER) database. The KNN, NB, LR, and MLP techniques exhibited low Acc, contrary to the RF and DT, which proved their high performance. Ensemble approaches, such as voting, stacking, bagging, and boosting, were used to increase the performance of single algorithms. Voting techniques provided the greatest improvement and Acc (99.99%) with NB, DT, and RF. For more information about the use of ML models in BC classification, readers are guided to the following surveys [22],[23].

Previous works presented several shortcomings. (1) Several researchers used an insufficient number of features, (2) most of the researchers disregarded FS as the primary algorithm for BC diagnosis, and (3) others focused on the same algorithm in the classification although most of the different types of data require the use of various types of algorithms to be classified. Most algorithms work best on text data and others on numeric data. Thus, combining two types of data using the same algorithm affects the classification performance. Therefore, we suggested the use of ensemble learning and FS to overcome these limitations and improve the Acc and performance for BC classification**.** Table I shows the characteristic of the reviewed studies.

TABLE I. MAIN CHARACTERISTICS OF THE REVIEWED STUDY

Ref.	Year	Method	Accuracy	Application	Dataset	Data used	Attributes
[10]	2022	SVM	97.3%	Prediction	569 instances from WDBC	Text	10 attributes
		NB	95.7%				
		ANN	95.7%				
		RF	97.3%				

Ref.	Year	Method	Accuracy	Application	Dataset	Data used	Attributes
		KNN	93.8%				
		DT	93%				
[1]	2021	KNN	87%	Classification	1222 samples (110 triple negative 992 non-triple negative; RNA-Seq Data [CrossRef] [PubMed]	Text	30 genes
		NGB	85%				
		DT	87%				
		SVM	90%				
[17]	2021	NLP + ML: J48	88.73%	Prediction	142 case patients from King Abdullah University Hospital in Jordan. Luminal A, 15%; luminal B, 46%; human epidermal growth factor receptor 2 positive, 24%; triple-negative BC, 15%	Text	13 attributes
		NLP + ML: NB	89.44%				
		NLP + ML: Bagging	92.25%				
		NLP + ML: Logistic	85.92%				
		NLP + ML: SVM	90.14%				
		NLP + ML: KNN	89.44%				
		NLP + ML: MLP	90.85%				
		NLP + ML: PART	89.44%				
		NLP + ML: OneR	90.14%				
[21]	2021	Voting ensemble methods (NB, DT, and RF)	99.99%	Classification	699412 cases (81386 in 2016 and 618026 in 2008–2015) from SEER; the National Cancer Institute from 2008 to 2016	Text	-15 attributes from 2008 to 2015 -11 attributes for 2016
[11]	2021	Curvelet transform - cubic SVM	93.3%	Classification	60 (30 normal, 30 abnormal) from visual laboratory, Brazil, Fluminense Federal University	Image	16 attributes
[19]	2021	Ensemble classifier (Bayesian network and RB function)	97%	Classification	699 instances, WBCD https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic).	Text	10 attributes
[15]	2020	(15-neuron network) with FS	96.4%	Classification	569 samples (357 normal, 212 cancer) from University of California Irvine (UCI) Learning Repository dataset for diagnosis of breast cancer in Wisconsin http:/archive.ics.uci.edu/ml/index.php	Text	32 attributes
[14]	2020	SVM	97.14%	Prediction	699 instances: 458 benign, 241 malignant (UCI) https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cance rwisconsin/breast-cancer-wisconsin.data.	Text	9 attributes
		K NN	97.14%				
		RF	95.71%				
		ANN	98.57%				
		LR	95.71%				
[20]	2020	Stacking ensemble methods (NB with LR and SMO)	97.8%	Prediction	699 instances (16 missing values, 444 benign, and 239 malignant cancers); Wisconsin Breast Cancer Database (WBC) available from UCI between 1989 and 1991	Text	10 attributes
[12]	2019	SVM	97.2%	Prediction	458 benign and 241 malignant (UCI ML Repository)	Text	10 attributes
		RF	96.6%				
		NB	97%				
[13]	2019	ANN	86.95%	Classification	64 patients with breast cancer and 52 healthy controls (UCI ML Repository) http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra	Text	10 attributes
		NB	83.54%				
[18]	2019	Ensemble	95.9943%	Classification	699 instances (444 benign, 239 malignant, and 16	Text	9

Ref.	Year	Method	Accuracy		Application	Dataset	Data used	Attributes
		hoffeding tree and NB				missing) WBC		attributes
[16]	2018		With out (PSO)	With (PSO)	Prediction	198 instances (UCI from the Wisconsin Prognostic Breast Cancer) https://archive.ics.uci.edu/ml/datasets/Breast±Cancer± Wisconsin± (Prognostic	Text	34 attributes
		KNN	66.3 %	75.5 %				
		NB	70%	81.3 %				
		REP	76.3 %	80%				

III. Methods and Materials

A. Data Set

In this section, we used the Wisconsin Diagnostic Breast Cancer (WDBC) obtained from UCI ML Repository [24]. This data set was used to detect malignant and benign tumors. All features in the WDBC were extracted from fine-needle aspirate images, which describe the features of the nucleus. The WDBC included the features of 569 patients (357 benign and 212 malignant cases) and 32 attributes in Wisconsin hospitals. Each attribute represents a measurement result. The first and second attributes refer to the identifier number and patient diagnosis status, respectively. The remaining attributes correspond to standard error, mean, and the least of red nucleus features. Table II provides details of the dataset attributes and their description.

TABLE II. Dataset Attributes and their Description

NO	Attribute	Description	Domain value
1	**Id**	Alphanumeric code is used in a health record system to uniquely identify a patient.	8670 to 912 m
2	**Diagnosis**	Process of identifying the condition of the disease	0 or 1
3	**Radius_mean**	Average of distances between the center and points around the perimeter	6.99 to 28.2
4	**Texture_mean**	Gray-scale value standard deviation	9.73 to 39.4
5	**Perimeter_mean**	Average tumor size	44 to 189
6	**area_mean**	Mean area inside the boundary of core tumor	144.1 to 2500
7	**smoothness_mean**	Local variation's mean in the radius lengths for the measurement of the smoothness of breast cells	0.06 to 0.17
8	**compactness_mean**	Calculated using the equation (perimeter^2 / area - 1.0)	0.03 to 0.35
9	**concavity_mean**	Mean of the degree to which the cell contour is concave	0 to 0.44
10	**Concave points _mean**	Mean of the quantity of concave contour parts	0 to 0.4
11	**symmetry _mean**	Mean of similar area of tumor parts that match	0.11 to 0.3
12	**Fractal_ dimension _mean**	Mean for "coastline approximation"-1	0.05 to 0.2
13	**radius_se**	Error of standard for the average length from the center to the perimeter points	0.11 to 2.88
14	**texture_se**	Error of standard for the grayscale values' standard deviation	0.38 to 4.88
15	**perimeter_se**	Error of standard for the tumor perimeter	0.77 to 22
16	**area_se**	Error of standard for the mean area inside the boundary of the core tumor	6.9 to 543
17	**smoothness_se**	Error of standard for variations in radius lengths on a local scale	0 to 0.03
18	**compactness_se**	Error of standard for perimeter^2/area-1.0	0 to 0.15
19	**concavity_se**	Error of standard for the degree to which the cell contour is concave	0 to 0.5
20	**concave points_se**	Error of standard for the quantity of concave contour parts	0 to 0.06
21	**symmetry_se**	Error of standard for the mean of similar area of matching tumor parts	0.01 to 0.08
22	**fractal_ dimension_se**	Error of standard for "coastline approximation"	0 to 0.04
23	**radius_worst**	Biggest or worst mean value for the average distance between the center and points on the perimeter	7.94 to 36
24	**texture_worst**	Biggest or worst mean value for gray-scale value standard deviation	12.1 to 49.6
25	**perimeter_worst**	Biggest or worst mean value size of the core tumor	50.6 to 251
26	**area_worst**	Biggest or worst mean value for the mean area inside the boundary of the core tumor	185 to 4255
27	**smoothness_worst**	Biggest or worst mean value for local variations of radius lengths	0.08 to 0.23
28	**compactness_worst**	Biggest or worst mean value for (perimeter^2 / area - 1.0)	0.03 to 1.07
29	**concavity_worst**	Biggest or worst mean value for the degree to which the cell contour is concave	0 to 1.27
30	**concave points_worst**	Biggest or worst mean value for the quantity of concave contour parts	0 to 0.30
31	**symmetry_worst**	Biggest or worst mean value for a similar area of matching tumor parts	0.17 to 0.77
32	**fractal_dimension_ worst**	Biggest or worst mean value for "coastline approximation"-1	0.07 to 0.22

B. Machine Learning

1) Single machine learning: In this paper, we used popular classification models to test the capability of data features in the classification of BC; these models included SVM [25]–[28], LDA [29], [30], [31] KNN [32]–[35], LR [35]–[38], DT[39], [40], and NB. The selection of classifiers was based on two main reasons: (1) popularity in medical domains (2) and their diversity in terms of classification structure. Table III shows the selected classifiers and their types.

TABLE III. CLASSIFICATION MODELS

Classifier Type	Classifier Name	Label
Tree based	Decision Tree	DT
Linear-based model	Logistic regression	LR
	Support Vector classifier	SVC
Nonparametric-based model	K-Nearest Neighbor	KNN
Statistical based model	Linear discriminant analysis	LDA
	Naïve Bayes	NB

2) Ensemble machine learning: Ensemble classifiers integrate the decision of a set of base classifiers (single classifier) using either weighted or unweighted techniques to obtain a model that outperforms all base classifiers. Table IV describes the ensemble classifiers. Ensemble learning is an initiative technique that simulates the human nature of humans and considers several perspectives before making the final decision [31].

In traditional learning, only a single classifier is used to make the decision and solve a problem, whereas in ensemble learning, several classifiers are used to solve a problem [41]. The two main types of ensemble classifiers include (1) homogeneous classifiers, which use the same classifier such as RF, and (2) heterogeneous classifiers, which use a set of diverse classifiers such as LDA and DT. The sequence of running base classifiers and the decision combination are significant in building ensemble classifiers. Three major kinds of learning and aggregation exist.

First, bagging considers a homogenous classifier in which learning occurs independently (parallel approach). Then, the results are combined using the averaging process as in the following equation $E= (\Sigma e_i)/n$, where E refers to the final classifier, and e indicates the base classifier (i.e., RF) [42].

Second, boosting is a homogenous classifier that works sequentially (self-learning technique). Boosting works by assigning weights for every classifier (equal weights in the first stage). Then, the weights are updated every iteration based on the classifier performance. The final model is averaged using a weighted average with the following equation $e= ((\Sigma e_i w_i)/*\Sigma w_i))/n$, where e is the base model, and w is the weight (i.e., AdaBoost, light GBM, GBM, and XGBM). Fig. 1 shows the bagging, boosting, and subspace models [43], [44].

TABLE IV. DESCRIPTION OF ENSEMBLE CLASSIFIERS

Classifier name	Description	Classifiers	Aggregating result technique
RF	Ensemble of DT mode	DT	Majority voting
Bagging	Fits base classifiers on each random subset and aggregates the results (parallel)	LR	Average of five weak classifiers
Boosting	Trains data on a model and then uses the second algorithm to correct errors from the first model (sequentially)	LR	Weighted average of five weak classifier
Voting	Runs several classifiers on all dataset and then outputs class depending on the highest majority voting.	LR, SVC, KNN, and LDA	Soft voting
Stacking	Splits data into several data subsets; each subset runs on different classifiers, and the final results are aggregated.	LR, SVC, KNN, DT, LDA, and NB	Meta model (Level 2 learning)

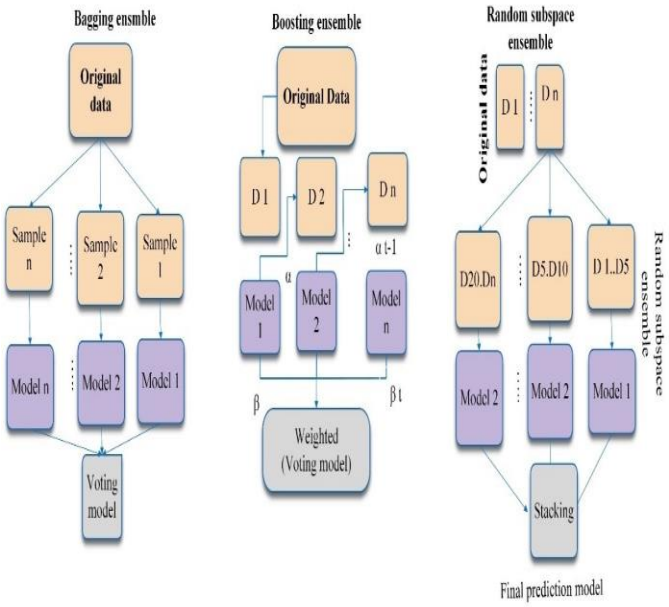


Fig. 1. Bagging, boosting, and subspace models.

Third, stacking considers heterogeneous classifiers that may run in two different ways, including the following. (1) All data are divided into feature subsets, with each subset running on all classifiers (several diverse classifiers) and a replacement for the selection of the most suitable classifier based on performance. Finally, the most accurate classifier is selected for each subset, and the final ensemble model is created. (2) For the second approach, all data sets are run in all classifiers (level 0); then, a meta learner is used to learn the best combination of predictions from all models. Fig. 2 displays the stacking ensemble model.

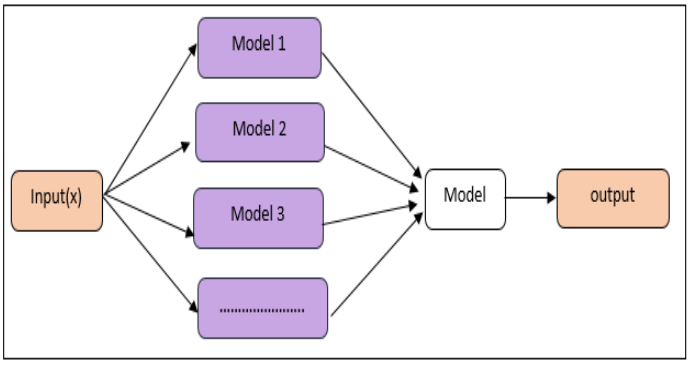


Fig. 2. Stacking ensemble learning.

IV. Proposed Framework

Based on the literature review, it was found that in the majority of the cases disregarded the role of FS algorithm for BC diagnosis. Others focused on the same or single algorithm in the classification although most of the different types of data require the use of various types of algorithms to be classified.

The steps of data pre-processing are as follows:

A. Data Pre-processing

This step aimed to represent data efficiently and improve the quality of the medical data collected. Data processing is

The objective of our proposed work is to construct an ensemble model to aggregate the benefits of different single algorithms until we can come up with an ensemble model which can give the best accuracy. We also use the method of FS that helps for improving the quality and making the process of model more efficient.

This framework is introduced a stacking ensemble classifier for the classification of BC based on tumor features (numeric features). It is a binary classification task in which 0 = benign and 1 = malignant. The model has been evaluated based on the WDBC dataset. As shown in Fig. 3, the proposed framework has several steps. The first step collects the WDBC data and implements the preprocessing steps to improve the quality of data, including (1) data normalization for data scaling to hasten the training process and (2) data balancing to increase the number of cases in the minority class. The second step uses FS to remove the weakest feature. The third step evaluates the performance of six single ML classifiers (KNN, LDA, DT, NB, SVM, and LR) with and without a FS step for BC classification task. The fourth step builds an ensemble model from a set of single classifiers with and without a FS step and compares its performance with other models that do not use ensemble learning. The tested ensemble models included RF, AdaBoost, bagging, stacking, and voting models.

We proposed a customized stacking ensemble model with high performance and Acc rate based on six optimized basic Classifiers required for ML to resolve several types of problems, such as noisy data, redundancy data, and missing data values, making it suitable for building and training ML mode:

1) Data normalization: In this step, we scaled all the data from their original range to be within the range of 0 to 1 through the following equation for normalization. This step speeds up the training processing.

$$y = \frac{(X - \min(x))}{(Max(x) - \min(x))} \quad (1)$$

2) Data balancing: As we mentioned in Section 3.1, WDBC included the features of 569 patients (357 benign and 212 malignant cases). The main problem of imbalanced data is that it affects the performance of ML algorithms. Algorithms, such as DT and KNN, will treat the minor class as not important as the major class and place more concern to the major class. Oversampling and undersampling are the main approaches used to deal with imbalanced data.

Oversampling increases the number of the training set in the minor class, whereas undersampling removes data from the minor class [45]. In this paper, given the shortage of the dataset, we decided to use an oversampling technique known as random oversampling [46]. This technique is used to randomly duplicate training examples in the minor class. The final dataset included 607 patients (357 benign and 250 malignant).

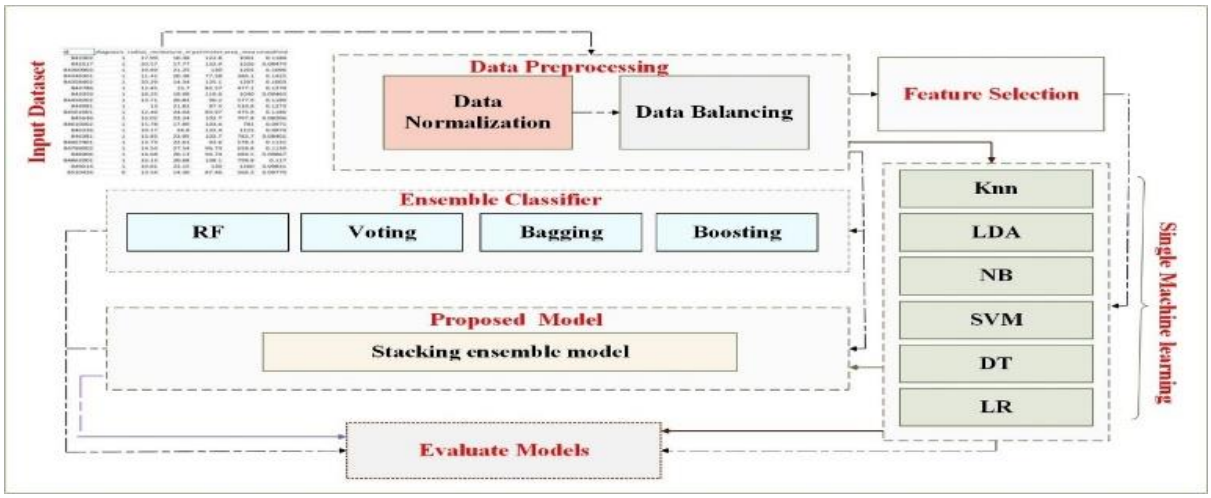


Fig. 3. Proposed framework.

B. Feature Selection

We used recursive feature elimination (REF) for FS. This method is used to reduce the input variable of a model by selecting relevant data only and cutting down the data noise [47].

The REF algorithm is based on filter and removes the weakest feature from our dataset to produce a new subset of features ranked by importance[48]. This process reduced the number of features from 32 to 24.

V. Results

A. Experimental Setup

All experiments were implemented on a PC workstation with an Intel Core i9, two tera hard disks, 32 GB RAM, and Windows 10 64 bit. All ML algorithms were implemented using Python language. Single and ensemble ML models were run using the Scikit-learn library.

All models were trained using 10-fold cross-validation (CV), which is a technique for validating the efficiency of a model in which unobserved testing data are used to evaluate the model performance and ensure the generalization performance of the training models.

We ensured that the training and testing datasets contained no data to prevent the algorithm from memorizing data and giving a better performance in the testing phase without ensuring its generalization capability. Fig. 4 shows the CV process.

B. Evaluation Performance

The effectiveness of ML algorithms was assessed using a set of performance measure parameters. The matrix of confusion, which included true positive (TP), False Positive (FP), true negative (TN), and FN results for actual and predicted data, was created to evaluate the parameters. In our study, we used the following parameters to evaluate terms by their formulas to evaluate our performance study.

The following equations were used to evaluate the comparison study's performance:

Accurcy (Acc): The correct ratio of samples that is classified to the total samples.

$$(\text{Acc}) = \frac{(TP+TN)}{(\text{TP+TN+FP+FN})} \quad (2)$$

Specificity (Sp): The relationship between negative examples observed and all other negative examples; shows the rate of projected existence in the presence of BC in all samples.

$$(\text{Sp}) = \frac{\text{TN}}{(TN+FP)} \quad (3)$$

Precision (P): The division of samples that were positive among all the samples that we anticipated to be positive.

$$(\text{P}) = \frac{\text{TP}}{(TP+FP)} \quad (4)$$

Recall (R): The rate of the case of positive impression divided by the total positive of cases.

$$(\text{R}) = \frac{\text{TP}}{(TP+FN)} \quad (5)$$

F1 score (F1): The mean between P and R

$$(\text{F1}) = 2 * \frac{P*R}{P+R} \quad (6)$$

Area under the ROC curve (AUC): Measures the performance and how the evaluated model distinguishes between models.

$$\text{AUC} = \frac{s_p - n_p + (n_{n+1)/2}}{n_p n_n} \quad (7)$$

C. Experimental Results

Four extensive experiments have been conducted to determine the efficiency of the proposed model using the WDBC dataset. WDBC provided 30 features for 560 patients and summarized the tumor status to determine benign or malignant tumors.

1) Single model

a) Without feature selection: In this section, we evaluated six diverse classifiers, including KNN, LDA, DT, NB, SVC, and LR. We carefully selected the most efficient classifier used frequently on the medical side. Table V shows the performance of all models represented by CV Acc, in addition to details about the testing performance in terms of P, R, F1, AUC, and SP.

From Table V, (1) SVC generated the least performance, with testing results of Acc=78.9%, AUC = 77.88%, and F1 = 69.2%, followed by LR with testing performance results of Acc = 92.1%, AUC = 90.49, and F1=87.32%. DT, KNN and NB performances improved by approximately 1%–2%, whereas LDA provided the best performance, with Acc = 98.2%, AUC = 97.97%, and F1 = 97.22%.

b) With feature selection: As shown in Table VI, in this section [31], [49], we explored the importance of using the FS technique to improve the ML model performance. In this paper, we used REF for FS. This technique works by fitting a model and then removing the weakest features. This process was repeated until the optimum feature set with 24 features was reached. FS techniques improved the Acc of the developed models by about 3%–6%.

SVC enhanced by about 6% and achieved Acc, AUC, and F1 of 89.4%, 82.71%, and 86.6, respectively. NB, KNN, and LR attained 92.9%, 91.22%, and 91.22%, respectively.

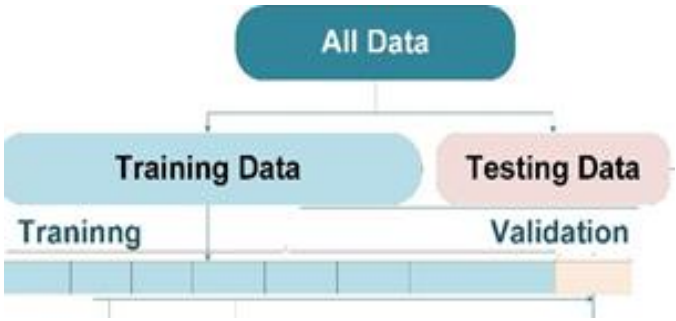


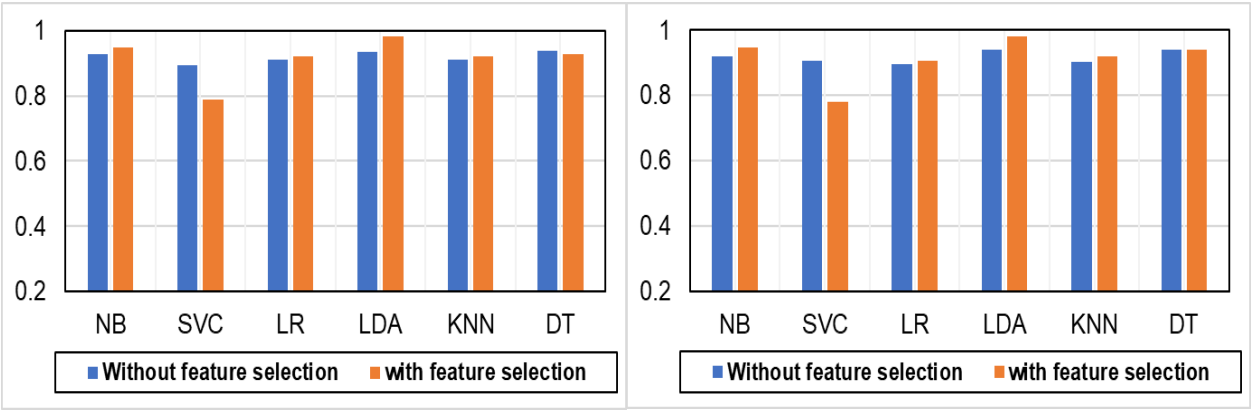
Fig. 4. Process of 10-fold CV.

TABLE V. SINGLE-CLASSIFIER RESULTS WITHOUT FEATURE SELECTION

Algorithm name	Training score %	Testing score %	Acc %	Sp %	P %	R %	F-score %	AUC %
NB	94.06	94.73	94.73	94.87	89.47	94.44	91.89	94.65
SVC	81.75	78.94	78.9	80.7	64.2	75.01	69.2	77.88
LR	91.64	92.10	92.1	92.1	88.57	86.11	87.32	90.49
LDA	96.92	98.24	98.2	98.71	97.22	97.22	97.22	97.97
KNN	94.94	92.10	92.1	92.30	84.61	91.66	87.99	91.98
DT	96.9	92.98	92.9	91.02	83.33	97.22	89.7	94.12

TABLE VI. SINGLE-CLASSIFIER RESULTS WITH FEATURE SELECTION

Algorithm name	Training score %	Testing score %	Acc %	Sp %	P %	R %	F-score %	AUC %
NB	94.505	92.98	92.9	95.8	92.3	92.3	87.8	91.8
SVC	90.54	89.47	89.4	86.3	71.61	85.11	86.6	82.71
LR	92.08	91.2	91.22	95.8	91.8	82.9	87.1	89.4
LDA	96.9	95.6	93.5	95.6	80.80	87..0	93.5	93.9
KNN	94.72	91.2	91.22	95.5	93.01	85.1	88.8	90.3
DT	98.02	93.8	93.8	94.0	91.6	93.06	92.6	93.8



(a) (b)

Fig. 5. (a) F-score of single classifiers with and without FS. (b) Acc result of single classifiers with and without FS.

The best performance was obtained by the DT, with Acc = 93.8%, F1 = 92.6%, and AUC = 93.8%. From the previous tables (Tables V and VI), we arrived at the following observations: (1) the use of FS improved the total performance with all classifiers; (2) DT worked better with less features. Fig. 5 compares F-score and Acc result of all single algorithms with and without FS.

2) Ensemble models

a) Without Feature Selection: In this section, we investigated the role of popular ensemble models, including voting, bagging, boosting, and RF, and compared them with the proposed stacking algorithm. The proposed stacking algorithm used three diverse classifiers: LDA, DT, and LR. LR was used as a meta classifier at level 2 of learning. These classifiers were selected because their diversity increases their search space and Acc accordingly. Table VII shows the CV Acc and testing performance of the used ensemble models. These results showed that (1) ensemble models improved the results by approximately 2%–4%; RF generated the least performance with the testing results of Acc = 93.8%, AUC = 92.8%, and F1=91.7%.

This finding may be attributed to the structure of RF, which selected a random subset to run in each branch. The other homogenous ensemble classifiers, including bagging and boosting classifiers, achieved better performances, with Acc = 94.3%, AUC = 96.01%, and F1 = 96.5% for the boosting classifier and Acc = 94.73%, AUC = 94%, and F1 = 93% for the bagging classifier. Regarding heterogeneous ensemble classifiers (voting and stacking), stacking provided the best performance, with Acc = 97.2%, AUC = 96.2%, and F1 = 97.72%. The performance improved by about 3%–5% over single classifiers and 1%–2% over homogenous ensemble classifiers.

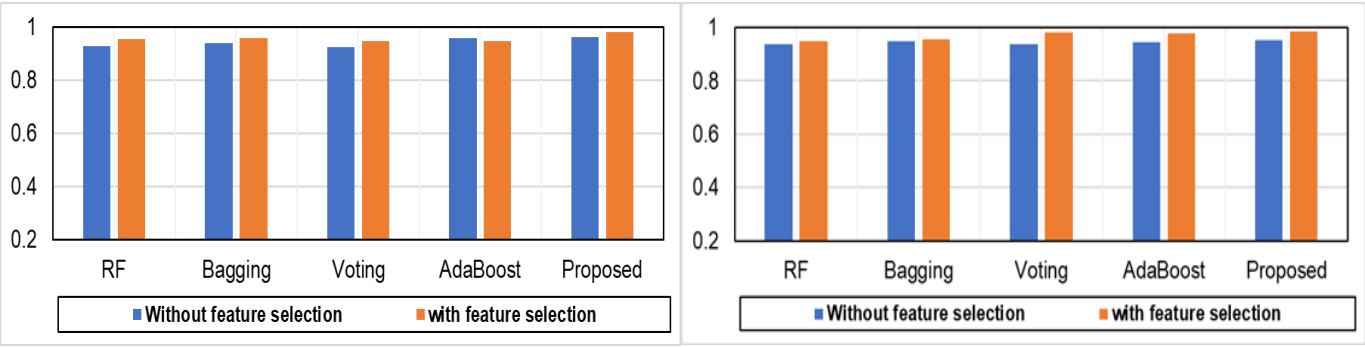
b) With Feature Selection: In this section, we explored the performance after using the FS technique with ensemble classifiers to improve the ML model performance. As shown in Table VIII, the FS techniques improved the Acc of the developed models by about 2%–3%. The RF enhanced by about 3% and achieved Acc, AUC, and F1 of 94.73%, 95.4%, and 92.1%, respectively.

TABLE VII. ENSEMBLE CLASSIFIER RESULTS WITHOUT FEATURE SELECTION

Algorithm name	Training score %	Testing score %	Acc %	Sp %	P %	R %	F-score %	AUC %
RF	96.4	93.85	93.8	97.1	95.1	88.6	91.7	92.8
Bagging	98.5	94.7	94.73	97.14	95.2	90.9	93.0	94.0
Voting	95.9	93.8	93.85	98.5	97.4	86.3	91.5	92.4
AdaBoost	95.7	97.36	94.3	98.5	97.6	95.45	96.5	96.01
Stacking	99.4	98.2	97.2	98.5	97.72	97.7	97.72	96.2

TABLE VIII. ENSEMBLE CLASSIFIER RESULTS WITH FEATURE SELECTION

Algorithm name	Training score %	Testing score %	Acc %	Sp %	P %	R %	F-score %	AUC %
RF	96.7	94.73	94.73	93.64	87.5	97.2	92.1	95.4
Bagging	99.7	95.61	95.6	94.8	89.7	97.2	93.3	96.04
Voting	96.7	98.2	98.2	98.7	97.22	97.22	97.22	94.9
AdaBoost	94.41	95.40	97.7	94.8	89.4	94.44	91.89	94.6
Stacking	97.81	95.61	98.6	94.8	89.7	97.22	96.33	98.1



(a) (b)

Fig. 6. (a). F-score of the ensemble classifiers with and without FS. (b) ACC result of the ensemble classifiers with and without FS.

The best performance was obtained from our stacking compared with other studies with F1 score = 9 6.33%, Acc = 98.6%, and AUC = 98.1%. From the previous tables (Tables VII and VIII), we arrived at the following observations; (1) the use of FS improved the total performance with all classifiers; (2) the combination of FS techniques with diverse stacking model improved the overall performance of the model. Fig. 6 compares F-score and Acc result of all ensemble algorithms with and without FS.

3) Model evaluation: We used the Nemenyi critical difference (CD) diagram to calculate the rank of each model [29], [45]; then, we evaluated and compared the performances of all the developed models (stacking, boosting, RF, LDA, and DT) with FS. The diagram in Fig. 7 shows that all algorithms were connected using horizontal lines, and they were ranked from 1 to 5, with rank 1 being alloted for the most accurate model and 5 for the least.

This ranking is equal to the average of ranking results of Friedman's test. We obtained complete information and explored the role of all features, such as radius_mean, area_mean, etc., in BC classification. This diagram proves that the stacking model achieved the best rank with a value 1.6. We calculated the feature importance with two different techniques, including information gain and RF. Table IX shows this importance numerically.

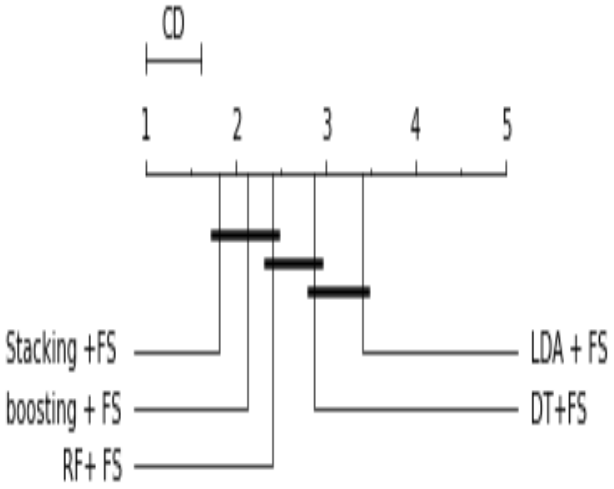


Fig. 7. CD between all classifiers with FS.

TABLE IX. RESULT OF THE FEATURE IMPORTANCE WITH INFORMATION GAIN AND RF

Feature Name	Correlation coefficient	RF
radius_mean	0.02537	0.02848
texture_mean	0.00337	0.00484
perimeter_mean	0.04407	0.05925
area_mean	0.05315	0.03405
smoothness_mean	0.00045	0.00001
compactness_mean	0.00049	0.01874
concavity_mean	0.03083	0.03261
concave points_mean	0.10136	0.14685
symmetry_mean	0.00118	0.00023
fractal_dimension_mean	0.00001	0.00108
radius_se	0.03708	0.02408
texture_se	0.00103	0.00028
perimeter_se	0.02726	0.03035
area_se	0.03464	0.03296
smoothness_se	0.00001	0.00052
compactness_se	0.01012	0.00001
concavity_se	0.00368	0.00435
concave points_se	0.00608	0.00799
symmetry_se	0.00054	0.00031
fractal_dimension_se	0.00001	0.00026
radius_worst	0.16231	0.12677
texture_worst	0.00825	0.00358
perimeter_worst	0.14621	0.15511
area_worst	0.17461	0.16965
smoothness_worst	0.00182	0.00064
compactness_worst	0.01091	0.00912
concavity_worst	0.01233	0.01436
concave points_worst	0.08065	0.09012
symmetry_worst	0.00612	0.00189
fractal_dimension_worst	0.00831	0.00121

VI. DISCUSSION

We introduced an ensemble ML classifier that can distinguish between malignant and benign BC cells. Our model was examined on the WDBC dataset. Our work faces the contribution of accelerating the training process, using data normalization for data scaling and increasing the number of instances in the minority class by data balancing. One of the main challenges in using FS is to remove the weakest feature in our model.

In order to test how well the suggested model works, we performed four experiments with and without the FS method for both single ML and ensemble ML.

In our research, we found that employing ensemble models with FS resulted in better overall performance across all classifiers. Based on six optimized basic classifiers (LR, SVC, KNN, DT, LDA, and NB), we proposed a customized stacking ensemble model that achieved the best performance and accuracy rate.

VII. COMPARISON WITH LITERATURE

In this section, we compared our proposed models with those of other literature. For fairness of comparison, we conducted this comparison only with studies that used the same dataset.

For example, [50] utilized SVM classifier in their studies to achieve an Acc of 97.47%. Similarly, in [51], a single classifier was utilized, and a classification performance of 96.5% was achieved in terms of Acc when using NB and ANN classifiers. Amrane et al. [52] used KNN and attained an Acc of 97.5%. Although these studies also used the same of classifiers, our results were superior with regard to the use of single classifiers. Such a conclusion can be attributed to the use of FS technique, which was used to extract the most important features. Bazila et al. [53] presented an ensemble known as tree augmented NB (TAN) with a boosting technique, and it achieved an Acc of 94.11%. Haifeng et al. [54] focused on an ensemble SVM learning approach named weighted area under the receiver operating characteristic curve ensemble (WAUCE). Their approach obtained an Acc of 97.68%. Table X summarizes this comparison.

Our proposed ensemble model used a heterogeneous ensemble classifier. This type of ensemble provides a high diversity between classifiers, which increases the overall classification performance. Our study achieved a superior Acc (98.6%) compared with other existing studies. On the other hand, our proposed model did not demonstrate a significant variation in the performance between training and testing. The proposed model can be potentially more stable and trustworthy and be used as an alternative to BC prediction.

TABLE X. COMPARISON OF THE ACC OF VARIOUS METHODS FOR BC CLASSIFICATION (EXPERIMENTS CARRIED OUT ON WDBC)

Ref.	YEAR	Method	Accuracy
[50]	2022	SVM	97.47%
[51]	2021	NB-ANN	96.5%
[55]	2021	KNN	97.15
[4]	2019	PCA with K-NN algorithm	95.6 %
[53]	2018	GBM -TAN	94.11%
[54]	2018	SVM -WAUCE model	97.68 %
[52]	2018	KNN	97.5%
[56]	2016	FS+ANN	97.3%
Proposed Model	2022	Stacking Ensemble ML	98.60%

VIII. CONCLUSION AND FUTURE WORK

In this paper, we proposed ML models for BC classification. We focused on the comparison of Acc when using single-classifier models, such as NB, KNN, LDA, LR, DT, and SVC, and ensemble models, such as RF, bagging, AdaBoost, stacking, and voting models, with and without using FS on WDBC. The key results from the study included the following.

In a single model, the SVC generated the least performance with testing results of Acc = 78.9%, AUC = 77.88%, and F1 = 69.2%. When using FS techniques, the total performance was improved, and SVC enhanced by about 6%,

with Acc, AUC, and F1 of 89.4%, 82.71%, and 86.6%, respectively.

For the ensemble model, we compared traditional ensemble models combining FS techniques with our stacking proposed model. The results of our experiments revealed that the use of the ensemble model and FS improved the Acc of all classifiers. The performance improved by about 3%–5% over single classifiers and 1%–2% over homogenous ensemble classifiers. The results also showed that the best Acc (98.6%) was achieved by the stacking ensemble ML with FS techniques. The proposed model can help cancer specialists to recognize BC.

In the future, we plan to expand our model to handle different types of data using different feature selection methods. We will investigate the role of machine learning models in dealing with time series data. Finally, we can apply classification with deep learning algorithms to investigate the accuracy enhancement of the classification.

REFERENCES

[1] J. Wu and C. Hicks, "Breast cancer type classification using machine learning," J. Pers. Med., vol. 11, no. 2, pp. 1–12, 2021, doi: 10.3390/jpm11020061.

[2] A. Ed-daoudy and K. Maalmi, "Breast cancer classification with reduced feature set using association rules and support vector machine," Netw. Model. Anal. Heal. Informatics Bioinforma., vol. 9, no. 1, pp. 1–10, 2020, doi: 10.1007/s13721-020-00237-8.

[3] G. Hindley, O. B. Smeland, O. Frei, and O. A. Andreassen, "Big data and the goal of personalized health interventions," in Mental Health in a Digital World, Elsevier, 2022, pp. 41–61.

[4] H. Rajaguru and S. R. Sannasi Chakravarthy, "Analysis of decision tree and k-nearest neighbor algorithm in the classification of breast cancer," Asian Pacific J. Cancer Prev., vol. 20, no. 12, pp. 3777–3781, 2019, doi: 10.31557/APJCP.2019.20.12.3777.

[5] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis," Procedia Comput. Sci., vol. 83, no. Fams, pp. 1064–1069, 2016, doi: 10.1016/j.procs.2016.04.224.

[6] Y. D. Austria, M. L. Goh, L. Sta. Maria Jr., J.-A. Lalata, J. E. Goh, and H. Vicente, "Comparison of Machine Learning Algorithms in Breast Cancer Prediction Using the Coimbra Dataset," Int. J. Simul. Syst. Sci. Technol., no. March 2021, 2019, doi: 10.5013/ijssst.a.20.s2.23.

[7] T. A. Assegie, "An optimized K-Nearest Neighbor based breast cancer detection," J. Robot. Control, vol. 2, no. 3, pp. 115–118, 2021, doi: 10.18196/jrc.2363.

[8] J. Razmara, M. Salehi, and S. Lotfi, "Development of an Ensemble Multi-stage Machine for Prediction of Breast Cancer Survivability," J. AI Data Min., vol. 0, no. 3, pp. 371–378, 2020, doi: 10.22044/JADM.2020.8406.1978.

[9] R. Damaševičius, A. Venčkauskas, J. Toldinas, and Š. Grigaliūnas, "Ensemble-based classification using neural networks and machine learning models for windows pe malware detection," Electron., vol. 10, no. 4, pp. 1–26, 2021, doi: 10.3390/electronics10040485.

[10] B. F. Abdulrahman, S. M. N. M. R, S. W. Kareem, and Z. R. Ahmed, "Comparative Evaluation of Machine Learning Algorithms in Breast Cancer," Qalaai Zanist Sci. J., vol. 7, no. 1, pp. 878–902, 2022, doi: 10.25212/lfu.qzj.7.1.34.

[11] R. Karthiga and K. Narasimhan, "Medical imaging technique using curvelet transform and machine learning for the automated diagnosis of breast cancer from thermal image," Pattern Anal. Appl., no. 0123456789, 2021, doi: 10.1007/s10044-021-00963-3.

[12] G. Engineering, "An Outline of Machine Learning Techniques for Breast Cancer Prediction," vol. 3, no. 3, pp. 125–130, 2019.

[13] M. M. Saritas and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," Int. J. Intell. Syst. Appl. Eng., vol. 7, no. 2, pp. 88–91, 2019.

[14] M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, and M. N. Kabir, "Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques," SN Comput. Sci., vol. 1, no. 5, pp. 1–14, 2020, doi: 10.1007/s42979-020-00305-w.

[15] M. A. Rahman and R. C. Muniyandi, "An enhancement in cancer classification accuracy using a two-step feature selection method based on artificial neural networks with 15 neurons," Symmetry (Basel)., vol. 12, no. 2, 2020, doi: 10.3390/sym12020271.

[16] S. B. Sakri, N. B. Abdul Rashid, and Z. Muhammad Zain, "Particle Swarm Optimization Feature Selection for Breast Cancer Recurrence Prediction," IEEE Access, vol. 6, pp. 29637–29647, 2018, doi: 10.1109/ACCESS.2018.2843443.

[17] A. Alzu'bi, H. Najadat, W. Doulat, O. Al-Shari, and L. Zhou, "Predicting the recurrence of breast cancer using machine learning algorithms," Multimed. Tools Appl., 2021, doi: 10.1007/s11042-020-10448-w.

[18] R. A. Ibrahem Alhayali, M. A. Ahmed, Y. M. Mohialden, and A. H. Ali, "Efficient method for breast cancer classification based on ensemble hoffeding tree and naïve Bayes," Indones. J. Electr. Eng. Comput. Sci., vol. 18, no. 2, pp. 1074–1080, 2020, doi: 10.11591/ijeecs.v18.i2.pp1074-1080.

[19] M. A. Jabbar, "Breast cancer data classification using ensemble machine learning," Eng. Appl. Sci. Res., vol. 48, no. 1, pp. 65–72, 2021, doi: 10.14456/easr.2021.8.

[20] T. E. Mathew, K. S. A. Kumar, and K. S. Kumar, "Breast Cancer Diagnosis using Stacking and Voting Ensemble models with Bayesian Methods as Base Classifiers," vol. IX, no. Ii, pp. 108–121, 2020.

[21] H. Saoud, A. Ghadi, and M. Ghailani, "Breast cancer diagnosis using machine learning and ensemble methods on large seer database," J. Theor. Appl. Inf. Technol., vol. 99, no. 3, pp. 594–604, 2021.

[22] G. Chugh, S. Kumar, and N. Singh, "Survey on Machine Learning and Deep Learning Applications in Breast Cancer Diagnosis," Cognit. Comput., no. 0123456789, 2021, doi: 10.1007/s12559-020-09813-6.

[23] S. J. S. Gardezi, A. Elazab, B. Lei, and T. Wang, "Breast cancer detection and diagnosis using mammographic data: Systematic review," J. Med. Internet Res., vol. 21, no. 7, pp. 1–22, 2019, doi: 10.2196/14464.

[24] "UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set." https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29 (accessed Sep. 14, 2022).

[25] Y. Ding, Y. Wang, and D. Zhou, "Mortality prediction for ICU patients combining just-in-time learning and extreme learning machine," Neurocomputing, vol. 281, pp. 12–19, 2018, doi: 10.1016/j.neucom.2017.10.044.

[26] H. Yao et al., "Severity Detection for the Coronavirus Disease 2019 (COVID-19) Patients Using a Machine Learning Model Based on the Blood and Urine Tests," Front. Cell Dev. Biol., vol. 8, no. July, pp. 1–10, 2020, doi: 10.3389/fcell.2020.00683.

[27] A. P. James and B. V. Dasarathy, "A review of feature and data fusion with medical images," Multisens. Data Fusion From Algorithms Archit. Des. to Appl., pp. 491–507, 2017, doi: 10.1201/b18851.

[28] H. Chiroma et al., "Neural networks optimization through genetic algorithm searches: A review," Appl. Math. Inf. Sci., vol. 11, no. 6, pp. 1543–1564, 2017, doi: 10.18576/amis/110602.

[29] M. Karimi Moridani and Y. Haghighi Bardineh, "Presenting an efficient approach based on novel mapping for mortality prediction in intensive care unit cardiovascular patients," MethodsX, vol. 5, no. 29, pp. 1291–1298, 2018, doi: 10.1016/j.mex.2018.10.008.

[30] G. Alfian, M. Syafrudin, M. F. Ijaz, M. A. Syaekhoni, N. L. Fitriyani, and J. Rhee, "A personalized healthcare monitoring system for diabetic patients by utilizing BLE-based sensors and real-time data processing," Sensors (Switzerland), vol. 18, no. 7, 2018, doi: 10.3390/s18072183.

[31] N. El-rashidy, S. El-sappagh, and H. M. E. Abuhmed, Tamer, Smair Abdelrazek, "Intensive Care Unit Mortality Prediction : An Improved Patient-Specific Stacking Ensemble Model," vol. 8, 2020, doi: 10.1109/ACCESS.2020.3010556.

[32] V. Srilakshmi, K. Anuradha, and C. S. Bindu, "Optimized deep belief network and entropy-based hybrid bounding model for incremental text

categorization," Int. J. Web Inf. Syst., vol. 16, no. 3, pp. 347–368, 2020, doi: 10.1108/IJWIS-03-2020-0015.

[33] S. Sareen, S. K. Gupta, and S. K. Sood, "An intelligent and secure system for predicting and preventing Zika virus outbreak using Fog computing," Enterp. Inf. Syst., vol. 11, no. 9, pp. 1436–1456, Oct. 2017, doi: 10.1080/17517575.2016.1277558.

[34] A. K. Das, S. Mishra, and S. S. Gopalan, "Predicting CoVID-19 community mortality risk using machine learning and development of an online prognostic tool," PeerJ, vol. 8, pp. 1–12, 2020, doi: 10.7717/peerj.10083.

[35] R. Sadeghi, T. Banerjee, and W. Romine, "Early hospital mortality prediction using vital signals," Smart Heal., vol. 9–10, pp. 265–274, 2018, doi: 10.1016/j.smhl.2018.07.001.

[36] R. Sadeghi, T. Banerjee, and W. Romine, "Early hospital mortality prediction using vital signals," Smart Heal., vol. 9–10, no. March, pp. 265–274, 2018, doi: 10.1016/j.smhl.2018.07.001.

[37] N. Mordvanyuk, F. Torrent-Fontbona, and B. López, "Prediction of glucose level conditions from sequential data," Front. Artif. Intell. Appl., vol. 300, pp. 227–232, 2017, doi: 10.3233/978-1-61499-806-8-227.

[38] I. U. Khan et al., "Computational Intelligence-Based Model for Mortality Rate Prediction in COVID-19 Patients," Int. J. Environ. Res. Public Health, vol. 18, no. 12, 2021, doi: 10.3390/ijerph18126429.

[39] D.-C. Li, C.-W. Liu, and S. C. Hu, "A learning method for the class imbalance problem with medical data sets," Comput. Biol. Med., vol. 40, no. 5, pp. 509–518, 2010, doi: https://doi.org/10.1016/j.compbiomed.2010.03.005.

[40] A. Ferraz, J. H. Brito, V. Carvalho, and J. Machado, "Blood type classification using computer vision and machine learning," Neural Comput. Appl., vol. 28, no. 8, pp. 2029–2040, 2017, doi: 10.1007/s00521-015-2151-1.

[41] K. Thangairulappan and P. Rathinasamy, "Ensemble Neural Network in Classifying Handwritten Arabic Numerals," J. Intell. Learn. Syst. Appl., vol. 08, no. 01, pp. 1–8, 2016, doi: 10.4236/jilsa.2016.81001.

[42] Y. Tounsi, L. Hassouni, and H. Anoun, "An Enhanced Comparative Assessment of Ensemble Learning for Credit Scoring," vol. 8, no. 5, 2018, doi: 10.18178/ijmlc.2018.8.5.721.

[43] B. Peter, Bagging , Boosting and Ensemble Methods Bagging , Boosting and Ensemble Methods, no. January 2012. 2015.

[44] Y. Wang, Y. Chen, and C. F. J. Huang, "Applying Neural Network Ensemble Concepts for Modelling Project Success," no. January, 2015.

[45] S. Ketu and P. K. Mishra, "Empirical Analysis of Machine Learning Algorithms on Imbalance Electrocardiogram Based Arrhythmia Dataset for Heart Disease Detection," Arab. J. Sci. Eng., vol. 47, no. 2, pp. 1447–1469, 2022, doi: 10.1007/s13369-021-05972-2.

[46] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results," 2020 11th Int. Conf. Inf. Commun. Syst. ICICS 2020, no. May, pp. 243–248, 2020, doi: 10.1109/ICICS49469.2020.239556.

[47] S. Dehdar, K. Salimifard, R. Mohammadi, M. Marzban, S. Saadatmand, and M. Dianatinasab, "Predicting Breast Cancer Diagnosis Delay Using Machine Learning Techniques," SSRN Electron. J., vol. 2022, 2022, doi: 10.2139/ssrn.4088355.

[48] A. Thalor, H. Kumar Joon, G. Singh, S. Roy, and D. Gupta, "Machine learning assisted analysis of breast cancer gene expression profiles reveals novel potential prognostic biomarkers for triple-negative breast cancer," Comput. Struct. Biotechnol. J., vol. 20, pp. 1618–1631, 2022, doi: 10.1016/j.csbj.2022.03.019.

[49] S. El-Sappagh, M. Elmogy, F. Ali, T. ABUHMED, S. M. R. Islam, and K.-S. Kwak, "A Comprehensive Medical Decision–Support Framework Based on a Heterogeneous Ensemble Classifier for Diabetes Prediction," Electronics, vol. 8, no. 6, p. 635, 2019, doi: 10.3390/electronics8060635.

[50] I. Qureshi, B. Mohammad, and M. A. Habeeb, "A MACHINE LEARNING MODEL FOR BREAST CANCER DETECTION USING SUPPORT VECTOR MACHINE," J. Xi'an Shiyou Univ. Nat. Sci. Ed., vol. 65, no. 03, pp. 66–78, 2022, doi: 10.17605/OSF.IO/TSC6J.

[51] İ. ATEŞ and T. T. BİLGİN, "The Investigation of the Success of Different Machine Learning Methods in Breast Cancer Diagnosis," Konuralp Tıp Derg., vol. 13, no. 2, pp. 347–356, 2021, doi: 10.18521/ktd.912462.

[52] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensari, "Breast cancer classification using machine learning," in 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), 2018, pp. 1–4.

[53] A. Bazila Banu and P. Thirumalaikolundusubramanian, "Comparison of bayes classifiers for breast cancer classification," Asian Pacific J. Cancer Prev., vol. 19, no. 10, pp. 2917–2920, 2018, doi: 10.22034/APJCP.2018.19.10.2917.

[54] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," Eur. J. Oper. Res., vol. 267, no. 2, pp. 687–699, 2018, doi: 10.1016/j.ejor.2017.12.001.

[55] S. Varadhan, N. Jeswani, V. Sajnani, and P. Jaiswal, "Analysis of Breast Cancer Diagnosis and Prognosis Using Machine Learning Algorithms," in Lecture Notes in Electrical Engineering, 2021, vol. 700, pp. 3197–3211, doi: 10.1007/978-981-15-8221-9_298.

[56] S. Aalaei, H. Shahraki, A. Rowhanimanesh, and S. Eslami, "Feature selection using genetic algorithm for breast cancer diagnosis: Experiment on three different datasets," Iran. J. Basic Med. Sci., vol. 19, no. 5, pp. 476–482, 2016, doi: 10.22038/ijbms.2016.6931.