SN Computer Science (2020) 1:270
https://doi.org/10.1007/s42979-020-00296-8

ORIGINAL RESEARCH

Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer

Vikas Chaurasia[1] · **Saurabh Pal**[1]

Received: 29 July 2020 / Accepted: 8 August 2020 / Published online: 14 August 2020

Abstract

This article compares six machine learning (ML) algorithms: Classification and Regression Tree (CART), Support Vector Machine (SVM), Naïve Bayes (NB), K-Nearest Neighbors (KNN), Linear Regression (LR) and Multilayer Perceptron (MLP) on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset by estimating their classification test accuracy, standardized data accuracy and runtime analysis. The main objective of this study is to improve the accuracy of prediction using a new statistical method of feature selection. The data set has 32 features, which are reduced using statistical techniques (mode), and the same measurements as above are applied for comparative studies. In the reduced attribute data subset (12 features), we applied 6 integrated models AdaBoost (AB), Gradient Boosting Classifier (GBC), Random Forest (RF), Extra Tree (ET) Bagging and Extra Gradient Boost (XGB), to minimize the probability of misclassification based on any single induced model. We also apply the stacking classifier (Voting Classifier) to basic learners: Logistic Regression (LR), Decision Tree (DT), Support-vector clustering (SVC), K-Nearest Neighbors (KNN), Random Forest (RF) and Naïve Bays (NB) to find out the accuracy obtained by voting classifier (Meta level). To implement the ML algorithm, the data set is divided in the following manner: 80% is used in the training phase and 20% is used in the test phase. To adjust the classifier, manually assigned hyper-parameters are used. At different stages of classification, all ML algorithms perform best, with test accuracy exceeding 90% especially when it is applied to a data subset.

Keywords Classification · Linear regression · Machine learning · Multilayer perceptron · k-Nearest neighbors · Support vector machine · Ensemble · Stack

Introduction

The abnormal growth of human cells is widely known as a cancer that attacks healthy cells. Abnormal growth of breast cells will invade cells around the breast more quickly and spread to other parts of the body. Breast cancer occurs when a malignant tumor (mass of tissue) occurs in the breast. Two types of breast cancer are: non-cancerous or benign and cancerous or malignant [1].

In machine learning, many researchers start their work from here to discover the severity of breast cancer, that is, whether the tumor is cancerous or non-cancerous. To find answers to these questions, two things are important: what is the role of the machine, and how does the machine learning combine medical data to predict the severity of the disease. Machine learning is the way to make data decisions with minimal human intervention. It is part of AI (artificial intelligence), which can learn from data, make decisions, discover patterns and build analytical models through data analysis. Clinical or medical data is part of information related to human health, which is based on routine patient care or clinical trial plans. It includes patient electronic health records based on patient health information. AI can obtain information from health-related data, process the data, and provide clear output to end users. This process is done through machine learning [2]. The algorithm used by this technique recognizes the data pattern and gives its own logic. The main goal of the algorithm used by AI is to find

This article is part of the topical collection "Advances in Computational Approaches for Artificial Intelligence, Image Processing, IoT and Cloud Applications" guest edited by Bhanu Prakash K N and M. Shivakumar.

✉ Saurabh Pal
drsaurabhpal@yahoo.co.in

[1] Department of Computer Applications, VBS Purvanchal University, Jaunpur, India

out the relationship between prevention or treatment and the patient's prognosis [3].

In this study, the main goal is to obtain the accuracy of the data set, and the features of dataset will be reduced due to the statistical method mode. Finally, on the reduced feature data subset, we apply ensemble techniques to combine multiple models constructed from a single learning algorithm by systematically changing training data.

The rest of this article is described as: (2) Literature review, which contains previous studies, based on many basic learners and their combine techniques ensemble stacking methods to produce a single result by different researchers; (3) Explained the suggested technical details of the model, including data investigation and preprocessing, statistical technique mode and overall framework; (4) Shows propose the model on data set and data subset and verify the balance comparative analysis of method and overall structure; at the end, (5) Get conclusion and (6) Future work discussion.

Literature Review

In this part of the article, we introduce previous research related to breast cancer detection and different types of classifiers that have been used to find accuracy. Table 1 shows a summary of the literature review.

Methodology

For this study, the dataset used the "Wisconsin Breast Cancer (Diagnostic) Data Set" with 569 instances and 32 attributes. This data set was created by Dr. William H. Wolberg of the University of Wisconsin to diagnose breast cancer, i.e., (M = malignant, B = benign). The dataset is located *archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic %29*.

Data Explanation

The breast cancer clinical data set contains 569 cases (357 benign, 212 malignant) reported on November 1, 1995, the patient ID number and diagnosis (malignant/benign) of each case. The remaining attributes contain 10 real-valued features for each cell nucleus. All information about attributes is discussed in detail in Table 2. Figure 1 shows the number of patients.

Data Preprocessing

In the diagnostic data set for breast cancer, attribute "diagnosis" is replaced by B's 0 and M's 1. In the data set, when the units of measurement are different, we need to standardize the data. Standardization is the process of rescaling one or more attributes so that their mean is 0 and the standard deviation is 1. Without standardization, variables measured at different scales will not contribute equally to the analysis and may produce deviations. Clinical data collected from different organizations for different purposes may be recorded in different formats. In order for these records to have the same format, they must be standardized [20]. To obtain a standardized value (z score), of the remaining attributes of the dataset we use the following formula:

$$z = \frac{X - \mu}{\sigma},$$

where X observation, μ mean and σ standard deviation.

Molding

Figure 2 shows the flow chart of model formation. The whole process can be divided into four parts.

1. The diagnostic breast cancer data set has 31 attributes, excluding the patient's ID number. Feature extraction techniques are used to extract relevant features with high scores from the dataset.
2. A feature selection technique mode is applied, which provides only prominent features from multiple attributes that have precise meanings.
3. Now, different accuracy measures from different classifiers are applied on the reduced subset of data.
4. For comparison, all the above classifiers are applied to a data set with all features to extract performance conclusions and use a reduced data subset for comparative studies.

Feature Extraction Techniques

We often pay attention to certain features that contribute the most to predictors or outputs. The process of selecting this output variable is called a feature selection method [21]. The existence of unrelated attributes in the data set may affect the accuracy of the data set.

Before data modeling, the importance of feature extraction may be helpful. Which are: improve accuracy, reduce over fitting, and reduce training time.

Following are the feature extraction techniques used in this research paper.

- Univariate or χ^2 test

Chi^2 test is often used in hypothesis testing. The chi^2 statistic is a test used to measure the comparison between

Table 1 Literature review

Author	Year of publication	Classifiers/Ensemble methods	Area of application/Disease	Accuracy achieved
Elsayad [4]	2010	Ensemble of Bayesian classifiers(multilayer perceptron neural network)	Severity of breast masses	91.83% on training subset and 90.63% of test
Huang et al. [5]	2010	Neural Network classifier	Breast cancers classification	98.83%
Lavanya and Rani [6]	2011	Decision tree algorithm	Breast cancer detection	92.97%
Bekaddour and Chikh [7]	2012	ANFIS (Adaptative Neuro-Fuzzy Inference System)	Breast cancer diagnosis	98.25%
Al-Bahrani et al. [8]	2013	Ensemble voting scheme	Prediction model for colon cancer	90.38%, 88.01%, and 85.13%
Zheng et al. [[9]	2014	K-means and support vector machine (K-SVM)	Tumor detection	97.38%
Vikas et al. [10]	2014	Naive Bayes, Support Vector Machine-Radial Basis Function (SVM-RBF) kernel, Radial Basis Function neural networks, and Decision trees	Breast cancer	SVM-RBF 96.84%
Zhang et al. [11]	2015	Ensemble decision approach(recursive partition tree) (Four molecular subtypes: Luminal-A, Luminal-B, HER2-amplified and Triple-negative.)	Breast cancer	83.8%, 77.4%, 87.9% and 92.7%
Hazra et al. [[12]	2016	Naïve Bayes, Support Vector Machine, Ensemble classifier	Breast cancer classification	97.3978% each
Nilashi et al. [13]	2017	Expectation Maximization (EM) and classification and regression trees (CART) to generate fuzzy rules	Breast cancer	93.20%
Chaurasia et al. [14]	2018	Naive Bayes, RBF network, J48	Breast cancer prediction	97.36%, 96.77%, and 93.94%, respectively
Emami and Pakzad [15]	2018	Affinity Propagation (AP) clustering for instances reduction, Adaptive Modified Binary Firefly Algorithm (AMBFA) for selection related predictor and Vectors Machine (SVM) technique for prediction	Breast cancer diagnosis	98.606%
Kadam et al. [16]	2019	Feature ensemble learning based on Sparse Autoencoders and Softmax Regression	Breast Cancer (prediction benign & malignant)	98.60%
Saritas and Yasar [17]	2019	Artificial neural networks and Naïve Bayes classifiers	Estimation of having breast cancer	86.95% 83.54, respectively
Rahman and Muniyandi [18]	2020	15-neuron network	Diagnostic Breast Cancer	99.4%

expected and actually observed data [22]. We use χ^2 (chi^2) test for feature selection to calculate χ^2 between each feature and the target and select the desired number of features with the best χ^2 scores. The following formula is used to estimate the χ^2 value:

$$\chi^2 = \sum_{i=1}^{n} \frac{O_i - E_i}{E_i},$$

where O_i observations in class i and E_i observations in class i if there was no relationship between the feature and target.

Table 2 Attribute information [19]

(1) ID number
(2) Diagnosis
M = malignant, B = benign
(3–32) Ten real-valued features are computed for each cell nucleus:
(a) radius (mean of distances from center to points on the perimeter)
(b) texture (standard deviation of gray-scale values)
(c) perimeter
(d) area
(e) smoothness (local variation in radius lengths)
(f) compactness (perimeter^2/area − 1.0)
(g) concavity (severity of concave portions of the contour)
(h) concave points (number of concave portions of the contour)
(i) symmetry
(j) fractal dimension ("coastline approximation" − 1)

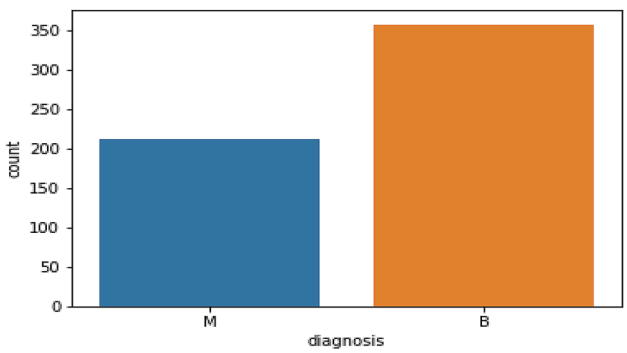


Fig. 1 Number of patients with Malignant (M) cancerous and Benign (B) non-cancerous cells

- Extra Tree (ET)

Extra Tree classifier is associate degree ensemble machine learning technique that may summarize the results of multiple unrelated decision trees collected within the forest and output its classification results. The original training samples in every decision tree derive further forests of trees. Then, at every check node, every tree is supplied with a random sample of k options from the feature-set from that every call tree should choose the simplest feature to separate the information supported some mathematical criteria generally the Gini Index. This random sample of options results in the creation of multiple uncorrelated decision trees [23].

- Recursive feature elimination (RFE)

Recursive feature elimination is largely a backward selection of predictors. This method first builds a model on the complete set of predictors to calculate the importance score for each predictor. Then Delete the area unit, rebuild the model, and calculate the importance score area unit again. First, the formula fits the model to all or any predictor variables [24]. Each predictor is a stratified victim, which is important for the model. Let S be an ordered sequence of numbers, which is a candidate for the number of predictors to keep ($S_1 > S_2, \ldots$). In each function of selection iteration, S_i high-level predictors are retained, the model is adjusted and performance is evaluated. The value of S_i with the simplest performance is determined; therefore, the high S_i predictor will match the final model.

- Random forest (RF)

Random forest may be a supervised learning algorithm rule, which is also used for regression in each category. However, it is mainly used for classification problems. As we all know, forests are composed of trees, and many trees mean many solid forests. Similarly, the rules of the random forest algorithm will create a decision tree on the knowledge samples, so as to obtain predictions from each knowledge sample, and finally select the simplest solution by voting [25]. It is a better correlation integration technique than a decision tree, because it can reduce over fitting by averaging results.

Statistical Feature Selection Technique (Mode)

Mode is derived from French word LaMode, which means 'most fashionable item'. Mode is the value which occurs largest time in a series. That is, mode in that point, where the frequencies in a distribution are maximum. At this point items tend to most heavily concentrated. There are two methods for calculating mode, i.e., mode by inspection and mode by grouping. Here we used mode by grouping method for selecting prominent features from Table 3 of all features.

Feature Selection by Grouping Method If attributes are concentrated at more than one value, we find the attributes of concentration by the method of grouping [26]. In this method we prepare a table in which the attributes are first arranged by finding different feature selection methods (χ^2, ET, RFE, RF) and their frequencies are written. The grouping table consists of the following columns.

Column 1	The given frequencies are written and highest frequency is marked.
Column 2	The frequencies in col.1 are grouped by two's and highest total is marked.
Column 3	Leaving the first frequencies of col.1 and grouping the remaining frequencies by two's and highest total is marked.

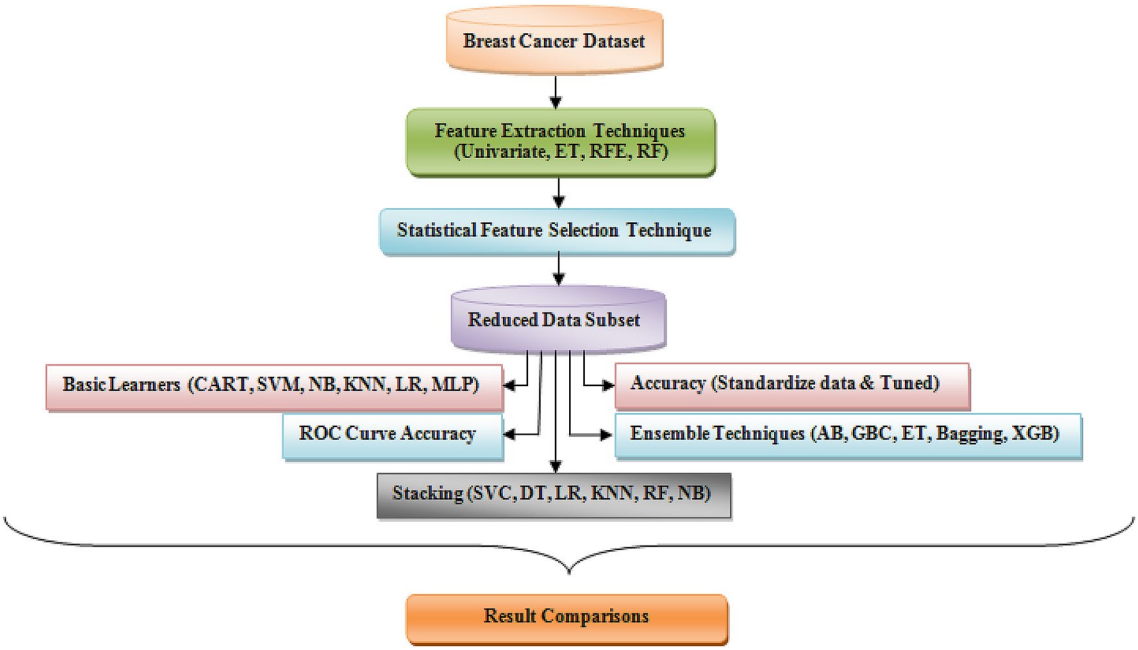


Fig. 2 Flow of proposed molding

Column 4 Starting from the first frequency of col.1, the frequencies are grouped in three's and highest total is marked.

Column 5 Starting from the second (leaving first) frequency of col.1, the frequencies are grouped in three's and highest total in marked.

Column 6 Starting from the third (leaving first two) of col.1, the frequencies are grouped in three's and highest total is marked.

After completing the grouping table, and analysis table is formed to find out attributes which are appearing the highest number of times. We tick ($\surd$) in the values used in the maximums of each column.

Accuracy of Classifiers

We have found a simplified data set from the above statistical methods. To find that the reduced data set has sufficient information related to the patient category (benign/malignant), we can apply different accuracy measures, such as basic learners, the accuracy of standardized and tuned data sets, ROC curve, ensemble methods and stacking [27].

Experiment

This section revolves around survey arrangements, methodology In addition, the results of this model on the diagnostic breast cancer dataset. The division of training and test sets follows the ratio is 80:20 and is chosen arbitrarily. The training set comes from these two datasets (reduced by feature selection and containing all features) are processed and executed by different accuracy measures for comparative study. All the analysis process was performed using Python 3.6.

Feature Extraction

Two levels of feature extraction methods are applied to the diagnostic breast cancer data set. First, we use univariate or χ^2 test, extra trees, recursive feature elimination, and random forest to select the best features from the data set [28]. By selecting 15 features from each method, we now have a total of 60 features. All these features are shown in Table 3

Mapping Features

These features need to be mapped for abbreviation, to determining the rank of each feature for analysis. After mapping the features from Table 3 in Table 4, the first column shows the attribute name corresponding to the abbreviated form in the second column. Now, the corresponding columns 3–6 show the attributes that are repeated in different feature selection techniques. Finally, column 7 represents the rank obtained by different features. After assigning features rank, these 60 features are reduced into 18 features.

Table 3 Extracted attribute

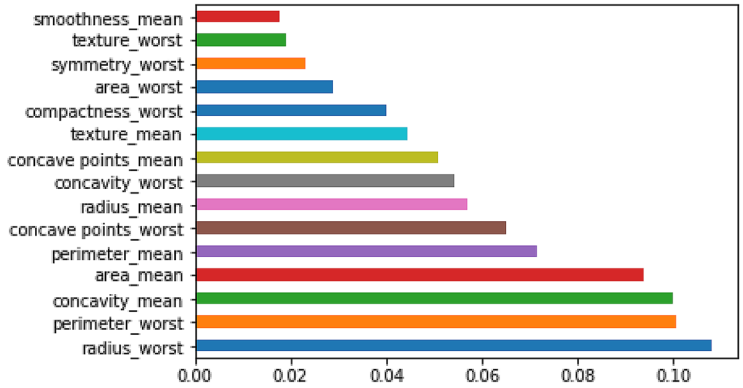
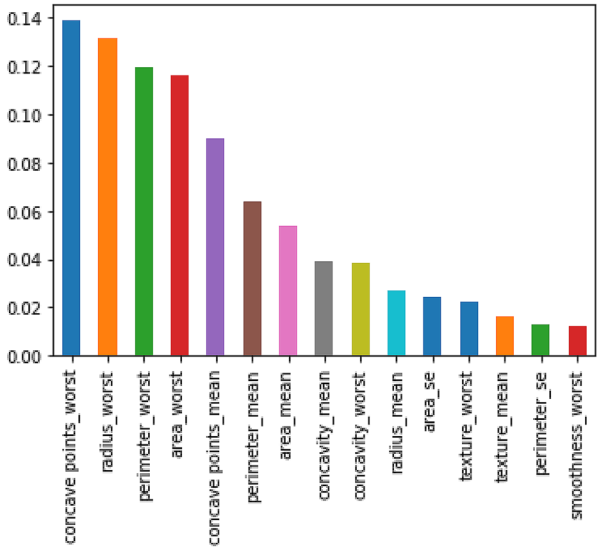
Methods	Results
Univariate or chi^2	Attributes Score 23 area_worst 112598.431564 3 area_mean 53991.655924 13 area_se 8758.504705 22 perimeter_worst 3665.035416 2 perimeter_mean 2011.102864 20 radius_worst 491.689157 0 radius_mean 266.104917 12 perimeter_se 250.571896 21 texture_worst 174.449400 1 texture_mean 93.897508 26 concavity_worst 39.516915 10 radius_se 34.675247 6 concavity_mean 19.712354 25 compactness_worst 19.314922 27 concave points_worst 13.485419
Extra Tree	
RFE	Num Features: 15 Selected Features: [True True False False False True True True True False True True True False False False False False False True True False False False True True True True False] Feature Ranking: [1 1 3 8 11 1 1 1 1 14 1 1 1 5 16 9 4 12 13 15 1 1 6 10 7 1 1 1 1 2]
RF	

Table 4 Abbreviation and rank distribution of attributes

Attribute Name	Abbreviated as	Univariate	Extra Tree	RFE	RF	Rank
area_worst	f1	f1	f1	f1	f1	4
area_mean	f2	f2	f2	f2	f2	4
area_se	f3	f3	f3	f3	f3	4
perimeter_worst	f4	f4	f4	f4	f4	4
perimeter_mean	f5	f5	f5	f5	f5	4
radius_worst	f6	f6	–	f6	f6	3
radius_mean	f7	f7	f7	f7	f7	4
perimeter_se	f8	f8	–	–	f8	2
texture_worst	f9	f9	f9	f9	f9	4
texture_mean	f10	f10	–	–	f10	2
concavity_worst	f11	f11	f11	f11	f11	4
radius_se	f12	f12	f12	f12	–	3
concavity_mean	f13	f13	f13	f13	f13	4
compactness_worst	f14	f14	f14	f14	f14	4
concave points_worst	f15	f15	f15	f15	f15	4
concave points_mean	f16	–	f16	f16	f16	3
compactness_mean	f17	–	f17	f17	–	2
smoothness_mean	f18	–	f22	–	–	1

Table 5 Grouping table of attributes

Attribute	Frequency					
	I	II	III	IV	V	VI
f1	**4**	**8**		**12**		
f2	**4**		**8**		**12**	
f3	**4**	**8**				**12**
f4	**4**		**8**	11		
f5	**4**	7			11	
f6	3		7			9
f7	**4**	6		10		
f8	2		6		8	
f9	**4**	6				10
f10	2		6	9		
f11	**4**	7			11	
f12	3		7			11
f13	**4**	**8**		**12**		
f14	**4**		**8**		11	
f15	**4**	7				9
f16	3		5	6		
f17	2	3				
f18	1					

Mode

After determining the level of each selected attribute in the above table, we will record these ranks as frequencies in column I of the next Table 5. The subsequent columns II, III, IV, V, and VI are frequency sum, as described in the "Statistical Feature Selection Technique" section.

The analysis in Table 6 is formed to find the attribute with the highest number of occurrences. We tick (√) the value used in the maximum value of each column.

In Table 7, at the end, we obtain the following 11 attributes from Analysis, Table 6. Now, these 11 (+ 1 target) attributes will be used for further analysis to conduct a comparative study with all attributes to find accuracy indicators.

Table 6 Analysis table

Column	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15	f16	f17	f18
I	√	√	√	√	√		√		√		√		√	√	√			
II	√	√	√	√									√	√				
III		√	√	√	√									√	√			
IV	√	√	√										√	√	√			
V		√	√	√														
VI			√	√	√													
No. of occurrences	3	5	6	5	3	–	1	–	1	–	1	–	3	4	3	–	–	–

Table 7 Abbreviated table of attributes

f1	Aea_worst
f2	Area_mean
f3	Area_se
f4	Perimeter_worst
f5	Perimeter_mean
f7	Radius_mean
f9	Texture_worst
f11	Concavity_worst
f13	Concavity_mean
f14	Compactness_worst
f15	Concave points_worst

Accuracy Metrics

To compare the data set with 30 attributes and 1 target attribute and the data subset with 11 attributes and 1 target attribute. We estimated the accuracy of the basic classifier, standardized data and adjusted data, ROC curve, Ensemble technique and stacking.

Table 7 lists the performance of the 6 basic classifiers on the data set with 31 attributes and the data subset with 12 attributes. For comparison, the performance of the same classifier is appended to the table. In terms of accuracy, the performance of the classifier logistic regression is the best. The LR classifier is better on a subset of data with 12 attributes (94.9614% < 95.1836%).

After the data set was standardized, LR achieved higher accuracy again in all 6 classifiers. Tuned accuracy is another measure of classifier accuracy. If there are too many false positives in the model, we start to set the sensitivity level to "narrow". Fine-tuning machine learning prediction models are used to improve the accuracy of prediction results. In both cases, the adjusted accuracy of LR is better than the other classifiers.

The ROC curve of the basic classifier and the proposed subset of data were analyzed. The ROC curve takes the "false positive rate" (showing the level of misalignment in the positive category) as x axis and the "true positive rate" (showing the level of correct classification in the positive category) as y axis. The area under the ROC polyline (AuROC) shows that the classifier gives a higher probability of prediction in the case of true positive than in the case of a true negative. Since the representation of each classifier is acceptable, it is difficult to identify the ROC curve of each classifier in the graph. Each of the two figures shows the comparison accuracy to better enhance the visualization. Of the two ROC curve in Table 8, LR has the highest accuracy, i.e., 99%.

Ensemble Techniques

This first attempt was to fluctuate application information and merge various copies of separate learning algorithms applied to each subset of the data. The basic inspiration for joining the model is to reduce the possibility of misclassification that relies on any single excitation model by mixing the specialized topics of the framework by mixing. To be sure, an understandable hypothesis determined by the model in metalearning is that there is an ideal learning algorithm for each assignment [29].

In the reduced data subset, we use AdaBoost, GradientBoosting, RandomForest, ExtraTrees, Bagging, and XGBoost as ensemble models. Table 8 shows the ROC curve accuracy of the ensemble model. The ExtraTrees classifier achieved the highest score, 95.1739%, followed by XGBoost 95.1691% and AdaBoost 94.7343% (see Table 9).

Stacking Classifier

Stack utilization differences among learners. They clearly performed two stages of learning: applying the learner to the basic level of the work that needs to be done, and applying another learner to the meta level of the information obtained from the basic learning [30].

At present, we have started all the models required in the Level-0 and stacked models at meta layer. We finally started

Table 8 Different accuracy metrics

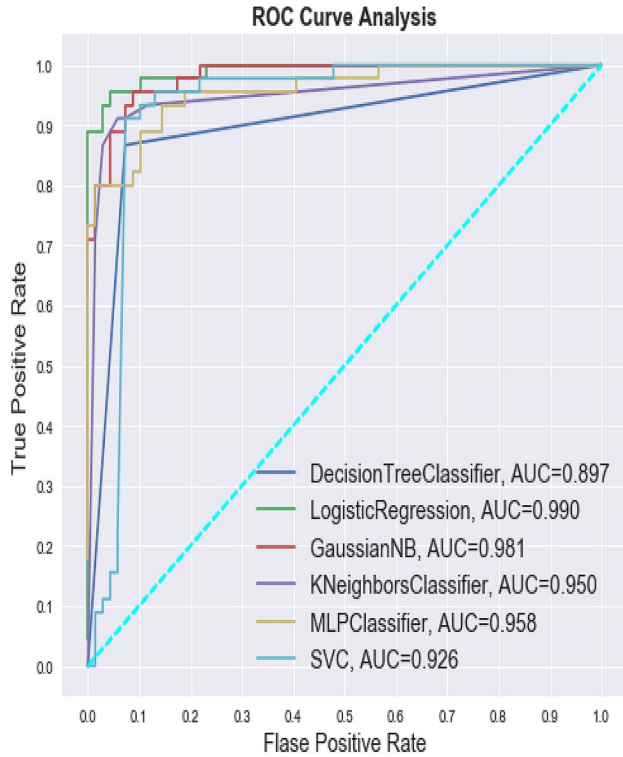
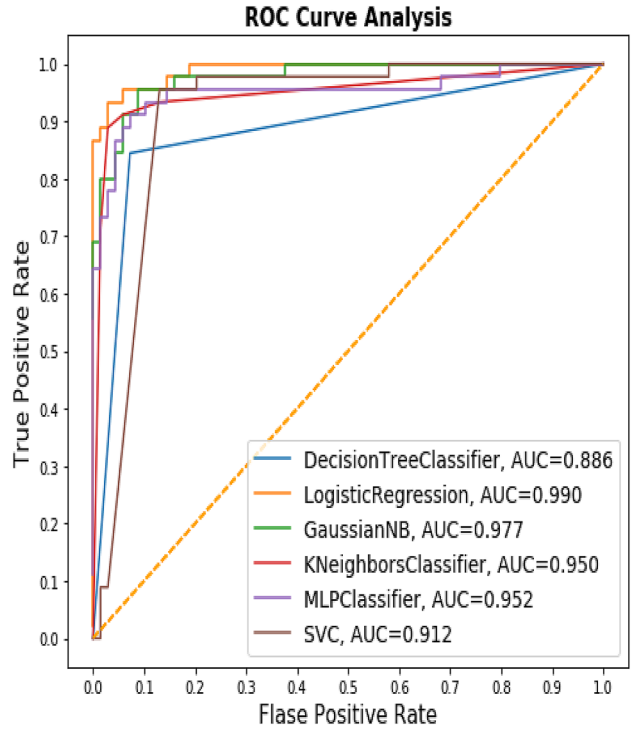
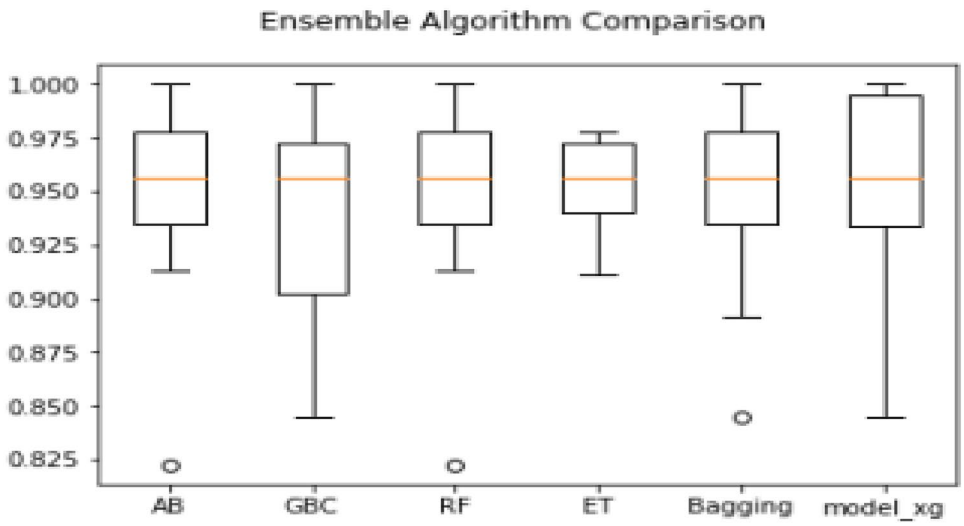
Metrics	Dataset with 31 attributes	Data subset with 12 attributes
Accuracy of basic classifiers	CART: 0.912077 (run time: 0.268554)	CART: 0.925362(run time: 0.501953)
	SVM: 0.619614 (run time: 0.520508)	SVM: 0.619614 (run time: 0.417969)
	NB: 0.940773 (run time: 0.040039)	NB: 0.938647 (run time: 0.031250)
	KNN: 0.927729 (run time: 0.151367)	KNN: 0.925507 (run time: 0.106445)
	LR: 0.949614(run time: 0.178711)	LR: 0.951836 (run time: 1.174805)
	MLP: 0.788068 (run time: 1.350586)	MLP: 0.740386 (run time: 0.810547)
Standardized data accuracy	ScaledCART: 0.914396 (run time: 0.854492)	ScaledCART: 0.923092 (run time: 0.049805)
	ScaledSVM: 0.964879 (run time: 0.078125)	ScaledSVM: 0.958261 (run time: 0.049805)
	ScaledNB: 0.931932(run time: 0.031250)	ScaledNB: 0.942995 (run time: 0.031250)
	ScaledKNN: 0.958357 (run time: 0.076172)	ScaledKNN: 0.949565 (run time: 0.031250)
	ScaledLR: 0.969324 (run time: 0.118164)	ScaledLR: 0.964928 (run time: 0.050781)
	ScaledMLP: 0.967101 (run time: 7.107422)	ScaledMLP: 0.958309 (run time: 8.562500)
Tuned accuracy	(LR) 0.977333 Run Time: 0.013672	(LR) 0.974912 Run Time: 0.832031
ROC curve		

Table 9 Ensemble accuracy of classifiers

Classifier	Accuracy	Box Plot
AB	94.7343	
GBC	93.8599	
RF	94.7295	
ET	95.1739	
Bagging	94.5169	
XGBoost	95.1691	

to use a stacked model with AdaBoost, random forest, extra trees, logistic regression, and decision trees on the 0th layer, and a voting classifier on the meta layer. We currently anticipate the relevant variables in the Test dataset and check the accuracy of this stacked model based on these expectations. We obtain 92.9824% accuracy from this model.

Conclusion

In the feature extraction and prediction technique, malignant growth is the disease with the second highest analysis frequency.

The classification of the classifier shows an incredibly great significance, especially for the identification of malignant cases. This inspection proposes a feature selection method (mode) using a basic classifier, an ensemble model with stacking classifiers to classify the instances with all attributes in comparison to reduced data subset. It is described as benign or malignant, and achieves an overall accuracy of 99% through the basic classifier. On the WBCD data set, it is 95.1739% in the ensemble model and 92.9824% in the stack classifier. By comparing the data set and the data subset, the basic classifier is recognized for its legitimacy in stack and ensemble model in terms of accuracy, accuracy at standardized data, tuned accuracy and AuROC.

Unnecessary attributes need not appear in the data set. These attributes may affect the accuracy of the data set, may produce over fitting and consume time for prediction. Following these ideas, the legitimacy and clinical estimates of the ensemble model and stack model proposed in this study were confirmed.

Discussion

The main idea using in this study is statistical technique for features selection to eliminate redundant attributes from data set. The survey can be used to compare situations, such as the type of diabetes, cervical malignant growth endurance rate, identifiable evidence of disease tumor cells, and quite different areas, such as sentiment analysis, drug classification, facial recognition, car driving Pedestrian identification, credit score, or spam discovery, where the attributes of data set is necessarily irrelevant or less relevant, indicates a difference from the specification. In addition, by passing the important classifier with stacking, ensemble and mode, it does allow the modularity of the entire model. After basic information preprocessing, the data set with reduced features and binary classification can directly utilize this study procedure. During this period, the model still has some shortcomings. Clinical and clinical data less dedicated for classification, containing more missing values and anomalies, and more data that may affect the performance of classification. When managing high-dimensional data sets, precision and specificity, confusion matrix and other indicators should be

considered. These problems make the proposed model not directly applicable to the clinic. Similarly, the choice of feature selection method, the decision of the type and number of pattern classifiers may additionally affect the execution of the performance, just like the time efficiency of allocation. Future work may include a system to check whether the standard classifier is indeed ideal and try to build it if necessary. With higher dimensions and more examples, deep learning strategies may also help to achieve better classification performance.

Compliance with Ethical Standards

Conflict of Interest Authors declare no conflict of Interest.

References

1. https://www.nationalbreastcancer.org/about-breast-cancer/, 2019.
2. Luca M, Kleinberg J, Mullainathan S. Algorithms need managers, too. Brighton: Chapman & Hall Ltd; 2016.
3. Coiera E. Guide to medical informatics, the Internet and telemedicine. London: Chapman & Hall Ltd; 1997.
4. Elsayad AM. Predicting the severity of breast masses with ensemble of Bayesian classifiers. J Comput Sci. 2010;6(5):576–84.
5. Huang M, Hung Y, Chen W. Neural network classifier with entropy based feature selection on breast cancer diagnosis. J Med Syst. 2010;34:865–73. https://doi.org/10.1007/s10916-009-9301-x.
6. Lavanya D, Rani DK. Analysis of feature selection with classification: Breast cancer datasets. Indian J Comput Sci Eng (IJCSE). 2011;2(5):756–63.
7. Bekaddour F. A neuro-fuzzy inference model for breast cancer recognition. Int J Comput Sci Inf Technol. 2012;4(5):163–73.
8. Al-Bahrani R, Agrawal A, Choudhary A (2013) Colon cancer survival prediction using ensemble mining on SEER data. In: Proceeding of IEEE International Conference on Big Data, pp 9–16.
9. Zheng B, Yoon SW, Lam SS. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Syst Appl. 2014;41(4):1476–82.
10. Chaurasia V, Pal S. Data Mining techniques: to predict and resolve breast cancer survivability. IJCSMC. 2014;3:10–22.
11. Zhang L, Li J, Xiao Y, et al. Identifying ultrasound and clinical features of breast cancer molecular subtypes by ensemble decision. Sci Rep. 2015;5:11085. https://doi.org/10.1038/srep11085.
12. Hazra A, Mandal S, Gupta A. Study and analysis of breast cancer cell detection using Naïve Bayes, SVM and ensemble algorithms. Int J Comput Appl. 2016;145(2):0975–8887.
13. Nilashi M, Ibrahim O, Ahmadi H, Shahmoradi L. A knowledge-based system for breast cancer classification using fuzzy logic method. Telemat Inf. 2017;34(4):133–44.
14. Chaurasia V, Pal S, Tiwari BB. Prediction of benign and malignant breast cancer using data mining techniques. J Algorithms Comput Technol. 2018;12(2):119–26.
15. Emami N, Pakzad A. A new knowledge-based system for diagnosis of breast cancer by a combination of affinity propagation clustering and firefly algorithm. J AI Data Min. 2018;7:59–68.
16. Kadam VJ, Jadhav SM, Vijayakumar K. Breast cancer diagnosis using feature ensemble learning based on stacked sparse Autoencoders and Softmax Regression. J Med Syst. 2019;43:263. https://doi.org/10.1007/s10916-019-1397-z.
17. Saritas M, Yasar A (2019) Performance Analysis of ANN and Naive Bayes classification algorithm for data classification. In: IJISAE, 2019, vol. 7, no. 2, pp. 88–91.
18. Rahman MA, Muniyandi RC. An enhancement in cancer classification accuracy using a two-step feature selection method based on artificial neural networks with 15 neurons. Symmetry. 2020;12:271.
19. Dua D, Graff C. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science. 2019.
20. Batyrshin I. Constructing time series shape association measures: Minkowski distance and data standardization. In: BRICS CCI 2013, Brasil, Porto de Galhinas. 2013. http://arxiv.org/pdf/1311.1958v3.
21. Kavitha R, Kannan E. An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining. in: IEEE Int. Conf. on Emerging Trends in Engineering Technology and Science (ICETETS), 2016, pp 1–5.
22. Uysal AK, Gunal S, Ergin S. The impact of feature extraction and selection on SMS spam filtering. Electronics and Electrical Engineering. 2013;19(5):67–72.
23. Maier O, Wilms M, von der Gablentz J, Krämer UM, Münte TF, Handels H. Extra Tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences. J Neurosci Methods. 2015;240:89–100.
24. Li L, Cui X, Yu S, Zhang Y, Luo Z, Yang H, Zhou Y, Zheng X. PSSP-RFE: accurate prediction of protein structural class by recursive feature extraction from PSI-BLAST profile, physical-chemical property and functional annotations. PLoS One. 2014;9:e92863.
25. Scanlon P, Kennedy IO, Liu Y. Feature extraction approaches to RF fingerprinting for device identification in femtocells. Bell Labs Tech J. 2010;15(3):141–51.
26. Kwac K, Lee H, Cho M. Non-Gaussian statistics of amide I mode frequency fluctuation of N-methylacetamide in methanol solution: linear and nonlinear vibrational spectra. J Chem Phys. 2004;120:1477–90.
27. Labatut V, Cherifi H Accuracy measures for the comparison of classifiers. 2012. http://arxiv.org/abs/1207.3790.
28. Guyon I, Gunn S, Nikravesh M, Zadeh L, editors. Feature extraction, foundations and applications. New York: Springer; 2006.
29. Araque O, Corcuera-Platas I, Sanchez-Rada JF, Iglesias CA. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Syst Appl. 2017;77:236–46.
30. Malmasi S, Dras M. Native language identification with classifier stacking and ensembles. Comput Linguist. 2018;44(3):403–46. https://doi.org/10.1162/coli_a_00323.