Improving Classification Accuracy of Breast Cancer Using Ensemble Methods

HAMMAM ABDELAAL[1], MOHAMED WAHBA[2], NEBAL OMRAN[3], HANY ANIS[4], MOTASEM ELSHOURBAGY[5,6]

[1]Department of Information Technology, Faculty of Computers and information, Luxor University
[2]Head of integration test dept. Egyptian Space Agency, Egypt
[3]Egyptian Company for blood transfusion services, Egypt
[4]Department of Computer and Electronics Engineering, Thebes Higher Institute of Engineering
[5]Department of Physics and Engineering Mathematics Mattaria, Faculty of Engineering, Helwan University, Cairo, Egypt
[6]Department of Software Engineering and Information Technology, Faculty of Engineering and Technology, Egyptian Chinese University, Egypt

Abstract:

Artificial intelligence plays an important role in medical sector, especially in improving healthcare for patients, in which the early detection and diagnosis of disease increasing the probability of recovery. Breast cancer ranks first among the most common types of cancer, globally, regionally. This paper with the help of machine learning technique proposes to present a non-invasive method for diagnosing and classify breast diseases based on mammograms and ultrasound images, to extract the statistical features of them (smoothness, perimeter, area, concavity, compactness, symmetry, size, diameter, concave and radius), to identify the breast tissue as malignant tumor, or a benign tumor and predicting in the future at the long term to prevent it. Learning algorithms are used mainly: support vector machine (SVM), multilayer perceptron (MLP), naïve Bayes (NB) and Decision tree (DT) algorithms to build model capable of classifying the breast tissue into malignant or a benign, based on several features reached up to 30 features. Ensemble methods were used in this study to improve the classification accuracy mainly: bagging, boosting and stacking on the same dataset that we have used it before in the classification using individual classifier. The Results showed that SVM achieved higher accuracy which is reached up to 97.89%, followed by MLP classifier with 95.61%, and the NB accuracy which is reached up to 92.62%. Also, the experimental results showed that the ensemble method is given higher accuracy than individual classifier, where the accuracy of Decision tree (DT) is increased from 93.15 as individual classifier to 97.71% using stacking algorithm

Keywords: Breast Cancer, Supervised Learning, SVM, MLP NB, DT, ensemble methods

1. INTRODUCTION

The field of artificial intelligence has advanced to solve several problems in various areas of life, especially in medical sector. Where we can use artificial intelligence through supervised learning algorithms such as support vector machine (SVM), multilayer perceptron (MLP), naïve Bayes (NB) and Decision tree (DT) to make a model that searches for signs of breast cancer through mammograms and predict the breast cancer[1][2]. AI-based algorithms represent a promising way to improve imaging accuracy digital mammography. AI system is able to classify images and texts into normal, benign tumors or cancer after learning and testing them, so it is able to teach itself and gain experience like humans, then the second stage is to follow up on the condition and the development of its recovery, then the third stage is to add or implement the long-term prediction, the prediction feature at the level of the generation resulting from the individual, from measuring the long-term incidence of the disease in parents in addition to their family history, and then measuring the percentage of giving birth to a generation that has the later gets sick.

Building a model that is able to distinguish and classify breast cancer into benign or malignant tumor, based on statistical features extracted from mammograms or ultrasound images, these images are as an input for the system to extract some of important the features, based on these features the model can classify and differentiate image categories, and follow up on the recovery and treatment of the patient. In addition to predicting it's occur in the future (long-term) based on some other checkup such as pathological history or the genetic factor, also improving the classification accuracy using ensemble method such as the stacking algorithm, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any learning algorithm as individually.[3][4], which is usually infinite, a machine learning ensemble consists of only a concrete finite set of alternative models, but typically allows for much more flexible structure to exist among those alternatives. An ensemble algorithm has better accuracy than single classification techniques. The success of the ensemble approach depends on the variety in the individual classifiers with respect to misclassified instances [5].

In our experiments, we used popular tool WEKA, it is an important tool for applying machine learning algorithms, through results showed that using ensemble methods achieved higher accuracy for DT classifier, which is reached up to 97.72 % with stacking, followed by boosting, which is reached up to 95.78% and finally the bagging, which is reached up to 95.08%, it is more accuracy than using individual classifier. Improvement ratio for each classifier (stacking, boosting, and bagging) is 4.57%, 2.63%, and 1.93% respectively.

2. PROPOSED SYSTEM

Breast cancer prediction is a process based on a deriving a classifier model; this model can describe and distinguish Breast cancer classes. The derived model is based on the analysis of a set of mammograms whose class label are known, to extract the features form these images, and feature selection to select the most important features [6][7], which are more related to target class. Then classifier learning process, it aims to classify the Breast cancer into two categories, these categories are Benign and Malignant, according to features of the images that distinguish of each class of them. The proposed system is consisting of different stages as shown in figure 1, each stage contains some of steps, started with the read mammograms as input for the system and ended with the final decision of identify breast tissue as normal, or abnormal (malignant tumor, or a benign tumor). Read of image to extract the features, then these feature are selected, to select the most important features, which are more related to target class, using Information gain technique, it used to measure the dependence between features and labels and calculates the gain between the (i-TH) feature f_i and the class labels according to equation 1, after calculating both of the expected information needed to classify a tuple in D is given by equation2, and the expected information required for each feature according to equation3 [8]

$$Gain(f) = Info(D) - Info_f(D) \qquad (1)$$

$$Info(D) = -\sum_{i=1}^{n} P_i \ log_2 \ (P_i) \qquad (2)$$

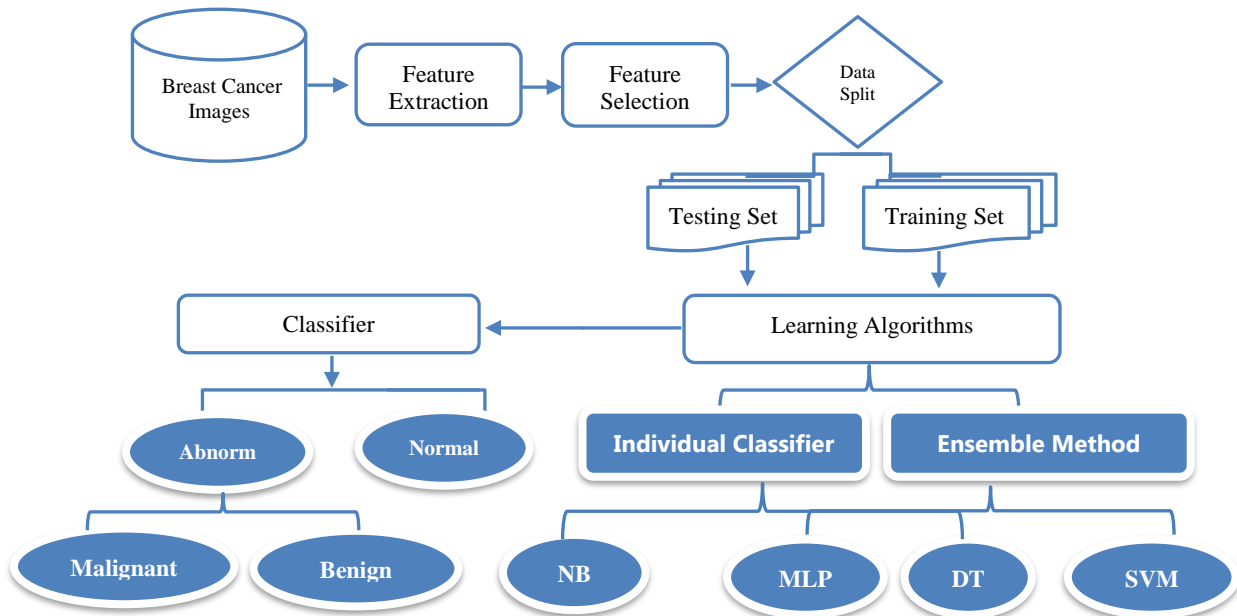$$Info_f(D) = \sum_{i=1}^{n} \frac{|Dj|}{|D|} \ x \ Info \ (Dj) \qquad (3)$$

Figure 1: Outline of the Breast cancer detection system

The data set is spited into training and testing using 10 folds cross-validation (CV) and percentage split (PS) methods. Data set contains of 570 samples, each sample has 30 features mainly: smoothness mean, perimeter mean, compactness mean, symmetry mean, perimeter worst and concavity worst. In cross validation [9], the dataset is divided into 10 folds. We use 9 of those parts for training and reserve one tenth for testing. We repeat this procedure 10 times each time reserving a different tenth for testing and calculate the accuracy for this iteration, and the end we get the overall accuracy, by calculate the mean these 10 measures. While in the percentage split, the dataset is divided randomly into 70% of the data set is used to train the model and 30% is used to test the model.

3. LEARNING ALGORITHMS

There are many algorithms that are used to train the classification model that can be used it to predict the classes of new images is unseen samples. In supervised learning, the data are labeled with pre-defined classes such as Multinomial, Logistic Regression, SGD Classifier, SVC, Linear SVC, Nu SVC, MLP, SVM and NB Classifiers. In this study we used the most of these algorithms to identify the best classifier gives high accuracy, to predict the class of breast cancer, based on mammography; it is a special type of x-ray imaging used to create detailed images of the breast. After the experimental results, we reported the best three classifiers of them, which have a highest accuracy mainly: SVM, MLP, VB and DT that have improved its accuracy using combination method (stacking algorithm). The learning process is two stages, the first Learning (training): Learn a model using the training data, and the second is testing, test the model using unseen test data to assess the model accuracy

4. CLASSIFIER EVALUATION

Classifier evaluation and testing are the final stage, it evaluates and testing of classifier; to validate of experimental results and determine the classifier capability to differentiating and classify image class to take the right final decision. The accuracy of the model is evaluated using different measures; mainly the Recall, Precision, and F-measure according to the following equations, based on determine the TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative)

TP: sick people correctly identified as sick, TN: Healthy people correctly identified as healthy
FP: Healthy people incorrectly identified as sick, FN sick people incorrectly identified as healthy.

$$(4) \qquad , \qquad Precision = \frac{T_P}{T_P + F_P} \qquad (5) \text{Acc} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}$$

$$Sensitivity = \frac{T_P}{T_P + F_N} \qquad (6) \quad , \qquad F - measure = \frac{2(precision * recall)}{precision + recall} \qquad (7)$$

5. EXPERIMENTAL RESULTS

Many classifiers have been used in this study, to identify the best classifier gives high accuracy. Table 1 and figure 2 show the overall accuracy using percentage split and cross-validation methods for each classifier (SVM, MLP, NB, and DT) as individual classifier, based on the number of instances that are correctly and incorrectly classified.

Method Classifier	Correctly Instances	Incorrectly Instances	Percentage split	Cross-validation
SVM	557	12	97.41%	97.89%
MLP	544	25	96.37%	95.61%
NB	527	42	91.19%	92.62%
DT	530	39	91.71%	93.15%

Table 1 Overall accuracy using percentage split and cross-validation methods for each classifier
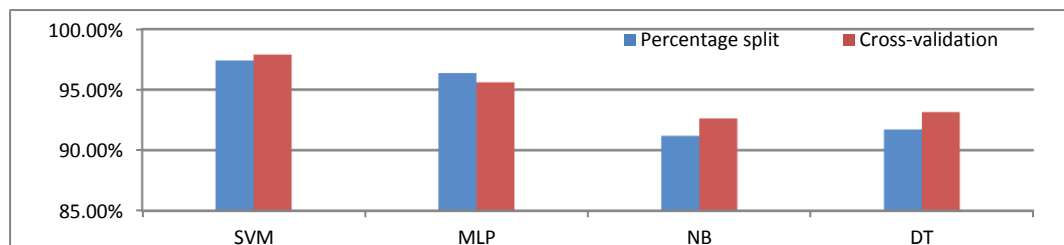

Figure 2 Overall accuracy using percentage split and cross-validation methods for each classifier

Figure 3 and figure 4 show the results of the F1-score, Recall, Precision, MCC and ROC for each category (Malignant and Benign), using cross-validation, which generated by the two classifiers mainly: SVM and MLP

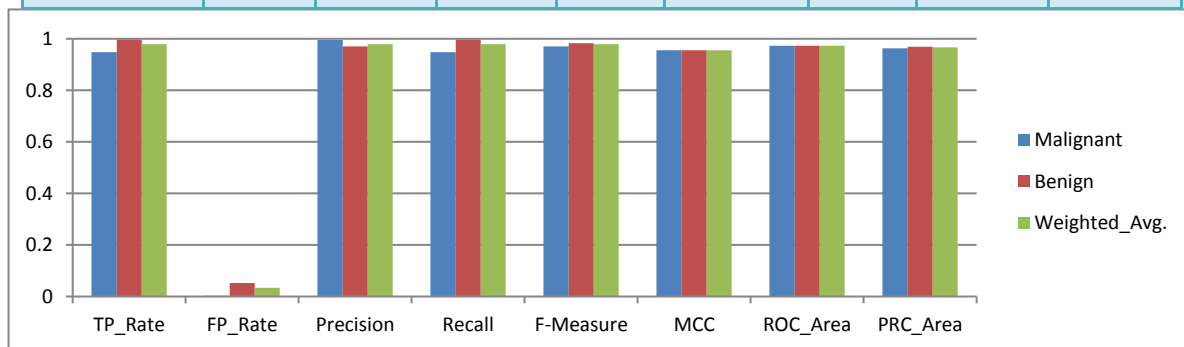| Class	TP_Rate	FP_Rate	Precision	Recall	F-Measure	MCC	ROC_Area	PRC_Area
Malignant	0.948	0.003	0.995	0.948	0.971	0.955	0.973	0.963
Benign	0.997	0.052	0.97	0.997	0.983	0.955	0.973	0.969
Weighted_Avg.	0.979	0.034	0.979	0.979	0.979	0.955	0.973	0.967

Figure 3 F-measure, Recall, Precision and MCC for each category using SVM Classifier using CV method

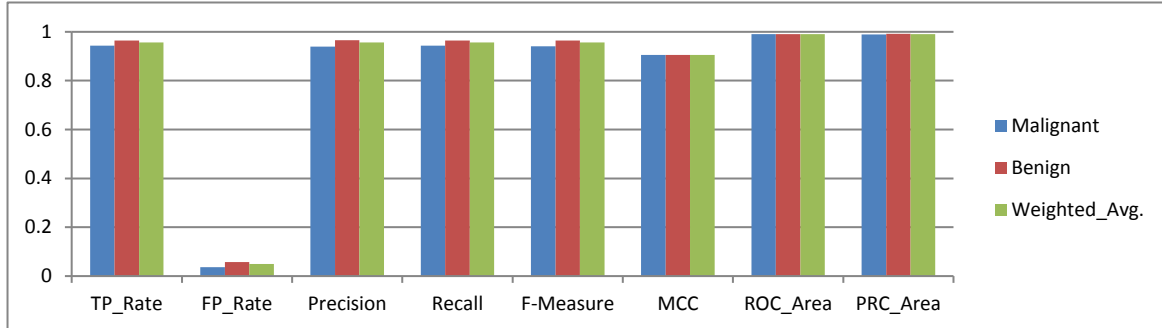| Class	TP_Rate	FP_Rate	Precision	Recall	F-Measure	MCC	ROC_Area	PRC_Area
Malignant	0.943	0.036	0.939	0.943	0.941	0.906	0.99	0.989
Benign	0.964	0.057	0.966	0.964	0.965	0.906	0.99	0.992
Weighted_Avg.	0.956	0.049	0.956	0.956	0.956	0.906	0.99	0.991

Figure 4F-measure, Recall, Precision and MCC for each category using MLP Classifier using CV method

6. IMPROVING CLASSIFICATION ACCURACY

Figure 5 shows the overall accuracy for DT (J48) and NB classifiers after and before applying the ensemble methods (bagging, boosting and stacking), to improve the classification accuracy, the accuracy of DT is increased from 93.15% to 97.72% and the accuracy of NB is increased from 92.62% to 96.78%, comparing to apply the DT, and NB as a single classifier. Three algorithms are used to improve the accuracy, the first is bagging. Bagging algorithm is based on the idea that bootstrap samples of the original training set will present a small change with respect to the original training set, but sufficient difference to produce diverse classifiers, each member of the ensemble is trained using a different training set, and the predictions are combined by averaging or voting[10][11]

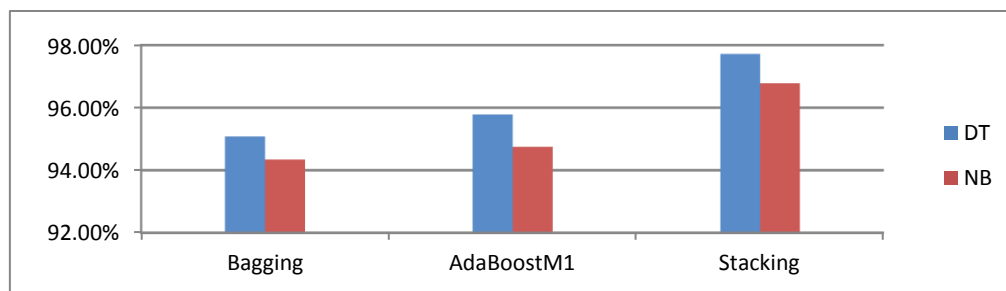| Classifier	Individual Classifier	Ensemble method		
		Bagging	AdaBoostM1	Stacking
DT	93.15%	95.08%	95.78%	97.72%
NB	92.62%	94.34%	94.75%	96.78%

Figure 5 the overall accuracy for DT and NB after applying ensemble method

The second is boosting, it is a technique to calculate the output using many different methods and then average the result using a weighted average technique. Finally, the third algorithm is stacking, it is a different technique of combining multiple classifiers depending on choosing different Meta classifiers, and the number of stacking folds, it are two levels of classification, the first level used as base classifiers, and the second level it learnt a Meta classifier based on the results of base classifiers to give the final classification result, therefore the stacking algorithm takes more time than other classifiers as bagging and boosting to build a model[12]

7. CONCLUSION

Artificial intelligence systems are used extensively in medical applications, such as the diagnostic and treatment, in additional to predict the disease before it occurs or discovers in its early stages, using learning algorithms for building a classifier model. In this study, we build classifier model using SVM, MLP, NB and DT and DT. They can describe and distinguish breast cancer type into benign or malignant tumor, based on the analysis of a set of mammograms. The Results showed that SVM achieved higher accuracy which is reached up to 97.89%, followed by MLP classifier with 95.61% % respectively, and improved the classification accuracy for NB and DT classifiers, they reached up to 96.78% and 97.72% respectively. Also the ensemble algorithms are able to improve the accuracy compared to using the classifier as individual. Finally, the model is evaluated using two methods mainly: 10 folds cross-validation (CV) and percentage split, but the best is cross-validation, it depends resampling that uses different fold of the data to test and train the classifier on different iterations, and thus gives correct accuracy when testing it using unseen samples.

RREFERENCES

1. Ng EYK et al (2007) Detection and differentiation of breast cancer using neural classifiers with first warning thermal sensors. Inf Sci 177(20):4526–4538

2. Eom G et al (2017) Nanogap-rich Au nanowire SERS sensor for ultrasensitive telomerase activity detection: application to gastric and breast cancer tissues diagnosis. Adv Funct Mater 27(37):1701832

3. Polikar, R, "Ensemble based systems in decision making". IEEE Circuits and Systems Magazine 6 (3): 21– doi:10.1109/MCAS.2006.1688199, S2CID 18032543- 2006

4. Rokach, L. (2010). "Ensemble-based classifiers". Artificial Intelligence Review. 33 (1–2): 1–39. doi:10.1007/s10462-009-9124-7, S2CID 11149239

5. Lee, K.C, Cho, H, "Performance of Ensemble Classifier for Location Prediction Task", Emphasis on Markov Blanket Perspective, International Journal of u and e-Service, Science and Technology, Volume-3, No. 3, September 2010

6. H. M. Abdelaal, B. R. Elemary, and H. A. Youness, "Classification of hadith according to its content based on supervised learning algorithms," IEEE Access, vol. 7, pp. 152379-152387, 2019.

7. Yousif A. Alhaj, Wiraj Udara, H M. Abdelaal, "Efficient Feature Representation Based on the Effect of Words Frequency for Arabic Documents Classification", Conference Paper: November 2018

8. Jiawei Han, Michelins Kamber, and Jian Pei, "Data Mining: Concepts and Techniques", the Third Edition University of Illinois at Urbana Champaign, 2012 Elsevier Inc.

9. H. M. Abdelaal and H. A. Youness, "Knowledge Discovery in the Hadith according to the reliability and memory of the reporters using Machine learning techniques, Digital Object Identifier 10.1109/ACCESS.2019.2944118 " IEEE Access,. September-2019

10. R. Lior, "Pattern Classification Using Ensemble Methods", New Jersey: World Scientific Publishing Co. Pte. Ltd, 2010

11. Bagging, Boosting, and Stacking. http://stats.stackexchange.com/questions/18891/bagging-boosting-and-stacking-in-machine-learning

12. H M. Abdelaal, and H A.Youness, 'Improve the automatic classification accuracy for Arabic tweets using ensemble methods', Journal of Electrical Systems and Information Technology, published by Elsevier, December 2018.

13. ABDELAAL, HAMMAM. AUTOMATIC CLASSIFICATION OF BREAST CANCER. International Journal of Artificial Intelligence and Emerging Technology, Vol. 5, No. 1, 2022, pp. 10-35.