IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11335

An Artificial Neural Network based approach along with Recursive Elimination Feature Selection Combined Model to detect Breast Cancer

Shiladitya Bose[1], Vishal Kumar Jha[2], Sk Tousif Hossain[3], Dr.Avijit Kumar Chaudhuri[4], Shulekha Das[5]

[1,2,3]UG - Computer Science and Engineering, Techno Engineering College Banipur, Banipur College Road, Banipur, Habra, West Bengal 743233

[4]Assistant Professor, Computer Science and Engineering, Techno Engineering College Banipur, Banipur College Road, Banipur, Habra, West Bengal 743233

[5]Assistant Professor, Computer Science and Engineering, Techno Engineering College Banipur, Banipur College Road, Banipur, Habra, West Bengal 743233

Abstract: Cancer is one of the deadliest diseases and of all the cancers breast cancer is the prime reason of cancer death among women as compared to men, today 1 in 8 women are suffering from breast cancer according to the American Cancer Society (Information collected from cancer.org website https://bit.ly/3uc13B3). As of now breast cancer accounts for 41 % of cancer deaths in women and this number is likely to increase by 2030, according to the world health organization (Link: https://bit.ly/3ioT8eh) in 2020 there were 2.3 million women diagnosed with breast cancer and the death toll was 685 ,000. Some of the factors that are contributing to this disease are usage of alcohol, exposure to cigarette smoke, no or minimal breastfeeding, lack of physical activity, family history of breast cancer, obesity, gene change and exposure to radiation. Through this research paper we are trying to investigate breast cancer triggering factors and applying data mining models to work on early detection based on patients' medical history and predicting where this disease will reoccur or not with accurate results. Our data mining model uses Recursive feature Elimination method with cross validation, Feature Selection and stacked with Artificial Neural Network to detect breast cancer. Furthermore, we have also compared the Artificial neural network classifier with other Machine learning algorithms to find the accuracy. To add some statistics into our model we have used concepts like Specificity, Sensitivity, ROC-AOC Score, Kappa-Cohen score, Wilcoxon Signed Rank test and other statistical parameters to check and compare the Artificial Neural Network model with other Machine Learning Classifier Algorithms. In our model we have represented other Machine learning classifier algorithms as Logistic Regression, Decision tree, Random Forest classifier, K-Nearest Neighbor, Support Vector Machine, Gradient Boosting and Naive-Bayes. In this Paper we have Proposed a stacked model which can outperform other Machine Learning Classifier Algorithm by calculating various statistical parameters and by conducting non-parametric test to prove our hypothesis.
In this research Paper the authors observed 98 percent accuracy by using Artificial Neural Network Based Approach along with Recursive Elimination Feature Selection combined model with Hyperparameter Tuning so that we observed sensitivity being 100 percent and specificity 99 percent also the ROC-AUC Score is 100 percent and the kappa score is 99 percent.

Keywords: Machine Learning, Deep Learning, Breast Cancer, Feature Selection

1. INTRODUCTION

Breast cancer is the most common cause of cancer death among women after lung cancer. Breast cancer begins when healthy cells in the breast change and grow out of control, forming a mass or sheet of cells called a tumour, a tumour can be cancerous or benign, a cancerous tumour is malignant meaning it can grow and spread to other parts of the body, a benign tumour means the tumour can grow but will not spread. (Collected from cancer.net, Link: https://bit.ly/3Jr2Qst)

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11335

Worldwide breast cancer comprises 10.4% of all cancer incidences among women, making it the second most common type of non-skin cancer (after lung cancer) and the fifth most common cause of cancer death. The risk factors of breast cancer include starting menopause at a later age, having no children or having first children after age 30, women with a previous history of breast cancer, not breastfeeding, lack of exercise and birth control pills.

Some Prevention of breast cancer is early detection as it is essential for treating the disease, not only this it can raise the potential for the cure by over 95% and reduce mortality by up to 30% also get screened for breast cancer regularly, know your family history of breast cancer, limit the amount of alcohol, avoid unnecessary medical radiation exposure, control your weight and do regular exercise.

There are many researchers who have mostly used machine learning methods to predict diseases [1,2,3,4]. Since then, many attempts have been made and skewed towards time to compute, and accuracy of prediction. For many years, there has been regular advancement pertinent to cancer research worldwide [5]. Several methods are applied by different scientists, for example, shielding in premature stage, in pursuance of finding types of cancer before the beginning of any symptoms. Furthermore, the scientists develop new strategies regularly to predict cancer beforehand to the treatment. At the same time with the advanced new technologies in medical field, an enormous amount of data in cancer field have been collated and readily available for medical research. On the other hand, predicting the disease outcome precisely is the most demanding and fascinating job for physicians [6]. There are many studies which have shown a high accuracy of prediction of breast cancer using various ML methods.

There are databases of diseases which contains lots and lots of data items, initially starting from demographic till diagnostic tests. Researchers [7,8,9,10,11,12,13] substantiated that prediction improves with the choice of right features. As information and storage technology are making exponential progress, data sets are now omnipresent in pattern analysis, data processing, machine learning (ML) and deep learning systems, with a vast range of variables or features [13]. Therefore, there is a need for choosing a subset which contains proper and ideal features that gives the best result.

In this paper, the authors introduce an artificial neural based network approach combined with recursive feature elimination (RFE) [35] method to identify the features and to predict the results using the right features.

Table.1. Data Set for Breast Cancer Performance Comparison

Year	Method	Conclusion
Sridevi &Murugan, 2014[14]	Multilayer perceptron (MLP)	Accuracy: 100%
Alickovic & Subasi, 2017[15]	Rotation Forest model classifies using GA	Accuracy: 99.48 ROC/AUC:0.993
Hamsagayathri & Sampath, 2017[16]	Priority based decision tree classifier	Accuracy: 93.63 Sensitivity/ Specificity: 0.936/0.982 Kappa Score:0.925 ROC/AUC: 0.929
Abdar et al., 2018[17]	SV-Naive Bayes-3MetaClassifiers	Accuracy: 98.07 Sensitivity/ Specificity: 0.981 ROC/AUC: 0.976
Zheng et al., 2014[18]	K-SVM	Accuracy: 97.38% ROC/AUC: Done
Sewak et al., 2007[19]	Ensemble SVM	Accuracy: 99.29% Sensitivity/ Specificity: 1/0.981
Jin et al., 2012[20]	Functional Trees (FT)	Accuracy: 97.33% Sensitivity/ Specificity: 0.946 Kappa Score: 98.85
Obaid, et al., 2018[21]	Quadratic Kernel Based	Accuracy: 98.1% ROC/AUC: 0.984305 for benign tumor and 0.988352 for malignant tumor
Kumari & Arumugam, 2015[22]	Hybrid Krill Herd	Accuracy: 87.89% Sensitivity/ Specificity: 0.975/ 0.718
A.K. Chaudhuri, et al. 2021[13]	Dataset-Centric-Approach (DCA)	Accuracy: 97% Sensitivity/ Specificity: 0.99/0.96 Kappa Score: 0.92 ROC/AUC:1 Wilcoxon: Done

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11335

2. RELEVANT LITERATURE

Data Mining and machine-learning techniques help us to generate automatically prognostic and diagnostic policies in a simple way. Data-mining analysis has shown an encouraging result, but one thing is sure that there the outcomes are not consistent with technologies as well as datasets. Researchers have tried most and tested different machine learning classifiers for predicting diseases, like SVM, RF, DT, LR, NB, GA and KNN.

There are many researchers who used many types of methods to predict the` Breast Cancer (BC) by using the Wisconsin (Diagnostic) Data Set. The names of those researchers mentioned in Table 1. Christobel and Sivaprakasam (2011) [23], used the classifiers DT, KNN, SVM & NB and compare the precision of their classification on the WDBC dataset for BC diagnosis. The accuracy they have achieved 96.99% of SVM, the highest accuracy among all the classifiers. the conclusion of their research restricts the use of a single classifier and a single dataset.

Lavanya and Rani (2011) [24], used the DT classification without using any feature selection techniques on the WDBC dataset for BC diagnosis. The classification techniques boost the accuracy to 70.63%, 96.99% and 92.09% respectively.

Keles et. al. (2011) [25], used fuzzy laws as a useful diagnostic tool on the WDBC dataset for BC diagnosis. The accuracy they have achieved is 97%.

Chen et. al. (2011) [26], used the rough-set vector classifier to diagnose BC. The conclusion of this research is that the algorithm identified five features that can help doctors classify BC and achieve high detection accuracy.

Salama et. al. (2012) [27] used NB, sequential-minimal optimization (SMO) along with DT(J48), Multi-Layer Perceptron (MLP) and instance-based KNN classifiers to determine the precision of the classification of various BC datasets. The conclusion of this research is that they suggested SMO most valid for the WDBC dataset.

Lavanya and Rani (2012) [28], used the ensemble classifier method on a BC dataset. The conclusion of this research is that pre-processing was used to improve efficiency and collection of features showed improved accuracy of the classification.

Kim et. al. (2012) [29], used the SVM technique on a BC dataset. The dataset consists of 679 records that included clinical, pathological and epidemiological data types. The conclusion of this research is that the precision achieved 99% with the role of local tumour.

Chaudhuri et. al. (2018) [30] analysed the dataset of BC to find a recurrence of the disease, using decision tree and discriminant analysis..

In this present research, the author used a simple framework to ensure all the conditions are met and to provide the better classification accuracy than Chaudhuri et. al. (2021) [13], the author used artificial neural network (ANN) [34] along with Recursive Feature Elimination (RFE) Feature Selection combined them in a Stacked Model to produce an accurate classification. Here the author compared the stacked model with other 8 ML algorithms LR, DT, RF, KNN, SVM, GB and NB. For both training and testing the dataset we have the combined RFE-ANN stacked model. In RFE, the author used Random Forest (RF) as an estimator to select the 15 best features from the dataset.

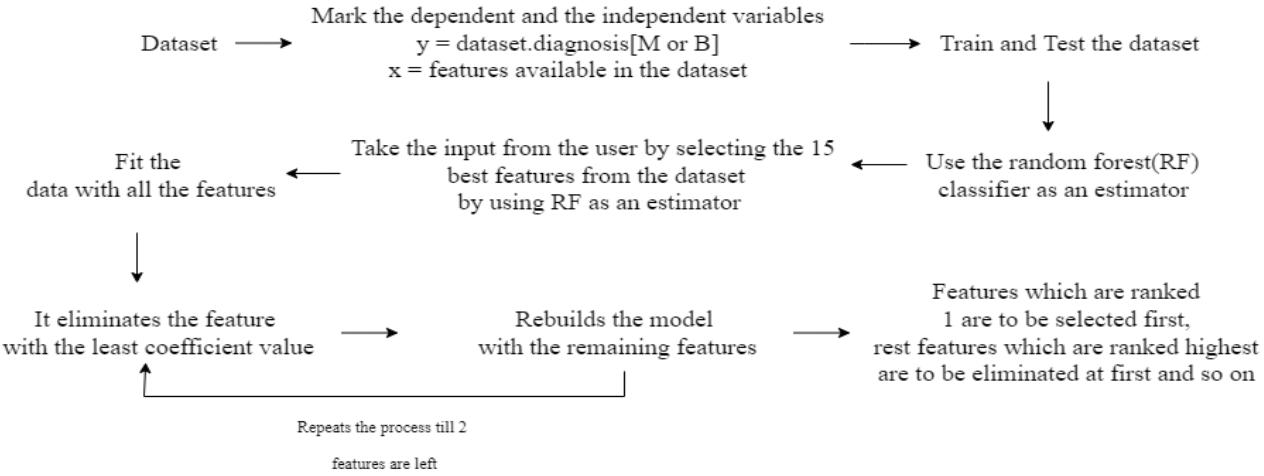


Fig.1. Flowchart of Recursive Feature Elimination (RFE) with RF Feature Selection

RFE is a wrapper-type feature selection algorithm. It uses filter-based feature selection internally. RFE works by searching for a subset of features by starting with all the features in the training dataset and successfully removing features until the desired number remains. Features are scored either using the provided machine learning model or by using a statistical method. Fig 1. Shows the working of RFE by using RF as an estimator.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11335

3. PROPOSED CLASSIFIER: RFE-ANN BASED MODEL

In this paper, the author used ANN as a classifier along with RFE feature selection supervised wrapper algorithm for detecting BC. ANN are algorithms based on brain function and are used to model complicated patterns and forecast issues. ANN is a deep learning method that arose from the concept of the human brain Biological Neural Networks. ANN algorithm accepts only numeric and structured data.(Link : https://bit.ly/3qk8WTV).

A single neuron is known as a perceptron[33]. It consists of a layer of inputs (corresponds to columns of a dataframe). Each input has a weight which controls the magnitude of an input. The summation of the products of these input values and weights is fed to the activation function. Activation functions are really important for an Artificial Neural Network to learn and make sense of something really complicated and Non-linear complex functional mappings between the inputs and response variable. They introduce non-linear properties to our Network. Their main purpose is to convert an input signal of a node in an ANN to an output signal. That output signal now is used as an input in the next layer in the stack. Specifically in ANN we do the sum of products of inputs(x) and their corresponding Weights(w) and apply an Activation function Φ(x) to it to get the output of that layer and feed it as an input to the next layer. Fig 2 given below shows the structure of neuron.

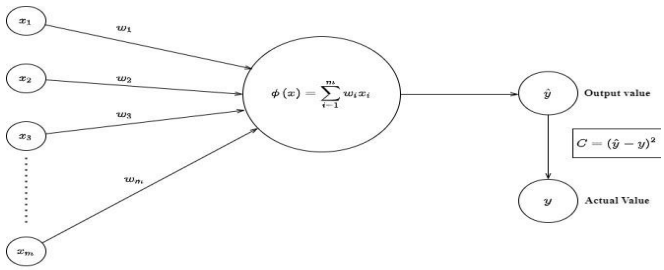


Fig.2. Neuron

In Fig 2., here C refers to the cost function, y is the actual output and ŷ is the predicted output from the neural networks. This formula is commonly known as Mean Squared Error (MSE) (1):

$$C = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \quad (1)$$

Activation functions are an extremely important feature of the ANN's. They basically decide whether a neuron should be activated or not. Generally, the formula of activation function can be described as (2).

$$y = \text{Activation}(\Sigma(\text{weight} * \text{input}) + \text{bias}) \quad (2)$$

In this research, the author used two non-linear functions, they are sigmoid and Rectified Linear Unit (ReLu). Non-Linear functions make it easy for the model to generalize or adapt with variety of data and to differentiate between the output.

The sigmoid function curve looks like S-shaped. The function reduces extreme values or outliers in data without removing them. It converts independent variables of near infinite range into simple probabilities between 0 and 1, and most of its output will be very close to 0 and 1. The formula for sigmoid function (3).

$$\phi(x) = \frac{1}{1+e^{-x}} \quad (3)$$

The ReLu is the most widely used activation function while designing networks. ReLu function can backpropagate the errors and have multiple layers of neurons being activated by the ReLu function. The formula for ReLu non-linear function (4)

$$\phi(x) = \max(x, 0) \quad (4)$$

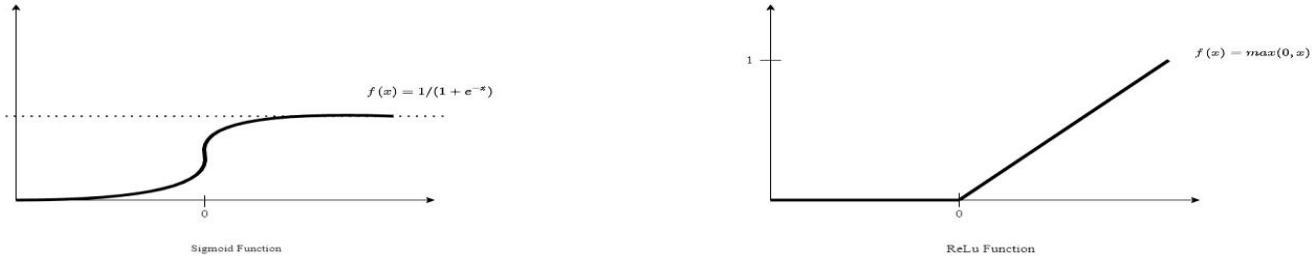


Fig.3. Activation Function graphs (Sigmoid and ReLu)

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11335

The RFE-ANN strategy used here is to use the RFE feature selection with Random Forest as estimator and ANN as a classifier to involve reducing error-rate production models. The framework of the ANN is shown in Fig. 4

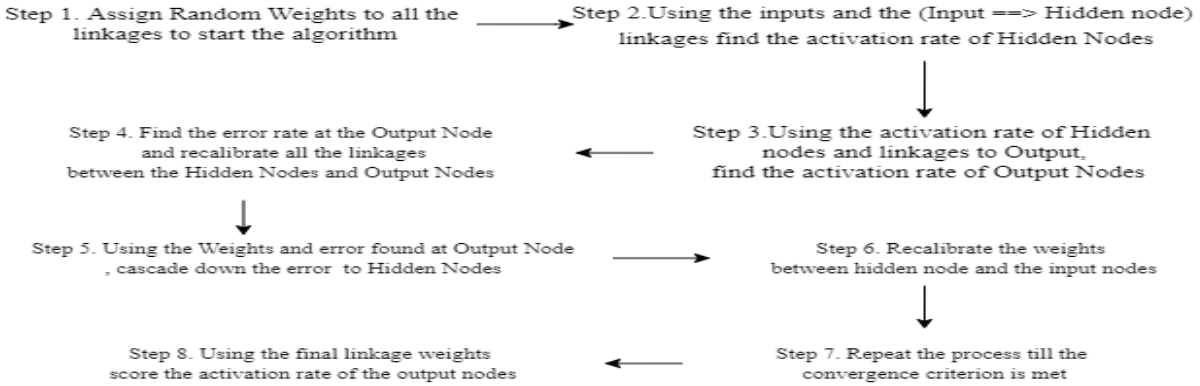


Fig. 4. Framework of Artificial Neural Network

4. ASSESSMENT OF PERFORMANCE OF ML ALGORITHMS

Evaluating the machine learning algorithm is an essential part, the model may give satisfying result when evaluated using a metric say accuracy but may give poor result when evaluated against other metrics such as sensitivity or any other such metric. The efficiency of any machine learning model is determined using measures such as true positive rate, false positive rate, true negative rate and false negative rate.
In this paper we have applied different types of evaluation metrics available to evaluate the classification performance of Machine learning algorithms.
Some of the metrics include.

4.1. Accuracy

When we use the term accuracy classification accuracy is what we actually mean, it is the ratio of number of correct predictions to the total number of input samples.

$$\text{Accuracy} = \frac{(\text{Tp} + \text{Tn})}{(\text{Tp} + \text{Tn} + \text{Fp} + \text{Fn})} \quad (5)$$

where, Tp, Tn, Fp and Fn indicates True Positive, True Negative, False Positive and False Negative respectively.

There are four important terms
True Positive: In this case our prediction is yes, and the actual output is also yes.
True Negative: In this case our prediction is no, and the actual output is no.
False Positive: In this case our prediction is yes, and the actual output is no.
False Negative: In this case our prediction is no, and the actual output is yes.

4.2 Kappa Statistics For each model

Cohen kappa statistics calculates the degree of agreement between a pair of variables, it is frequently used to test inter-rater reliability i.e., it most often deals with data that are a result of a judgement not a measurement.
The interpretation of kappa value is in such a way that if the value comes out to be less than 0 then no agreement is observed, if 0 < value < 20, slight interpretation is observed, if 0.41< value < 0.60, moderate interpretation is observed, 0.61 < value < 0.80, substantial interpretation is observed, and finally if the value comes out to be in between 0.81 and 1.0 then fair value is interpreted.
The Kappa score gives the measure of the classification in the range of [-1, 1] with 0 being no agreement and 1 being full agreement.

4.3 ROC Curve and AUC values

The Roc Curve (receiver operating characteristic curve) is a graph showing the performance measurement of classical problems at different thresholds. The ROC is a probability curve and AUC represent the degree or measure of separability . The ROC-AUC curve plots two parameters True Positive Rate (TPR) and False Positive Rate (FPR). The ROC curve is plotted with TPR against the FPR where TPR is on the y-axis and FPR is on the x-axis.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11335

An AUC value close to or equal to 1 means an excellent model as it has a good measure of separability while an AUC value close to 0 means it has a worst measure of separability . However when the AUC value is 0.5 it means that the model has no class separation. So by analogy we can say that higher the AUC the better is the model.

4.4 Wilcoxon ranked sum (WLS) test

The Wilcoxon signed-rank test is an example of a non-parametric test or distribution test to compare two identical and matched samples. The Wilcoxon signed-rank test is used if the differences between the pairs of data are non-normally distributed, it is used to test the null hypothesis that the median of a distribution is equal to some value. It compares the sample median against a hypothetical median.

5. DESCRIPTION OF THE DATASET

The authors use the Wisconsin (diagnostic) dataset for Breast cancer, created at Dr. William H. Wolberg General Surgery Dept, University of Wisconsin, clinical sciences center Madison, WI 53792 wolberg@eagle.surgery.wisc.edu available at the UCI ML Repository website . Dataset Website Link: https://bit.ly/3qjwCrq. The key highlights of this database is that it contains records of 569 patients out of which 212 are malignant and 357 are benign. This dataset contains 32 columns, which represents the feature of this dataset.

6. TRAINING TESTING PARTITION

The training test partitioning usually involves the partitioning of the data into a training set and a test set in a specific ratio ,the data partitioning can be done randomly or in a fixed way while the fixed way is typically avoided as it may introduce systematic difference between the training test and the test set , in machine learning obtaining separate samples from an original set is a common practice for predictive models to be constructed and tuned the most important thing is to utilize the independent test specimen to make a precise model with a preparation test and then approve the model afterwards, hence the probability of over-fitting will be decreased and will provide a reasonable accuracy model estimate and improve the generalization when using the model on new data. Table 2 shows the training testing set partition.

Table 2: Training and Testing Set Partition

Training-Testing Partition	Total Training Records	Positive Records in Training Set	Negative Records in Training Set
50-50	284	111(39%)	173(61%)
40-60	341	129(38%)	212(62%)
30-70	398	149(37%)	249(63%)
20-80	455	165(36%)	290(64%)
10-fold cross validation	569	212(37%)	357(63%)

7. RESULTS AND DISCUSSION

The authors developed and simulated a proposed model by using the Python Programming language and Google Colab Notebook. In this model, the authors perform a comparative study between ML algorithms namely LR, DT, RF, KNN, SVM, GB and NB, and our proposed model RFE-ANN. Among these eight popular techniques, some techniques show better accuracy, whereas some performances of some other techniques are inferior.

Table. 3. Comparison of Accuracies

Training Testing Partition	LR	DT	RF	KNN	SVM	GB	NB	RFE-ANN
50-50	0.95	0.92	0.95	0.94	0.91	0.94	0.93	0.98
40-60	0.97	0.91	0.94	0.95	0.92	0.96	0.92	0.98
30-70	0.95	0.92	0.97	0.95	0.92	0.96	0.92	0.98

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11335

20-80	0.94	0.94	0.96	0.94	0.92	0.97	0.92	0.98
10 FCV	0.97	0.91	0.96	0.96	0.97	0.95	0.93	0.98

Our RFE-ANN outperformed all the algorithms with respect to the accuracy and have achieved 98%. Table 4 shows the comparison of the standard deviation among these classifiers.

Table 5.1 and Table 5.2 shows the comparision values of sensitivity and specificity result generated from confusion matrices by using the 8 classifier algorithms. The performance of our proposed model, along with other methods evaluated and purely based with respect to sensitivity, specificity and accuracy tests, consists of True Positive (TP), True Negative (TN), False Negative (FN) and False Positive (FP) terms.From the given table, the results demonstrate the potential of our proposed model in the classification of two classes. It is determined from the comparative results that our proposed classification technique has the highest sensitivity, specificity and accuracy values (accuracy = 98%, sensitivity = 0.98, specificity = 0.99) for BC dataset. Table 6 shows the ROC-AUC scores of the ML algorithm including our proposed model used in this research paper.

Table.4. Comparison of Standard Deviation

Training Testing Partition	LR	DT	RF	KNN	SVM	GB	NB	RFE-ANN
Std. Dev. (10-FCV)	0.027	0.013	0.029	0.02	0.02	0.032	0.042	0.0187

Table.5.1. Comparison of Specificity

Training-Testing Partition	LR	DT	RF	KNN	SVM	GB	NB	RFE-ANN
50-50	0.95	0.95	0.97	0.95	0.99	0.97	0.95	0.98
60-40	0.97	0.95	0.97	0.97	0.99	0.98	0.95	0.99
70-30	0.94	0.91	0.96	0.96	0.99	0.97	0.93	0.98
80-20	0.94	0.94	0.97	0.95	0.98	0.97	0.94	0.96
10FCV	0.98	0.92	0.98	0.97	0.98	0.95	0.95	0.97

Table.5.2. Comparison of Sensitivity

Training-Testing Partition	LR	DT	RF	KNN	SVM	GB	NB	RFE-ANN
50-50	0.95	0.87	0.89	0.91	0.77	0.88	0.9	0.97
60-40	0.97	0.84	0.86	0.92	0.8	0.89	0.89	0.96
70-30	0.96	0.95	0.96	0.93	0.81	0.96	0.9	0.98
80-20	0.95	0.95	0.95	0.93	0.85	0.95	0.91	0.94
10FCV	0.96	0.89	0.92	0.92	0.94	0.93	0.89	0.97

Table.6. Comparison of ROC-AUC scores

Training-Testing Partition	LR	DT	RF	KNN	SVM	GB	NB	RFE-ANN
50-50	0.95	0.91	0.93	0.93	0.88	0.93	0.93	0.98
60-40	0.97	0.90	0.92	0.95	0.89	0.94	0.92	0.98
70-30	0.95	0.93	0.97	0.95	0.90	0.97	0.92	0.98
80-20	0.95	0.95	0.96	0.95	0.92	0.96	0.93	0.99
10FCV	0.99	0.91	0.98	0.98	0.99	0.99	0.98	1

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

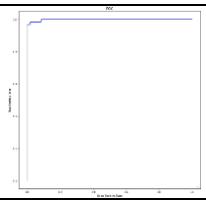
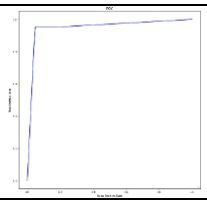
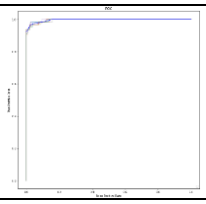
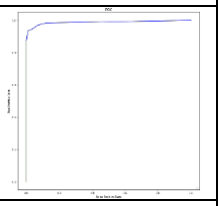
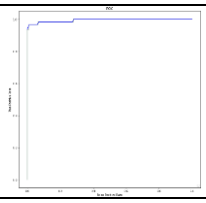
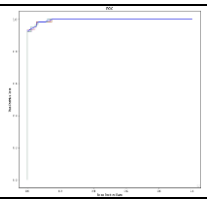
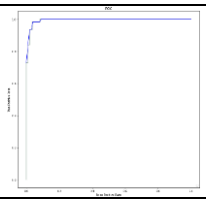
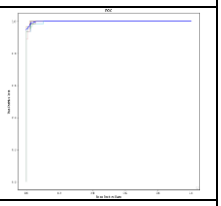
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11335

Our proposed classifier's analytical results seemed to be positive. In the absence of the disorder, fewer patients would need to be tested for BC due to higher specificity. At the same time, higher sensitivity value would also save money and shorten the waiting times of the actual ill patients that are critical to saving lives.

Table.7. Comparison of ROC curves and AUC values

Training-Testing Partition	LR	DT	RF	KNN
10-fold cross validation				
	0.99	0.91	0.98	0.98
Training-Testing Partition	SVM	GB	NB	RFE-ANN
10-fold cross validation				
	0.99	0.99	0.98	1

In this Table 7, eight ROC charts drawn in different parts for 10 cross-validation are shown. Experiments and our research shown that our proposed classifier model have outperformed all other seven ML techniques with respect to the ROC curve and AUC curves. In terms of cross validation accuracy, our proposed model's AOC score reached 1.

Comparing performances of different ML classifiers might generate an ambiguous result if it has been produced based only on accuracy-based metrics, for that reason we have used Cohen's Kappa Statistics. The Cohen's Kappa Statistics (CKS) value is used to help to produce error-free comparative efficiency of different classifiers. The cost of error must be considered in such evaluations. CKS is an excellent measure in this respect for inspecting classifications that may be due to chance. Usually, CKS takes a value between -1 to +1. As the classifiers calculated Kappa value approaches' 1,' the classifier's performance is assumed to be more realistic than 'by-chance'. Therefore, CKS value is a suggested metric for measurement purposes in the performance analysis of classifiers [31]. The formula for Kappa value is calculated by (6)

$$\mathrm{CKS} = \frac{(\mathrm{pa}-\mathrm{pbc}())}{(1-\mathrm{pbc}())} \qquad (6)$$

Table.8. Kappa Statistics for Each Mode

Training Testing Partition	LR	DT	RF	KNN	SVM	GB	NB	RFE-ANN
50-50	0.90	0.84	0.88	0.87	0.81	0.88	0.86	0.95
60-40	0.94	0.81	0.85	0.90	0.82	0.89	0.85	0.96
70-30	0.90	0.85	0.93	0.90	0.83	0.94	0.84	0.96
80-20	0.89	0.89	0.93	0.89	0.85	0.93	0.86	0.96
10 FCV	0.95	0.82	0.91	0.91	0.93	0.92	0.85	0.96

Where pa represents total agreement probability and pbc represents probability 'by-chance'. The results of CKS analysis of the seven popular ML techniques and our proposed model are shown in Table 8. According to this study, it can be proved easily that the proposed model performed much better than other classifiers (value=0.92).

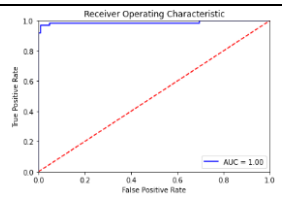
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11335

Table.9. Results with Hyperparameter Tuning

Parameter	Hyperparameter optimization
Accuracy/Sensitivity/Specificity	0.98/0.98/0.99
ROC_AUC Score: 1	

In this paper, the authors compare the results with wrapper-type feature selection algorithm-based hyperparameter optimization is used to derive an optimal solution for predicting breast cancer. This approach provided an appropriate level of accuracy. The authors considered a simple hidden layer neural network architecture, which is simply based on Multi-Layer Perceptron [33], used as a classifier to find out the classification accuracy. Fig.5. is an example of a simple hidden layer neural network architecture. In the classifier, the authors selected the 15 input features in the first input layer by using the ReLu activation function. In the second hidden layer, the authors used the ReLu activation function. In the last layer, sigmoid activation function is used by the author. In the classifier, Adam optimizer and binary cross-entropy cost function is used by the author. Optimizers are the algorithms or methods need to minimize an error function (loss function) or to maximize the efficiency of production. Adam (Adaptive Moment Estimation) is a method that computes adaptive learning rates for each parameter. It stores both the decaying average of the past gradients, similar to momentum and also the decaying average of the past squared gradients. A cost function is a measure of performance of a neural network with respect to its given training sample and the expected output. The formula for the binary cross-entropy cost function, also known as Bernoulli negative log likelihood, is (7)

$$C_{CE}(W, B, S^r, E^r) = -\sum_j [E_j^r \ln a_j^L + (1 - E_j^r) \ln(1 - a_j^L)] \quad (7)$$

Where W is our neural network's weight, B is our neural network biases, S^r is the input of a single training sample, E^r is the desired output of the training sample.

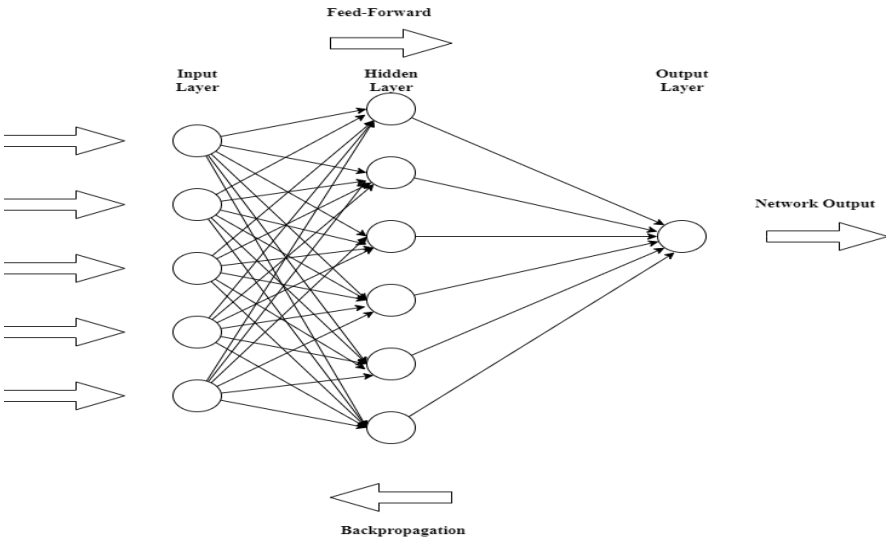


Fig.5. Formation of a simple hidden layer neural network architecture

Where W is our neural network's weight, B is our neural network biases, Sr is the input of a single training sample, E^r is the desired output of the training sample. The gradient of this cost function with respect to the output (a^L) of a neural network and some sample r is (8)

$$\nabla_a C_{CE} = \frac{(a^L - E^r)}{(1 - a^L)(a^L)} \quad (8)$$

By using this classifier, classification accuracy achieved is 98%. Here, the authors used Grid Search technique to estimate accuracy. The test set was set to 20%.

The authors have also used a statistical test proposed by Wilcoxon, which is a non-parametric test to compare the performance of different ML algorithms used in this study [32]. This hypothesis test is used to evaluate statistical differences among two populations by comparing the median of a single column of numerical values against a

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11335

hypothetical median. To handle this problem, authors utilized a one against one approach, which will break n number of classes into n(n-1)/2 binary classes containing a set of all possible pairs of n classes. In this study, 8- class classification is broken down into 28 binary sub-problems. Table 10 shows the Wilcoxon signed rank test table.

Table.10. Wilcoxon Signed-Rank Test

	DT	RF	KNN	SVM	GB	NB	RFE-ANN
LR	0.063	1	0.102	0.068	0.042	0.891	0.041
DT		0.041	0.068	0.853	0.461	0.042	0.041
RF			0.194	0.782	0.422	1	0.042
KNN				0.741	0.039	0.257	0.038
SVM					0.655	0.782	0.039
GB						0.042	0.033
NB							0.033

Table 10 list the p-values of the WLS test for pairs of accuracies from the analysis of different ML algorithms and our proposed algorithm model performed in this research work by splitting the BC dataset into various training and testing partitions.WLS test indicates that these algorithm pairs are mutually convergent. The authors cannot reject the null hypothesis for the above cases. However, our proposed model produces different results with seven other classifiers, and in all such cases, the p-value is less than 0.05, which means the medians of these two distributions differ. Thus, the null hypothesis H0 for all these pairs can be rejected, which indicates our proposed model outperforms the compared significantly.

Table.11. Wisconsin (Diagnostic) Data Set for Breast Cancer Performance using proposed model (RFE-ANN)

Method	Accuracy (%)	Sensitivity/Specificity	Kappa	ROC/AUC	Wilcoxon
RFE-ANN	98	0.98/0.99	0.96	1	□

8. CONCLUSION

This paper puts forth a new approach to the collection of features using RFE with Random Forest as an estimator. An increasing data quality, decreasing attributes and a dataset based trained classifier guides to achieving higher performance and prediction of breast cancer as shown in Table 11.

In summary, in relation to other healthcare datasets, the proposed model is competent of successfully performing medical decision support tasks for various diseases; however, it may assemble better classification accuracy with real-world datasets. Researchers can reorganize other boosting algorithms by using different weak classifiers as the base classifier to improve the classification accuracy for medical and non-medical datasets.

REFERENCES

1. M. U. Sarwar, M. K. Hanif, R. Talib, A. Mobeen, and M. Aslam, “A survey of big data analytics in healthcare,” Int. J. Adv. Comput. Sci. Appl., vol. 8, pp. 355-359, 2017.
2. E. W. Steyerberg, Clinical prediction models. Cham: Springer International Publishing, 2019, pp. 297-308.
3. R. Ramani, N.Suthanthira Vanitha,S. Valarmathy,"The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images", IJIGSP, vol.5, no.5, pp.47-54, 2013. DOI:10.5815/ijigsp.2013.05.06. Dekker, C. (1999). Carbon Nanotubes as Molecular Quantum Wires. Physics Today, 52(5), 22.
4. Prabhjot Kaur, Yashita Pruthi, Vidushi Bhatia, Janmjay Singh,"Empirical Analysis of Cervical and Breast Cancer Prediction Systems using Classification", International Journal of Education and Management Engineering(IJEME), Vol.9, No.3, pp.1-15, 2019.DOI: 10.5815/ijeme.2019.03.01.
5. Hanahan, D., Weinberg, R.A.: Hallmarks of cancer: the next generation. Cell 2011(144), 646–674 (2011).
6. Chaudhuri A.K., Sinha D., Thyagaraj K.S. (2019) Identification of the Recurrence of Breast Cancer by Discriminant Analysis. In: Abraham A., Dutta P., Mandal J., Bhattacharya A., Dutta S. (eds) Emerging Technologies in Data Mining and Information Security. Advances in Intelligent Systems and Computing, vol 813. Springer, Singapore. https://doi.org/10.1007/978-981-13-1498-8_46
7. A. K. Chaudhuri, D. Sinha, K. Bhattacharya, and A. Das, “An Integrated Strategy for Data Mining Based on Identifying Important and Contradicting Variables for Breast Cancer Recurrence Research,” Int. J. Recent Tech. Eng., vol. 8, March 2020.

DOI: 10.17148/IJARCCE.2022.11335

8. Bhagwati Charan Patel,G. R. Sinha,"Energy and Region based Detection and Segmentation of Breast Cancer Mammographic Images", IJIGSP, vol.4, no.6, pp.44-51, 2012.
9. C.D. Katsis, I. Gkogkou, C.A. Papadopoulos, Y. Goletsis, P.V. Boufounou, G. Stylios, "Using Artificial Immune Recognition Systems in Order to Detect Early Breast Cancer", International Journal of Intelligent Systems and Applications(IJISA), vol.5, no.2, pp.34-40, 2013.DOI: 10.5815/ijisa.2013.02.04.
10. D. Tripathi, I. Manoj, G. R. Prasanth, K. Neeraja, M. K. Varma, and B. R. Reddy, "Survey on classification and feature selection approaches for disease diagnosis," in Emerging Research in Data Engineering Systems and Computer Communications, Singapore: Springer, 2020, pp. 567-576.
11. M. Tubishat, N. Idris, L. Shuib, M. A. Abushariah, and S. Mirjalili, "Improved Salp Swarm Algorithm based on opposition based learning and novel local search algorithm for feature selection," Expert Syst. Appl., vol. 145, pp. 113122, 2020.
12. S. Maldonado, J. López, A. Jimenez-Molina, and H. Lira, "Simultaneous feature selection and heterogeneity control for SVM classification: An application to mental workload assessment," Expert Syst. Appl., vol. 143, pp. 112988, 2020.
13. Avijit Kumar Chaudhuri, Dilip K. Banerjee, Anirban Das, "A Dataset Centric Feature Selection and Stacked Model to Detect Breast Cancer", International Journal of Intelligent Systems and Applications(IJISA), Vol.13, No.4, pp.24-37, 2021. DOI: 10.5815/ijisa.2021.04.03
14. T. Sridevi and A. Murugan, "A novel feature selection method for effective breast cancer diagnosis and prognosis," Int. J. Comput. Appl., vol. 88, 2014.
15. E. Aličković and A. Subasi, "Breast cancer diagnosis using GA feature selection and Rotation Forest," Neural. Comput. Appl.,vol. 28, pp. 753-763, 2017.
16. P. Hamsagayathri and P. Sampath, "Performance analysis of breast cancer classification using decision tree classifiers," Int. J. Curr. Pharm. Res., vol. 9, pp. 19-25, 2017.
17. M. Abdar, M. Zomorodi-Moghadam, X. Zhou, R. Gururajan, X. Tao, P. D. Barua, and R. Gururajan, "A new nested ensemble technique for automated diagnosis of breast cancer," Pattern Recognit. Lett., vol. 132, pp. 123-131, 2020.
18. B. Zheng, S. W. Yoon, and S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," Expert Syst. Appl., vol. 41, pp. 1476-1482, 2014.
19. M. Sewak, P. Vaidya, C. C. Chan, and Z. H. Duan, "SVM approach to breast cancer classification," in Second International Multi-Symposiums on Computer and Computational Sciences (IMSCCS 2007), IEEE, August 2007, pp. 32-37.
20. S. Y. Jin, J. K. Won, H. Lee, and H. J. Choi, "Construction of an automated screening system to predict breast cancer diagnosis and prognosis," Basic Appl. Pathol., vol. 5, pp. 15-18, 2012. A Dataset Centric Feature Selection and Stacked Model to Detect Breast Cancer 37 Copyright © 2021 MECS I.J. Intelligent Systems and Applications, 2021, 4, 24-37
21. O. I. Obaid, M. A. Mohammed, M. K. A. Ghani, A. Mostafa, and F. Taha, "Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer," Int. J. Eng. Tech., vol. 7, pp. 160-166, 2018.
22. S. Kumari and M. Arumugam, "Application of bio-inspired krill herd algorithm for breast cancer classification and diagnosis," Indian J. Sci. Technol., vol. 8, pp. 30, 2015.
23. A. Christobel and Y. Sivaprakasam, "An empirical comparison of data mining classification methods," Int. J. Comput. Inf. Syst., vol. 3, pp. 24-28, 2011.
24. D. Lavanya and D. K. U. Rani, "Analysis of feature selection with classification: Breast cancer datasets," Indian J. Comput. Sci. Eng., vol. 2, pp. 756-763, 2011.
25. A. Keleş, A. Keleş, and U. Yavuz, "Expert system based on neuro-fuzzy rules for diagnosis breast cancer," Expert Syst. Appl., vol. 38, pp. 5719-5726, 2011.
26. H. L. Chen, B. Yang, J. Liu, and D. Y. Liu, "A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis," Expert Syst. Appl., vol. 38, pp. 9014-9022, 2011.
27. G. I. Salama, M. Abdelhalim, and M. A. E. Zeid, "Breast cancer diagnosis on three different datasets using multi-classifiers," Int. J. Comput. Inform. Tech., vol. 1, pp. 36-43, September 2012.
28. D. Lavanya and K. U. Rani, "Ensemble decision tree classifier for breast cancer data," Int. J. Inf. Technol. Converg. Serv., vol.2, pp. 17-24, 2012.
29. W. Kim, K. S. Kim, J. E. Lee, D. Y. Noh, S. W. Kim, Y. S. Jung, and R. W. Park, "Development of novel breast cancer recurrence prediction model using support vector machine," J. Breast Canc., vol. 15, pp. 230-238, 2012.
30. A. K. Chaudhuri, D. Sinha, and K. S. Thyagaraj, "Identification of the recurrence of breast cancer by discriminant analysis," in Emerging technologies in data mining and information security, Singapore: Springer, 2019, pp. 519-532.

DOI: 10.17148/IJARCCE.2022.11335

31. A. Ben-David, "Comparison of classification accuracy using Cohen's Weighted Kappa," Expert Syst. Appl., vol. 34, pp. 825-832, 2008.
32. F. Wilcoxon, "Individual comparisons by ranking methods," in Breakthroughs in statistics, New York: Springer, 1992, pp.196-202.
33. F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.", Psychological Review, volume 65, issue 6 (1958)
34. Xavier Glorot et al., 2011 Deep sparse rectifier neural networks
35. Chen, Xue-wen; Jeong, Jong Cheol, "Enhanced recursive feature elimination",[IEEE Sixth International Conference on Machine Learning and Applications (ICMLA 2007) - Cincinnati, OH, USA (2007.12.13-2007.12.15)] Sixth International Conference on Machine Learning and Applications (ICMLA 2007).