Breast Cancer Diagnosis (BCD) Model Using Machine Learning

Priyanka Israni

Abstract: ***In the recent years, breast cancer research has made a significant growth however there is still a scope of advancement. Breast cancer increases the statistics of mortality among women. In concern to this issue, treatment of cancer should be started at the earlier stage, to increase the chances of survival of the patient. Thus, there is a need to diagnose breast cancer at the early stage using the features from the mammograms. This paper proposes an efficient BCD model to detect breast cancer by using Support Vector Machine (SVM) with 10-fold cross validation. The complexity of the problem increases if there are many input features for the diagnosis of cancer. Thus, Principal Component Analysis (PCA) is used to reduce the feature space from a higher dimension to a lower dimension. Experiment result shows that the PCA increases the accuracy of the model. The proposed BCD model is compared with other supervised learning algorithms like Decision trees (DT), Random Forest, k- Nearest Neighbors(k-NN), Stochastic Gradient Descent (SGD), AdaBoost, Neural Network (NN), and Naïve Bayes. Evaluation parameters like F1 measure, ROC curve, Accuracy, Lift curve and Calibration Plot proves that proposed BCD model outperforms and gives the highest accuracy among other compared algorithms.***

Index Terms: ***Accuracy, Breast cancer, Classification, Features, Machine learning.***

I. INTRODUCTION

In the 21st century, the deadliest disease is considered to be cancer, which results in maximum mortality rate, especially among women. According to the report in [23] by the Times of India newspaper in 2017 "India has the third highest cancer patients among women". Most common cancer among women is breast cancer. The statistics of mortality due to cancer can be reduced by detecting cancer at an early stage and getting proper treatment for the same.

Breast Cancer is the abnormal growth of the old cells in the breast tissue that results in the cyst. Early detection of cancer can be easily analyzed and treated as only a few cells have undergone the process of abnormal growth. At an initial stage of cancer, it is more likely to be treated successfully. Hence, cancer should be detected as early as possible before it spreads into the various areas of the breast. The cancer research done by the United Kingdom in 2015 [22] has given the relationship between the survival rate of the women and the stage of the breast cancer at diagnosis. The survival rate decrease as the stage is increased. Thus, the appropriate test and various methods should be introduced to detect breast cancer.

Machine Learning and Data Mining methods are most commonly used for the early detection of the disease.

The classification algorithms along with preprocessing, feature selection and feature extraction produce promising results for cancer detection. Several experiments are done on the breast cancer data that includes combination of classification algorithms, genetic algorithm for optimization, feature extraction and selection etc.

In Machine Learning, the machine is trained by different learning types. The Fig 1 shows the hierarchy of different types of learning. Supervised learning is categorized as classification and regression. Classification is the process of predicting the data into discrete classes (i.e. prediction of email is spam or not). Regression is the supervised learning method that predicts the continuous values (i.e. Weather forecasting). \ The aim of the proposed work is to give the accurate classifier for diagnosis of breast cancer by comparing various algorithms of machine learning and data mining. The objectives of this work are stated as (1) Study of various Machine Learning algorithms and finding the appropriate algorithm for medical data. (2) Study of the dataset used for diagnosis and its features. (3) Preprocessing the data. (4) Implementation of algorithms on the dataset. (5) Dimensionality reduction by using Feature Extraction and Feature Selection techniques on the dataset. (6) Implementation of algorithms with new generated features. (7) Evaluating Performance measures. (8) Comparison of algorithms. Experiments are performed on the open source toolkit Orange tool (3.13) released under GPL.

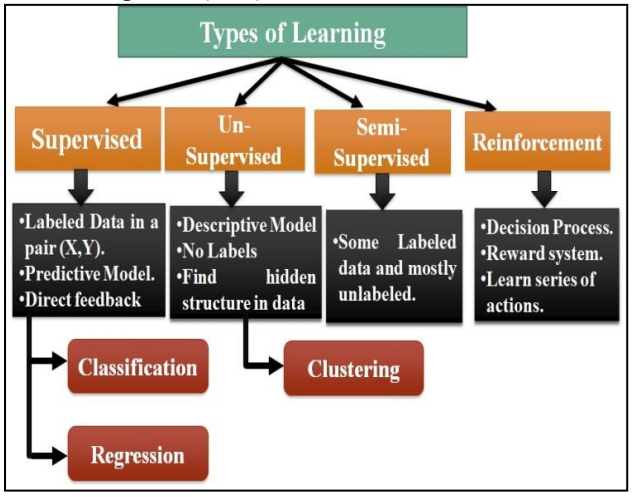

Fig. 1. Types of Learning

The paper is organized in various sections. The description about the dataset used for implementation of the proposed work is depicted in section II. In section III, literature survey for the proposed work is presented. Section IV gives the details of the proposed work with the description of machine learning techniques which are implemented. Section V shows the results of the experiments and the various graphs for comparison. Section VI

Revised Manuscript Received on August 05, 2019.

Priyanka Israni is an Assistant Professor in Computer Engineering Department in G H Patel College of Engineering & Technology.

gives the comparison of proposed BCD model with the other implemented models on the same dataset. Finally, conclusion of the research is presented.

II. DATASET DESCRIPTION

The dataset used for the implementation of the proposed work is Wisconsin Breast Cancer (Original) dataset. It is freely available dataset in the University of California, Irvine (UCI) Machine Learning Repository [21].

Table I. Description of the Breast Cancer Wisconsin Dataset.

Sr. No.	Attribute	Description	Domain
1	Sample code number	The unique id of the observation.	Unique id.
2	Clump Thickness	Cells that are grouped in monolayers are identified as benign cells, while the cells that are grouped in multilayer are identified as malignant i.e. cancerous cells.	1-10
3	Uniformity of Cell Size	The size of the cells differs from malignant cells to benign cells.	1-10
4	Uniformity of Cell Shape	The shape of the cells differs from malignant cells to benign cells.	1-10
5	Marginal adhesion	Normal cells are compact in nature while cancerous cells are loose in nature.	1-10
6	Single Epithelial Cell Size	Epithelial cells that are distended and quite large may be a cancerous cell.	1-10
7	Bare Nuclei	Nuclei in the cells that is not surrounded by cytoplasm is known as naked or bare nuclei. This is found in benign tumors.	1-10
8	Bland Chromatin	In benign cells, a fine and uniform texture of the nucleus is found while chromatin is coarser in cancer cells.	1-10
9	Normal Nucleoli	Generally, cells contain nucleoli which is hardly visible but in cancerous cells, the nucleoli are prominent.	1-10
10	Mitoses	Division process of normal cells varies from the cancerous cells.	1-10
11	Class	Response Attribute	It has two values, 2 and 4. 2 represents benign and 4 represents malignant

This dataset contains 699 instances, 10 input variables i.e. features and 1 output variable i.e. response. Description of the all the features and response is given in table 1. Dataset consists of 458 (65.5 %) instances of the Benign tumor and 241(34.5%) instances of the Malignant tumor. There are 16 instances in the dataset and each contains a single missing value. For simplicity, in the 'class' response attribute, 2 is replaced by 'benign' and 4 is replaced by 'malignant'.

III. LITERATURE SURVEY

Many Researchers have used machine learning and data mining approaches to classify the breast cancer wisconsin dataset available at UCI machine learning repository.

J. Ivancáková et al. [8] have compared many machine learning techniques for the breast cancer wisconsin dataset. They performed binary classification by using SVM, Naïve Bayes, Random Forests, C4.5, neural networks and k-NN. Authors have divided the dataset into three ratios for the training and testing sample i.e. 80:20, 70:30 and 60:40 to train the model and proved that the random forest and SVM with 10-fold cross validation gives the highest performance. A. F. Seddik and D. M. Shawky [16] have used logistic regression along with dimensionality reduction to classify malignant and benign cases on Wisconsin Diagnostic Breast Cancer (WDBC) dataset. Borges and L. Rodrigues [5] have used J48 (DT) and Bayesian Networks on the same dataset after discretizing the data and proved that the Bayesian Networks gives the highest accuracy than J48. G. D. Rashmi et al. [13] have used Naïve Bayes for classification and prediction of the same dataset. G. I. Salama et al. [15] have implemented Naïve Bayes, Multi-Layer Perceptron (MLP), SVM, k-NN and J48 on the same dataset. Authors also implemented fusions of 2,3 and 4 classification methods and proved that the fusion of fusion of J48 (DT), MLP, Sequential Minimal Optimization (SMO), and Instance Based for k-NN (IBK) is giving higher performance than the other compared classifiers. H. Asria et al. [4] have implemented SVM, Decision Tree (C4.5), NB and k-NN on the Wisconsin Breast Cancer (original) dataset and proved that the performance of SVM is highest among the models. A. F. Agarap [7] have implemented SVM, GRU-SVM, Linear Regression, MLP, Nearest Neighbor (NN) search and Softmax Regression on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset and proved MLP outperforms as compared to other models.

L. G. Ahmad et al. [2] have implemented SVM, Decision Tree (C4.5) and Artificial Neural Network (ANN) on the records of Iranian Center for Breast Cancer which contains 1189 records. Authors used 10-fold cross validation for modeling a general classifier for unbiased accuracy prediction and proved that the SVM classification predicts with the highest accuracy among the three. M. Montazeri et al. [12] have implemented 1-Nearest Neighbor (1NN), RBF Network (RBFN), NB, Random Forest, MLP machine learning, SVM and AdaBoost (AD) techniques with 10-fold cross validation for generalized model on the records of 900 patients. Authors through experimental results proved that the Random Forest model is superior than other compared models.

H. Jouni et al. [9] proposed an architecture of optimal artificial neural networks that classifies breast cancer through pattern recognition. Authors focus was to reduce classification error by finding the optimal activation function with fewer blocks. They proved that the hyperbolic tangent and log-sigmoid function without biases are the optimal solutions for their dataset. They have done the implementation by CMOS technology.

Many Researchers have extracted features from the images of the left and right breast for the detection of cancer. A. Lashkari et al. [10] have used imaging technique based on thermography for diagnosing breast cancer based on degrees. The features extracted and selected were classified using supervised learning algorithms like SVM, k-NN, AdaBoost probability neural network and Naïve Bayes. They proved that the best mean accuracy is achieved by

using 0-degree image with a combination of AdaBoost and mRMR and for 3 degrees image with a combination of AdaBoost and Genetic Algorithm. M. Li and Z. Zhou [11] proposed Co-Forest, a new semi-supervised learning algorithm for microcalcification detection to diagnose breast cancer diagnosis. It extends the co-training of ensemble method Random Forest. They also proved that testing data is good in building Computer Aided Design systems and Co-Forest is able to increase the accuracy of the model that is trained on a few training instances by using the testing data.

In spite of the great achievements of all state of the art machine learning techniques for diagnosis of breast cancer, some drawbacks are still to be resolved. While training the data to model the classifier, generalized classifier should be created to remove the biased prediction. Hence, in this paper 10-fold cross validation technique is used to model the generalized classifier. Many datasets contain large number of features which makes the model difficult to train. Thus, in this paper Principal Component Analysis for dimensionality reduction is integrated along with Support Vector Machine which reduces computational complexity and also prevents overfitting of the data.

IV. PROPOSED MODEL

As the dataset consists of 10 input attributes and one response variable which is discrete in nature, thus proposed BCD model uses different supervised learning classification algorithms for training the data and finding the accurate classifier for the dataset. Figure 2 represents the block diagram of the BCD model.

A. Sampling

Before training the data to model a classifier, the dataset is partitioned into two different subsets, for training and testing. 70 % of data is used for training the model to build a classifier and 30% for testing on the modeled classifier for prediction.

B. Preprocessing

The instances in the dataset contain some missing values, noisy data and inconsistencies that need to be cleaned before actually processing the data for training. The preprocessing of the data will perform data cleaning i.e. to fill the missing values of the instances, smoothing noisy data and resolving inconsistencies.

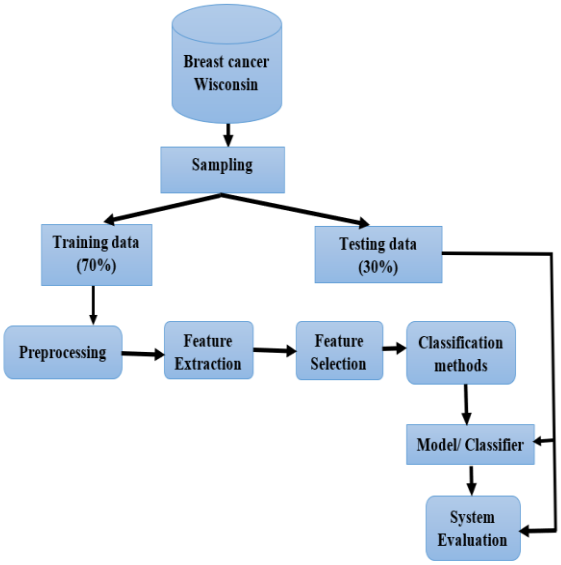

Fig. 2. Block Diagram for BCD model

C. FEATURE EXTRACTION AND FEATURE SELECTION

The preprocessed data contains some features which are not required or are not as important as the other features of the dataset. Then, such features need to be removed for the sake of curse of dimensionality. In machine learning, "dimensionality" refers to the number of features (i.e. input variables) in your dataset. The curve of dimensionality states that "when the number of features is very large relative to the number of observations i.e. instances in your dataset, certain algorithms struggle to train effective models". Feature Extraction and Feature Selection are the techniques for dimensionality reduction which can also avoid overfitting problem in training the data on the model.

Feature extraction is the technique for transforming the existing features into the new set of features. Here the new set of features is the function of the existing features.

Feature Selection is the technique for selecting the subset of existing features with any transformation. Fig.3a. and Fig. 3b, shows Feature Selection and Feature Extraction respectively.

The proposed model uses **Principal Component Analysis (PCA)** for feature extraction which is computationally competent for the reduction of features i.e. dimensionality reduction. PCA converts the data from higher dimension to lower dimension by orthogonal linear transformation. Thus, the first principal component i.e. first dimension is the one which gives highest variance among the data, the second principal component is the second highest variance among the data and so on.

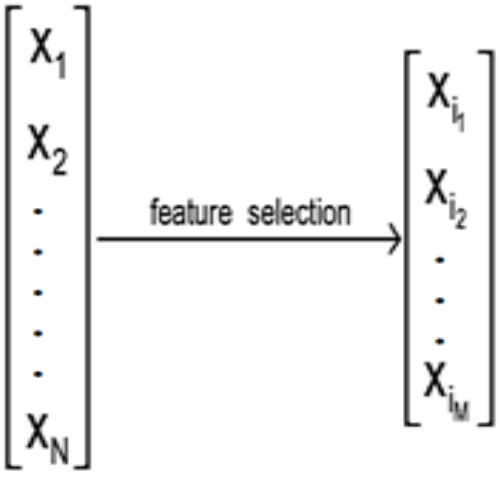

Fig.3a. Feature Selection

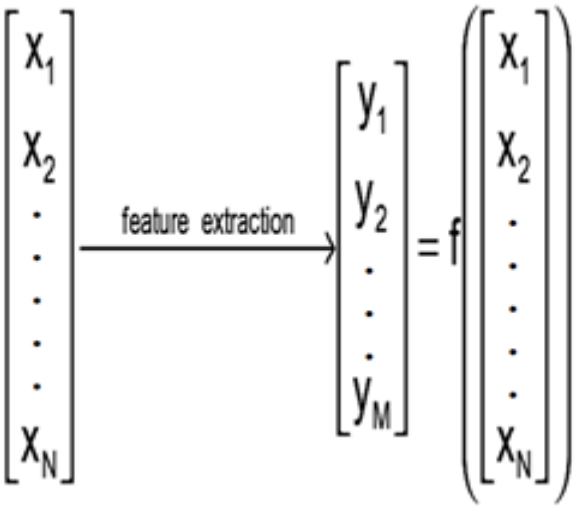

Fig.3b. Feature Extraction

This conversion is defined by the s-dimensional vectors of weight $\mathbf{w_{(k)}} = \{w_1, w_2, \ldots\ldots, w_s\}_{(k)}$ that map each instance vector $\mathbf{x_{(i)}}$ of $\mathbf{X}=\{x_1,x_2, \ldots, x_n\}$ to a new vector of principal component $\mathbf{C_{(i)}}=\{C_1,C_2,\ldots, C_m\}_{(i)}$ is given as in Equation (1).

$$C_{k(i)} = x_i \cdot w_k \qquad (1)$$

Here, i=1, 2, …, n and k=1,2, …. , m

The conversion is done in such a way that the variables $\mathbf{C_1, C_2, \ldots, C_m}$ of **C** considered over the data increasingly receive the highest variance from **x**, with each vector **w** which is a unit vector.

D. CLASSIFICATION METHODS

To train the data for a model, SVM along with 10-fold cross validation is used on the extracted features from the PCA. K-fold cross validation is the validation technique which is used for generalizing the model on the given dataset. It is used to analyze the predictive nature of the model i.e. how accurately the model will predict the unknown data. It is the solution to the problem of overfitting that rises due to many features or having noisy data. The first round of cross validation will partition the given data into k subsets and use k-1 subsets for training the data to build a model and one for validating i.e. testing the model. This process will be repeated for k times and the performance of all the k rounds are combined to give the estimate of the predictive performance of the model.

Support Vector Machines are the supervised learning method which is used to analyze the data for classification as well as regression. For the binary classification problem, SVM trains the model such that the data points belonging to one class are separated from the other class in the feature space by the clear margin, which is wide enough to detect the two different classes clearly. SVM maximizes the margin width to generalize the classifier and also to detect the outliers. The maximum margin separator is decided by the data points in the feature space and these data points are known as support vectors. Support vectors are used to decide the belongingness of the test data in either class.

The given training dataset of points are of the form $(x_1,y_1), (x_2,y_2), \ldots\ldots, (x_n, y_n)$. where y_i is either 1 or -1 which is the response attribute in which x_i i.e. input attribute belongs. x_i is the m- dimensional real vector. If the dataset is linearly separable, then two hyperplanes parallel to each other will separate the data such that the distance between the two, should be greater. The region between these two hyperplanes is known as margin. As shown in Figure 4 the margin hyperplane divides purple points into class 1 and orange points in class -1. This margin hyperplane is the separator which divides the two classes. This hyperplane needs to be maximized such that the distance between the hyperplane and the nearest point x_i belonging to either group is maximized.

Mathematically, the equation of the hyperplane can be written as the set of points x which satisfies the Equation (2).

$$\mathbf{w}.x - b = 0 \qquad (2)$$

Equation (4) will be zero, if the constraint in Equation (3) is satisfied i.e. *x* lies on the correct side of the margin in which it actually belongs else the function value is proportional to the distance from the margin.

$$\text{Minimize } \left[\frac{1}{n}\sum_{i=1}^{n} \max\left(0, 1 - y_i(w.x_i - b)\right)\right] + \lambda \|w\|$$

Here, λ is the trade-off between increasing size of margin and x_i lies on the proper side of the margin where it actually belongs.

$$w.x_i - b \geq 1 \text{ if } y_i = 1$$
$$w.x_i - b \leq -1 \text{ if } y_i = -1$$

These above equations can be rewritten as

$$y_i(w.x_i - b) \geq 1 \quad for\ all\ 1 \leq i \leq n \qquad (3)$$

If the data is not linear and cannot be separated by linear classifier, then hinge loss function is introduced to support non-linear data in SVM as in Equation (4).

$$\max\left(0, 1 - y_i\left(\mathbf{w}.x_i - b\right)\right) \qquad (4)$$

Equation (4) will be zero, if the constraint in Equation (3) is satisfied i.e. *x* lies on the correct side of the margin in which it actually belongs else the function value is proportional to the distance from the margin.

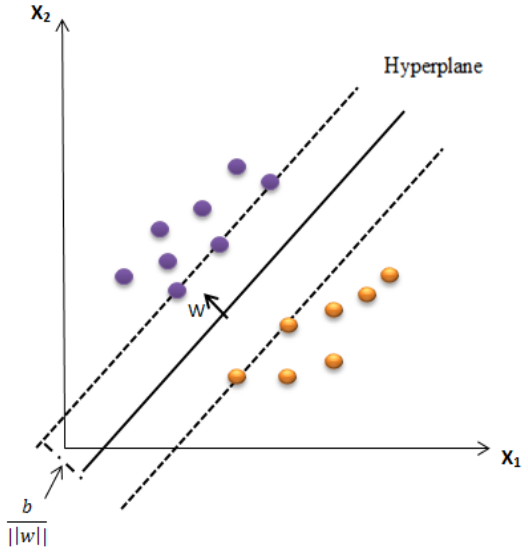

Fig.4. Support Vector Machine

Minimize (5)

$$\left[\frac{1}{n}\sum_{i=1}^{n} \max\left(0, 1 - y_i(w.x_i - b)\right)\right] + \lambda \frac{}{\|w\|}$$

Here, λ is the trade-off between increasing size of margin and x_i lies on the proper side of the margin where it actually belongs.

E. MODEL/CLASSIFIER

After providing training, the model or classifier is ready to predict the new instances. The test data is given as an input to the model or classifier to predict the class and accuracy of various models is compared to find the best-trained model for the given dataset.

F. SYSTEM EVALUATION

The evaluation of the BCD model is done by using various classification performance measures like confusion matrix, recall, precision, f1 measure, Area under Curve (AUC), classification accuracy (CA), ROC curve, Lift curve and Calibration plot.

a. Confusion Matrix

The confusion matrix for the binary classification and the terms related to it is shown in Figure 5. For example, total instances of patients available for examination are 165. Out of 165, 105 patients have a disease while 60 does not have the disease. When these 165 instances were examined on the classifier, classifier predicted 100 patients have disease while 65 don't have the disease.

- True Positive (TP): The instances which are predicted as true by the classifier and are actually as true.
- True Negative (TN): The instances which are predicted as false by the classifier and are actually as false.
- False Positive (FP): The instances which are predicted as true by the classifier but are actually False.
- False Negative (FN): The instances which are predicted as false by the classifier but are actually true.

$$\text{Classification Accuracy (CA)} = \frac{TP+TN}{TP+FP+TN+FN} \quad (6)$$

$$\text{Recall (True Positive Rate)} = \frac{TP}{TP+FN} \quad (7)$$

$$\text{Precision (Positive Predictive value)} = \frac{TP}{TP+FP} \quad (8)$$

$$\text{F1 score} = \frac{2*Precision*Recall}{Precision+Recall} \quad (9)$$

	Predicted: Yes	Predicted: No	Total of Actual instances
Actual: Yes	TP=95	FN=10	105
Actual: No	FP=5	TN=55	60
Total of Predicted instances	100	65	

Fig.5. Confusion Matrix

b. ROC Curve

Receiver Operating Characteristic curve is the performance measure used to examine the ability of classifiers by the variation of its discrimination threshold. A ROC curve is a graph which plots the false positive rate on the X-axis and the true positive rate on the Y-axis. At point (0,1) for all classifies, all the instances are correctly classified in their respective group, thus it is the perfect classifier. In most cases, a classifier has a changing parameter that can be varied such that the increase in TP will result in increase in FP or decrease in FP will result in decrease in TP. By the changing the value of parameter, we get a point (FP, TP) and the series of such points which are used to plot the ROC curve. A classifier which doesn't have changing parameter is represented by a single point (FP, TP), which is the ROC point.

In Binary Classification, instances are predicted into the classes by using some continuous random variable X, which is the value computed for the given observation i.e. instance like estimated probability. If T is the threshold parameter, then if X > T, the observation is classified as positive, otherwise negative. The probability density function for X is $f_1(x)$ if the observation is actually positive and $f_2(x)$ otherwise.

$$TPR(T) = \int_T^{\infty} f_1(x)\, dx \quad (10)$$

$$FPR(T) = \int_T^{\infty} f_2(x) dx \quad (11)$$

Thus, by changing the value of T, the ROC curve plots (FP, TP) pair i.e. True Positive Rate i.e. TPR(T) v/s False Positive Rate i.e. FPR(T). Figure 6 shows the ROC plot with varying threshold T. The diagonal in the Figure 6 shows the behavior of a random classifier.

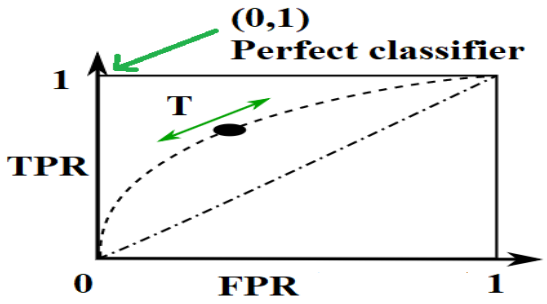

Fig.6. ROC Curve

c. Area under Curve (AUC)

AUC is the probability that the classifier while choosing a random observation, gives higher preference to the positive observation than a negative observation when the features used are normalized. Thus, the value of AUC for the perfect classifier is 1.

$$AUC = P(X_1 > X_2) \quad (12)$$

where X_1 is the value for the positive observation and X_2 is the value for the negative observation.

d. Lift curve

The Lift curve shows the relationship between, the number of observation which are positively predicted and actually positive observations. Thus, the performance of the chosen classifier is measured against a random classifier. The graph is constructed with the actually positive cases (in descending order of probability) on the X-axis and the number of positively predicted cases i.e. true positives on the Y-axis.

e. Calibration Plot

The Calibration Plot gives the relationship between the class probabilities against the predicted class by the classifiers. The diagonal of the graph represents optimal behavior of the classifier. Thus, the prediction probabilities are more accurate if the classifier's curve is closer.

V. EXPERIMENTS

The experiments are done on the supervised learning algorithms i.e. Decision Tree, AdaBoost, k-NN, SVM, Stochastic Gradient Descent (SGD), Neural Network, Naïve Bayes and Random Forest. Table 2 shows the comparison of evaluation parameters while training and testing the model without dimensionality reduction i.e. PCA. Table 3 shows the comparison of evaluation parameters while training and testing the model with dimensionality reduction i.e. PCA. These values are the average of the 10 observed values. Figure 7 shows the graphical comparison of all the models with PCA.

Table II. Comparison of Evaluation parameters for training and testing data without PCA

Method	**Training**					**Testing**				
	AUC	**CA**	**F1**	**Precision**	**Recall**	**AUC**	**CA**	**F1**	**Precision**	**Recall**
kNN	0.987	0.965	0.965	0.965	0.965	0.992	0.971	0.971	0.972	0.971
Tree	0.93	0.933	0.932	0.932	0.933	0.962	0.943	0.942	0.943	0.943
SVM	**0.993**	**0.963**	**0.963**	**0.964**	**0.963**	**0.995**	**0.971**	**0.971**	**0.971**	**0.971**
SGD	0.966	0.967	0.967	0.968	0.967	0.967	0.971	0.971	0.971	0.971
Random Forest	0.982	0.967	0.967	0.968	0.967	0.997	0.971	0.971	0.971	0.971
Neural Network	0.993	0.965	0.965	0.966	0.965	0.999	0.971	0.971	0.972	0.971
Naive Bayes	0.986	0.965	0.966	0.967	0.965	0.999	0.986	0.986	0.986	0.986
AdaBoost	0.935	0.949	0.949	0.949	0.949	0.929	0.938	0.937	0.938	0.938

Table III. Comparison of Evaluation parameters for training and testing data with PCA

Method	**Training**					**Testing**				
	AUC	**CA**	**F1**	**Precision**	**Recall**	**AUC**	**CA**	**F1**	**Precision**	**Recall**
kNN	0.985	0.955	0.955	0.955	0.955	0.986	0.962	0.962	0.962	0.962
Tree	0.979	0.961	0.961	0.961	0.961	0.968	0.947	0.947	0.948	0.947
SVM	**0.994**	**0.965**	**0.965**	**0.965**	**0.965**	**0.995**	**0.981**	**0.981**	**0.981**	**0.981**
SGD	0.958	0.961	0.961	0.961	0.961	0.972	0.971	0.971	0.972	0.971
Random Forest	0.986	0.961	0.961	0.961	0.961	0.992	0.967	0.967	0.967	0.967
Neural Network	0.994	0.957	0.957	0.957	0.957	0.992	0.967	0.967	0.967	0.967
Naive Bayes	0.983	0.943	0.943	0.944	0.943	0.978	0.933	0.933	0.933	0.933
AdaBoost	0.945	0.951	0.951	0.951	0.951	0.946	0.952	0.952	0.952	0.952

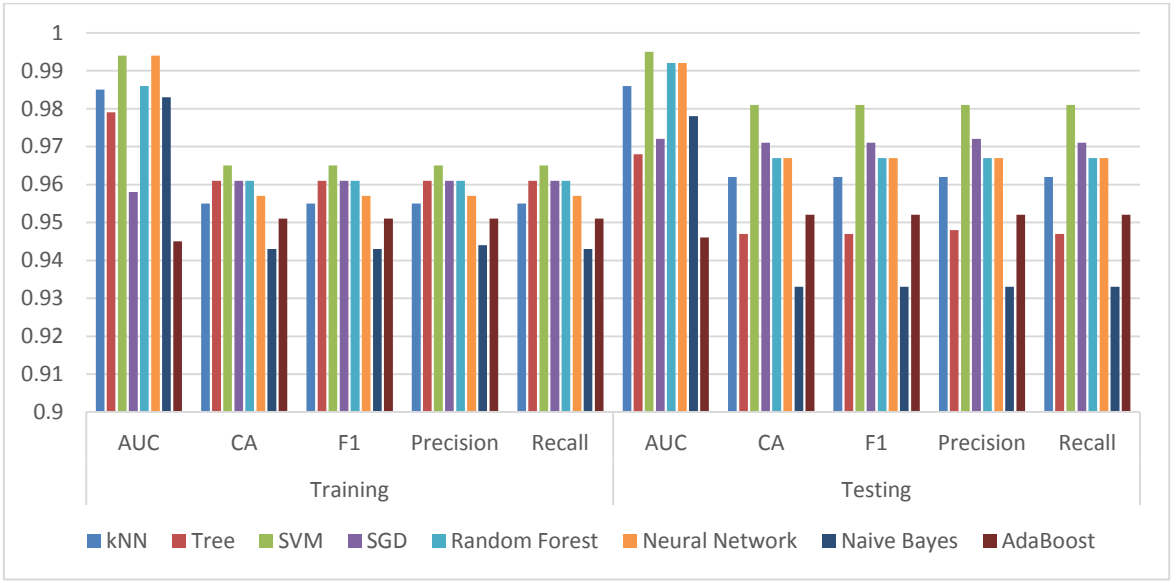

Fig.7. Graphical Comparison of all models

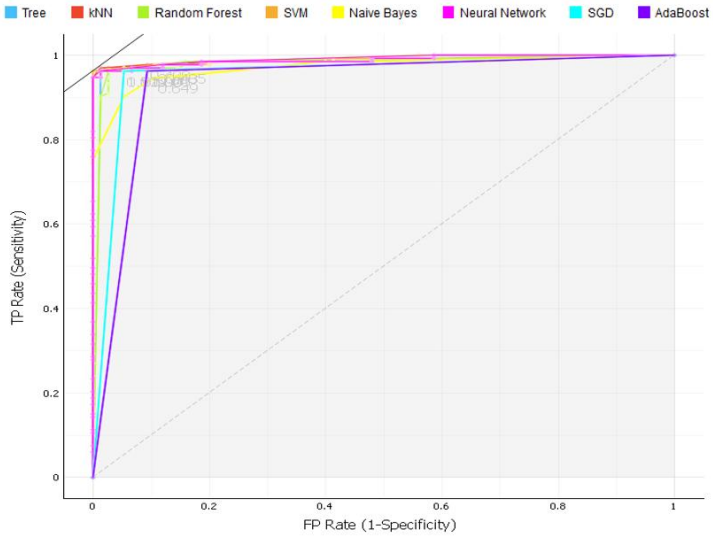

Fig.8. ROC curve of all models.

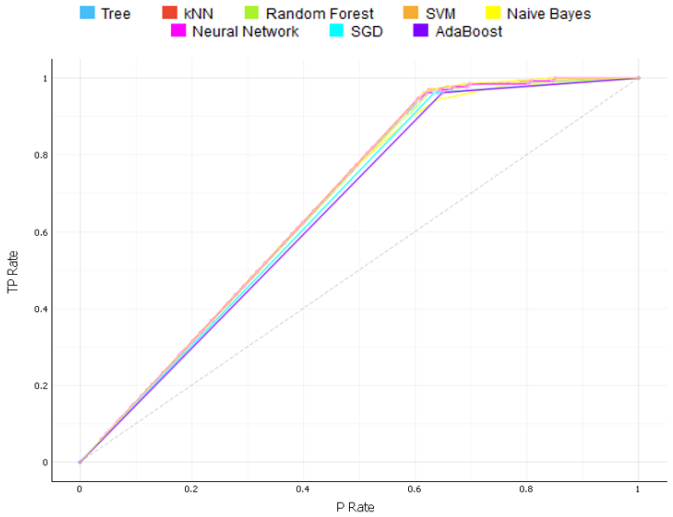

Fig.9. Lift curve of all models.

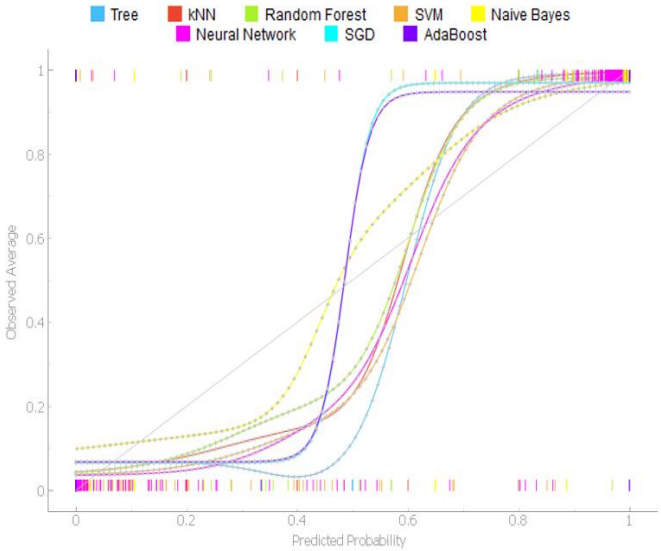

Fig.10. Calibration Plot of all models

Figure 8 shows the ROC curve of all the compared models along with BCD model. From the Figure 8 it is clear that SVM with PCA graph's point is nearer to the point (0,1), which indicates that the proposed model is giving highest accuracy than other compared models. It has the highest area under the curve of 0.995.

Figure 9 shows the lift curve for all the models. Lift curve gives the performance of the model compared to the random guess. The dotted line shown in the graph is the random classifier.

Figure 10 shows the calibration plot for all the compared models along with BCD model. Calibration Plot tests the prediction capacity of the model. The dotted line in the graph is the optimal classifier. Thus, the proposed model is overall close to the optimal classifier than the other models. Hence the prediction capacity of the model is good as compared to other models.

VI. RESULTS AND COMPARISON

Figure 7 depicts that BCD Model (SVM with PCA) outperforms other supervised learning algorithms. BCD model gives the highest F1 measure of 98.1 % and AUC as 0.995. Table 5 gives the comparison of the proposed BCD model with the other model implemented to diagnose breast cancer. The table shows the BCD model is giving the highest classification accuracy than other models.

Table 5. Comparison of BCD model with other models

Author	Model	Accuracy	AUC
Proposed Method	**SVM with 10-fold Cross Validation and PCA**	**98.1 %**	**0.995**
Yang & Xu [22]	SVM with PCA and DE passed parameter tuning	97.72%	0.993
Borges [7]	Bayesian Networks	97.80%	-
Salama et al. [9]	Fusion of SMO, J48, MLP and IBK	77.32%	-
Ahmad et al. [12]	SVM with 10-fold Cross-Validation	95.7%	-
Asria et al. [10]	SVM with 10-fold Cross-Validation	97.13%	-
Ch. Shravya et al.[20]	SVM	92.78%	-

VII. CONCLUSION

The proposed model is efficient to diagnose the breast cancer at the early stage. The proposed BCD model uses SVM along with 10-fold cross validation for generalizing classifier's performance and also to overcome the problem of overfitting of data. Principal Component Analysis is used for Feature Extraction. The experimental results prove that the proposed model outpaces the other models which were developed to detect breast cancer using the Wisconsin Breast Cancer (Original) dataset. BCD model is compared with other implemented models like AdaBoost, Random Forest, Naïve Bayes etc. by using various evaluation parameters like F1 measure, class Accuracy, ROC curve, AUC, Lift curve and Calibration plot. The accuracy of the proposed BCD model is 98.1 % and AUC is 0.995, which are the highest among the other implemented models.

REFERENCES

1. B. M. Abed, B.M., Shaker, K., Jalab, H.A., Shaker, H., Mansoor, A.M., Alwan, A.F. and Al-Gburi, "A hybrid classification algorithm approach for breast cancer diagnosis," in Proceedings of IEEE Industrial Electronics and Applications Conference (IEACon 2016), Malaysia, pp. 269-274, 2016.
2. L. G. Ahmad, A. T. Eshlaghy, A. Poorebrahimi, M. Ebrahimi and A. R. Razavi, "Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence," Journal of Health & Medical Informatics, vol. 4, no. 2, pp.1-3, 2013.
3. M. R. Al-Hadidi, A. Alarabeyyat and M. Alhanahnah, "Breast Cancer Detection Using K-Nearest Neighbor Machine Learning Algorithm," in Proceedings of the 9th International Conference on Developments in e Systems Engineering (DeSE), 2016.
4. H. Asria, H. Mousannif, H. Moatassimec and T. Noeld, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis," in Proceedings of the 6th International Symposium on Frontiers in Ambient and Mobile Systems, pp. 1064 – 1069, 2016.
5. Borges and Lucas Rodrigues, "Analysis of the Wisconsin Breast Cancer Dataset and Machine Learning for Breast Cancer Detection," in Proceedings of the XI Workshop de Visão Computacional, 2015.
6. P. Drotár and Z. Smékal, "Comparative study of machine learning techniques for supervised classification of biomedical data," in Proceedings of the Acta Electrotechnica et Informatica, vol. 14, no. 3, pp. 5-10, 2014.
7. A. Fred and M. Agarap, "On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset," in Proceedings of the 2nd International Conference on Machine Learning and Soft Computing, 2018
8. J. Ivancáková, F. Babic and P. Butka, "Comparison of Different Machine Learning Methods on Wisconsin Dataset," in Proceedings of the 16th IEEE World Symposium on Applied Machine Intelligence and Informatics, pp.173-178, 2018.
9. H. Jouni, M. Issa, A. Harb, G. Jacquemod and Y. Leduc, "Neural Network Architecture for Breast Cancer Detection and Classification," in Proceedings of the IEEE International Multidisciplinary Conference on Engineering Technology (IMCET),2016.
10. A. Lashkari, F. Pak, M. Firouzmand "Breast thermal images classification using optimal feature selectors and classifiers," The Journal of Engineering, vol. 2016, no. 7, pp. 237–248, 2016.
11. M. Li and Z. Zhou, "Improve Computer-Aided Diagnosis with Machine Learning Techniques Using Undiagnosed Samples," IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 37, no. 6, 2007.
12. M. Montazeri, M. Montazeri, M. Montazeri and A. Beigzadeh, "Machine learning models in breast cancer survival prediction," Technology and health care, vol. 24, no. 1, pp. 31-42, 2016.
13. G. D. Rashmi, A. Lekha and N. Bawane, "Analysis of Efficiency of Classification and Prediction Algorithms (Naïve Bayes) for Breast Cancer Dataset," in Proceedings of the IEEE International Conference on Emerging Research in Electronics, Computer Science and Technology, pp.108-113, 2015.
14. A. A. Raweh, M. Nassef and A. Badr, "A Hybridized Feature Selection and Extraction Approach for Enhancing Cancer Prediction Based on DNA Methylation," IEEE Access, vol. 6, pp. 15212 – 15223, 2018.
15. G. I. Salama, M. B. Abdelhalim, and M. A. Zeid, "Experimental Comparison of Classifiers for Breast Cancer Diagnosis," Proceedings of the 7th International IEEE conference on Computer Engineering & Systems, pp.180-185, 2012.
16. A. F. Seddik and D. M. Shawky, "Logistic Regression Model for Breast Cancer Automatic Diagnosis," in Proceedings of the SAI Intelligent Systems Conference, pp. 150-154, 2015.
17. L. Yang and Z. Xu, "Feature extraction by PCA and diagnosis of breast tumors using SVM with DE-based parameter tuning," International Journal of Machine Learning & Cybernetics, pp. 1-11, 2017.
18. W. Yue, Z. Wang, H. Chen, A. Paynand and X. Liu, "Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis", Journal of Engineering Design, vol. 2, no. 13, pp 1-17, 2018.
19. W. Zhanga and J. Chenb, "Relief Feature Selection and Parameter Optimization for Support Vector Machine based on Mixed Kernel Function," International Journal of Performability Engineering, vol. 14, no. 2, pp. 280-289, 2018.
20. Ch. Shravya, K. Pravalika and Shaik Subhani, "Prediction of Breast Cancer Using Supervised Machine Learning Techniques", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol.8, No.6, pp.1106-1110, 2019.
21. Fawcett, Tom, "An Introduction to ROC Analysis", Pattern Recognition Letter , Vol.27 , No.8, pp. 861–874, 2006.
22. "Breast Cancer Wisconsin Dataset", Available at: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+Original
23. "Cancer Research UK. Breast cancer survival statistics", Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/survival#heading-Three
24. "India now has 3rd highest number of cancer cases among women", Available at: https://timesofindia.indiatimes.com/india/india-now-has-3rd-highest-number-of-cancer-cases-among-women/articleshow/60812041.cm
25. Ting, Kai Ming (2011). Encyclopedia of machine learning. Springer. ISBN 978-0-387-30164-8.

AUTHORS PROFILE

Priyanka Israni is an Assistant Professor in Computer Engineering Department in G H Patel College of Engineering & Technology. Her research area includes Machine Learning, Big Data, Machine Vision and Healthcare.