EASR

Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, KhonKaen University, Thailand

Breast cancer data classification using ensemble machine learning

Meerja Akhil Jabbar*

Department of AI&ML, Vardhaman College of Engineering, Hyderabad, Telangana 501218, India

Received 14 January 2020
Revised 25 May 2020
Accepted 23 June 2020

Abstract

Breast cancer (BC) is the largest cause of death in women. Accurate classification of breast cancer data is important in cancer diagnosis and classification of Malignant and Benign tumors can prevent patients to take unnecessary tests. Breast cancer classification can also be used to determine suitable treatment. Classification of Benign and Malignant patient groups is widely recognized research in the medical field. Due to the advantage of detecting critical features from a medical data set, machine learning is widely used in Breast cancer Prediction. Recently there has been greater attention to the use of machine learning methods in medical diagnosis. These decision support systems are effective and helpful for medical experts in the healthcare domain. The objective of this work is to address the problem of the classification of breast cancer data using ensemble learning. Ensemble learning techniques are used to improve the performance of a classifier. This paper deals with building a decision support system using the ensemble model built with Bayesian network and Radial Basis Function. In this work, extensive experiments were carried out on the much- studied open access data set "Wisconsin Breast Cancer Data set (WBCD)". The data set is partitioned into training and testing. Various metrics like accuracy, sensitivity, specificity, positive predictive value, negative predicted value, Error rate, false-positive rate, Mathew's correlation coefficient were used to measure the performance of the model. Experimental results show that the proposed method records a remarkable accuracy of 97% to classify breast cancer data and outperformed the existing approaches. The proposed ensemble learning would be viable in helping cancer specialists in recognizing cancer tumors accurately and help the patients in taking the correct treatment.

Keywords: Breast cancer, Ensemble learning, Machine learning, Bayesian network, Radial basis function, Classification, Wisconsin breast cancer data set, Accuracy

1. Introduction

Breast cancer is one of the most dangerous reproductive cancers that affect mostly women worldwide. First Breast cancer case was recorded in Egypt in 3000 BC [1] Out of 8 female, one will get caught this cancer [2]. Cancers are abnormal cells that divide uncontrollably and are able to invade other tissues. Breast tumor is an abnormal growth of tissues in the breast, and it may be felt as a nipple or discharge or it may be change of skin texture around the nipple. Breast Cancer always had high mortality and incidence rate. According to the latest survey breast cancer is expected an account of 25% of all cancer diagnosis and is 15% of all deaths among women [3-5]. Early diagnosis of cancer is very necessary. Many automated classification methods for breast cancer has been developed Due to the efforts of researchers, in early diagnosis of Breast cancer, the mortality rate is declined over the years [6]. Soft computing techniques are widely used for cancer disease diagnosis due to their good performance in classification.

1.1 Signs and symptoms of breast cancer

Breast cancer is of two types 1) Invasive 2) Non Invasive. In invasive cancer cells spread over other parts of the body, whereas in second type of cancer, cancer cells remain in one location of the breast and will not spread to other parts. There are many different signs and symptoms of breast cancer. Few of them are

1) Change in appearance or shape and the size of a Breast
2) Dimpling of Breast
3) A newly inverted nipple
4) Redness over the breast
5) Scaling, Peeling, or flaking of the pigmented area of breast skin
6) A breast lump [7]

1.2 Risk factors of breast cancer

Risk factors for breast cancer are classified into preventable and non-preventable. Factors that are associated with risk of breast cancer are

1) Advanced age
2) A personal history of breast conditions
3) Personal history of cancer
4) Early menarche
5) Late menopause
6) A family history of breast cancer
7) Obesity
8) Exposure to radiation
9) Having first child after 30 years of age
10) Nulliparity
11) Drinking alcohol

*Corresponding author.
Email address: Jabbar.meerja@gmail.com
doi: 10.14456/easr.2021.8

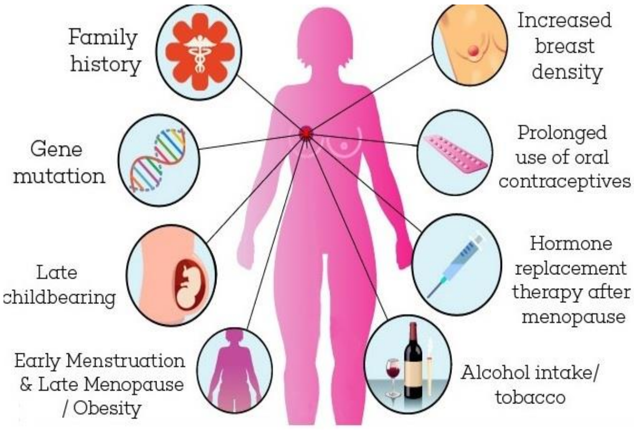

Figure 1 Causes and Risk Factors of Breast Cancer [8]

Causes and risk factors of breast cancer are pictorially shown in Figure 1.

1.3 Machine learning in disease prediction

ML is a form of Artificial Intelligence technique, which uses various probabilistic, statistical, and optimization methods to improve the performance from experiences and new data. Varieties of different ML techniques have been widely applied to disease prediction.

Machine Learning (ML) techniques played a remarkable role in the early diagnosis of Breast Cancer. As per survey report, most experienced physicians can diagnose breast cancer with 79% accuracy, while accuracy recorded by machine learning techniques is 91%. By using ML techniques, it is easy to distinguish people with Breast Cancer from others. Accurate classification will help clinicians to detect cancer in its early stage. Classification is a supervised and complex optimization problem. Different classification techniques such as K-NN, SVM, ANN, Naïve Bayes, the Bayesian network, CNN are used to classify cancer data.

The main contributions of our proposed approach are listed as follows.

1. This paper proposes a novel ensemble technique, built using heterogeneous classifiers such as Bayesian Network and RBF for the classification of Breast Cancer data.

2. We improved the performance of classifier to predict Breast Cancer.

3. We conduct experiments on WBCD dataset available from the UCI machine-learning repository.

The rest of the paper is organized as follows. Related work is discussed in section 2. The proposed method is described in section 3. Section 4 presents results and discussion. Finally, the paper is concluded in section 5.

2. Related work

This section discusses various literatures on classification of breast cancer data. Many Machine-learning algorithms have been applied for prediction of Breast Cancer.

In [9], Author applied K-NN on Breast Cancer dataset. They used DWT and BPNN tools for image filtering and processing respectively. The proposed method recorded on the accuracy of 98.80%.

Khuriwal et al proposed adaptive voting ensemble approach for Breast Cancer diagnosis [10]. Authors used chi-square feature selection method. ANN and Logistic Regression are combined and applied to Breast Cancer dataset. Their method recorded good accuracy when compared with other methods.

Cancer data classification using ML techniques was proposed in [11]. Authors used two-feature selection methods 1) Recursive feature elimination 2) RLR. Eight ML methods are applied to cancer dataset. Compared to other classification methods, SVM classifier recorded remarkable accuracy.

In [12] the authors employed four ML algorithms SVM, NB, K-NN and C4.5 on breast cancer data. Experiments were carried out using WEKA tool. Among four machine-learning algorithms, SVM classifier recorded an accuracy of 97.13%.

In [13] authors investigated the application of ML in breast cancer diagnosis. The WBCD dataset is taken for analysis. ANN, SVM, DT, and K-NN algorithms were used to analyze the dataset. Apart from ANN, several ML algorithms have dominated diagnosis of breast cancer [14-16].

Fu et al [17] performed a study using ML for detection of lymphedema among BC patients. Statistical and ML algorithms were used for data analysis. Total 355 patient's data was collected for the study. 18 features related to BC symptoms are used. Authors compared 5 classification algorithms like DT C4.5, DT C5.0, GBM, ANN, and SVM using RBF. Parameters are optimized using five cross validation. Among 5 classification algorithms ANN achieved an accuracy of 93.75% for detecting lymph edema. DT C4.5 recorded lowest accuracy of 76.31% for detecting lymphedema.

Detection of BC using CNN and SVM was proposed by Ragab et al [18]. Authors proposed a computer aided detection to classify tumors. CNN is used for feature extraction. Alexnet architecture is used to fine-tune the classes into two. DDSM and CBIS-DDSsM publicly available data sets were used for experimental analysis. SVM is used for classification. Proposed method recorded an accuracy of 87.2%, which will be used to detect abnormalities in breast cancer.

Agarap et al [19] presented a survey on applications of ML on breast cancer detection. Six ML algorithms namely GRU-SVM, MLP, NN, Linear regression, SVM and softmax regressions were used. Wisconsin Diagnostic breast cancer data set is used for experimental analysis. Data set is partitioned into training (70%) and Testing (30%). Among all the classifiers, MLP recorded an accuracy of 99.04%.

BC prediction using, SVM, Decision Tree and Naïve Bayes was proposed by Pritom et al [20]. Their method focused on finding whether breast cancer is a recursive or not.

Karthik et al [21] used Deep Neural Network (DNN) for classification of breast cancer data. DNN is applied on WBCD data set, which can be available in UCI repository. Recursive feature elimination is used to feature selection and to extract the best features and remove other features. DNN is used for classification. Their proposed method obtained an accuracy of 98.62% in DNN-RFE model.

Ebrahim Ali et al [22] proposed a method to classify breast cancer as Malignant or normal.SVM and ANN classification techniques are applied on WBCD data set. Authors claimed that ANN algorithm gives superior Accuracy=82.64%and Precision = 79% when compared to SVM.

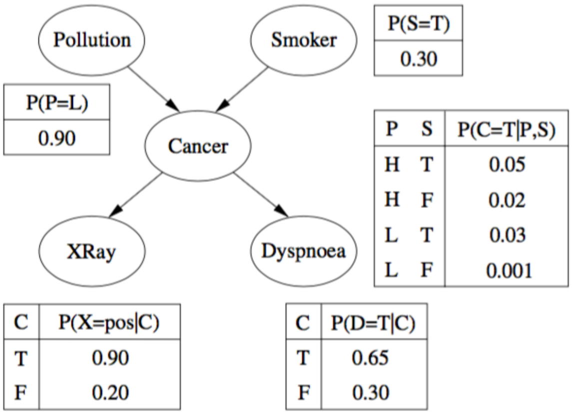

Figure 2 Bayesian Network Example

Singh et al [23] presented a paper to detect breast cancer using adaptive neural network .WBCD data set is used for experimental analysis. Adaptive Resonance Neural Network (ARNN) is used for detecting patience cancer stage. ARNN is a kind of Neural Network which performs unsupervised learning i.e. clustering .The proposed model recorded accuracy of 82.64% and with a precision of 79%.

Breast cancer classification based on mean radius and mean texture was proposed by Liu [24]. The proposed model obtained an accuracy of 90.48% .Logistic regression applied on breast cancer data received an accuracy of 90.5%

ANN and SVM based breast cancer detection was proposed by Wadkar et al [25]. Two ML classifiers SVM and ANN are applied on Breast cancer data which consists of 5000 images. Accuracy of their proposed method is recorded as 91% for SVM and 97% for ANN.

Cowsik and Clark proposed a ML model based on probabilistic perceptrons. Experiments were carried out on WBCD. Proposed model generates an estimate of probability of the cancer tumor being malignant or benign [26]. Accuracy of the model is recovered with 97% and with a standard deviation of 2%.

Classification model for breast cancer using 3 machine learning algorithms RF, LR, DT was implemented by Murugan et al [27] accuracy obtained using LR is reported as 84.14% where as for random forest it is 88.14%

Numerous machine learning algorithms [28] have been implemented for breast cancer prediction previously in the research cited above but most of the methods are not able to record remarkable accuracy in the diagnosis of breast cancer. Motivated from the existing work, this paper proposes an ensemble method to classify breast cancer data.

3. Materials and methods

In this paper, Ensemble classifier built with Bayesian Network and Radial Basis Function is applied for classification of breast cancer. Wisconsin Breast Cancer Data set (WBCD) is used for experimental analysis. Total 699 instances and 10 attributes are taken into consideration. WBCD is available from [29]. Bayesian network and RBFs are used in modeling ensemble classifier. The subsections will discuss about two individual classifier and the proposed ensemble classifier.

3.1 Bayesian network

Condition probability is very useful in the real world. Classical statistical models will not permit the introduction of prior knowledge into the model. Bayesian Network is used to represent the probabilistic relationships among items or objects. Bayesian Networks (BN) are directed acyclic graphs (DAG) where nodes represent variables, missing edges- conditional independence between variables. Each node is assigned to the probability function. Bayesian Networks are used to represent knowledge about an uncertain domain. Bayesian Networks are a combination of statistics, probability theory, and graph theory and computer science.

Bayesian Network represents a joint probability distribution over a set of random variables "A". The Bayesian Network is defined as B = <G, Ɵ> where G is DAG, which consists of nodes N1, N2….Nn and edges represent dependencies among variables.

Joint distribution for a Bayesian network defined as = P(node|parents(node)) for all nodes.

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i \middle| X_1, \dots, X_{i-1}) = \prod_{i=1}^{n} P(X_i|Parents(X_i)) \tag{1}$$

This joint distribution will reduce the computation when large networks are used. Figure 2 shows an example of Bayesian networks.

Motivation behind using the Bayesian network is the fact that they record good accuracy for complex and uncertain domains.

3.2 Radial Basis Function (RBF)

Radial basis function network is a supervised algorithm, which is derived from function approximation. RBF are 3 layered neural networks with I/P layer, O/P layer, and Hidden layer. Each Hidden unit has its own receptive field in input space. These units are called as radial centers which are represented by $V_1, V_2 \ldots V_n$ .The transformation from hidden units to output is linear whereas the transformation from input layer to the hidden unit is non-linear. RBF will establish a local mapping due to this RBF learns fastly. Twofold learning is used in RBF both weight and centers have to be learned.

RBF performs supervised learning by calculating the input's similarity to examples from the training set. To classify a new sample, each neuron computes the distance between the input and its prototype. Each RBF neuron computes similarity between the input and its prototype vector. Gaussian function is mostly used in RBF. One dimensional input Gaussian equation is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-u)^2}{2\sigma^2}} \tag{2}$$

Where x is the input, u is the mean, and sigma is the standard deviation.

The RBF neuron activation function is defined as

$$\varphi(x) = e^{-\beta\|x-\mu\|^2} \tag{3}$$

Various radial basis functions are

1) The Gaussian radial basis function
2) Quadratic
3) Inverse quadratic
4) Thin plate spline

Motivation behind the RBF is RBF learns fast compared to a simple feed forward network due to following good features like

1) Training is fast
2) Like MLP output nodes implement linear summation function
3) They are good at interpolation
4) Radial basis functions are like thin plate are implemented by hidden nodes

3.3 Ensemble learning

Ensemble learning is a supervised machine learning method, which uses multiple learning algorithms in order to obtain one optimal predictive model. It combines the diverse set of supervised learners to improve the predictive power of the model. Commonly used ensemble learning techniques are 1) Bagging 2) Boosting 3) Stacking.

Bagging is a way to decrease the variance. Bagging combines weak learners by averaging process. Bagging aims to **decrease variance**, not bias. This method is also called parallel ensemble learning. These are suitable for complex models especially suitable for high variance low bias models.

Boosting is a two-step approach, also called sequential ensemble learning and is used to decrease bias not variance. These models suitable for low variance and high bias models.

Stacking is similar to boosting and called as meta learning approach. Features are retrieved using ensemble which will be used by other layers. Stacking is normally used in competitions, where multiple algorithms are used to train the data and average of these results is taken.

Ensemble classifiers are constructed from a set of base (weak) classifiers and classify new sample by taking weight or majority voting.

Base learners are also called weak learners used by different algorithms. For example the individual base learners like naïve bayes, knn, SVM, Bayesian, Decision Trees, SVM, etc. Base learners are having different tuning parameters. Like in Random Forest ensemble, decision tree is base learning algorithm. Base learners are also called weak learners .weak learners records less accuracy often but not regularly. It is weak when learning the relationships between input and output. Ensemble learning is used to combine weak learners to get good accuracy. Selection of algorithm is based on the method.

Ensemble classifiers will handle three issues, like 1) Statistical 2) computational 3) representational, where weak classifiers can't handle these issues.

Ensemble learning is a powerful method in ML used to improve the performance of the model. The diverse sets of learners (Base Learners) are combined to improve the predictive power of the model. Choosing the right ensemble is most important for prediction of Breast Cancer.

3.4 Proposed algorithm

Stages in the proposed approach are

Algorithm: Breast Cancer Data Classification using Ensemble Machine Learning

Step 1: Input WBCD data set

Step 2: Apply Preprocessing Technique on WBCD

Step 3: Partition the data set as Training and Test data set

Step 4: Build ensemble learning (Majority voting) using Bayesian network and RBF with 10 Fold cross Validation

Step 5: For each new sample let say "S" Test data

Step 6: Ensemble classifier (test sample) = majority voting (Class (RBF), class (Bayesian Network))

Step 7: Record the accuracy of the proposed approach and evaluate using various metrics

Process flow diagram of proposed approach is shown in Figure 3.

Wisconsin breast cancer data is collected from the UCI machine-learning repository. This dataset preprocessed in step 2. Preprocessing in machine learning is used to remove redundant data and to transform raw data into an understandable format. After applying the preprocessing on WBCD, ensemble learning with majority voting using RBF and Bayesian network is applied. Cross-validation with 10 folds is chosen for experiments.

In our proposed approach, we used RBF and Bayesian Network as base learners and combined using ensemble learning method. You can find explanation about RBF and Bayesian Network section 3.

Majority voting can be expressed as

$$\sum_{t=1}^{T} \boldsymbol{d_{t,J}} = \max_{\boldsymbol{j=1}}^{\boldsymbol{C}} \sum_{t=1}^{T} \boldsymbol{d_{t,j}} \tag{4}$$

Where T is the data set and C is the class label.

4. Results and discussion

This section discusses data analysis using ensemble learning. We evaluated the effectiveness of ensemble classifier using various evaluation metrics like accuracy, precision, recall, f measure, TPR, TNR, MCC. These metrics are derived from the confusion matrix, which is represented in Table 1.

TN refers to a number of normal cases identified as normal, FN is the number of breast cancer cases incorrectly identified normal, TP is the number of breast cancer cases correctly identified as they are, and FP is the number of normal cases incorrectly identified as breast cancer. Metrics derived from Confusion matrix are

1. Sensitivity =Recall= TP/(FN+TP)
2. Specificity = TN/ (TN+FP)
3. Accuracy = (TP+TN)/ (TP+TN+FP+FN)
4. Positive Predicted Value (PPV) = precision =TP/TP+FP
5. Negative Predicted Value (NPV) = TN/TN+FN
6. F-measure= $\frac{2*precison*recall}{precison+recall}$
7. Error rate (ER) = $\frac{FP+FN}{TP+TN+FP+FN}$
8. False positive rate = FP/FP+TN

Table 1 Confusion matrix

	Predicted patients with disease	**Predicted patients with no disease**
Patient with disease	TP	FN
Healthy Persons (No Disease)	FP	TN

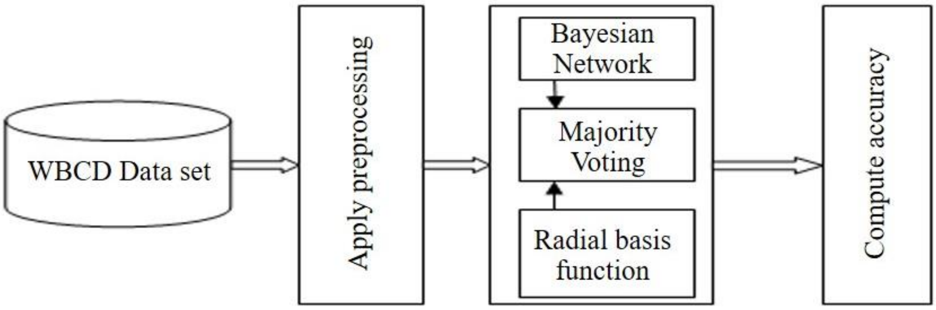

Figure 3 Process Flow Diagram of the Proposed Approach

Table 2 WBCD data set description

S. no	Attribute Name	Value
1	Clump Thinkness	1-10
2	Uniformity of Cell Size	1-10
3	Unifomity of Cell Shape	1-10
4	Marginal Adhesion	1-10
5	Single Epithelial Cell Size	1-10
6	Bare Nuclei	1-10
7	Bland Chromatin	1-10
8	Normal Nucleoli	1-10
9	Mitoses	1-10
10	Class	For Benign=2 Malignant=4

Table 3 Results obtained by proposed approach

SI. no	Metric	Value
1	Accuracy	97.42
2	Precision or PPV	96.72
3	Recall or Sensitivity	99.32
4	Specificity	94.07
5	NPV	98.75
6	F measure	98.00
7	FPR	0.06
8	TPR	0.99
9	MCC	0.44

Table 4 Comparison of proposed approach with existing approaches.

SI. no	Author	Technique used	Accuracy recorded
5	Fu MR [17]	ANN	93.75
1	Kumar et al [28]	SVM-NB	97.13
2	Karabatak and Ince [30]	AR-ANN	97.40
3	Seera and Lim [31]	FMM-CART	97.29
4	Chen et al [32]	RS-SVM	96.87
6	Proposed approach	BN+RBF	97.42

9. Mathews Correlation Coefficient (MCC) =

$$\mathbf{MCC} = \frac{\mathbf{TP} \times \mathbf{TN} - \mathbf{FP} \times \mathbf{FN}}{\sqrt{(\mathbf{TP} + \mathbf{FP}) \times (\mathbf{FN} + \mathbf{TN}) \times (\mathbf{TP} \times \mathbf{FN}) \times (\mathbf{FP}\ \mathbf{TN})}}$$

If MCC =+1 prediction was 100% accurate and prediction is wrong if MCC = -1. The proposed system is evaluated on WBCD data set. WBCD was used by a number of researchers in ML and Pattern Recognition.

Table 5 Accuracy obtained by various ML Techniques

SI. no	*ML Techniques*	*Accuracy*
1	SVM-NB	97.13
2	AR-ANN	97.40
3	FMM-CART	97.29
4	RS-SVM	96.87
5	ANN	93.75
6	GRU-SVM	93.75
7	Linear Regression	96.09
8	L1-NN	93.56
9	L2-NN	94.73
10	SVM	96.09
11	SVM-RBF	93.47
12	LP-SVM	97.33
13	EM-PCA-CART	93.2
14	Proposed Method(BN+RPF)	97.42

The WBCD dataset consists of 11 attributes and the first one is ID, which will be removed from the data set before classification and having 699 instances. This dataset contains two classes either benign or malignant. Data set description is presented in Table 2. Missing values are normalized and experiments were conducted using WEKA tool.

Table 3 describes results obtained by the proposed ensemble learning. 10 cross-validations is used to classify the data set. The accuracy of the proposed method is recorded at 97.42 %. FPR for the ensemble learning is 0.06.

The number of clusters used in RBF is 2 and simulated annealing search technique is used in the Bayesian network. Recall and TPR values are recorded highest in our proposed method with 99.32% and 99.0 respectively. Mathew's correlation coefficient (MLCC) is a balanced measure, which is used when the classes are imbalanced. MCC has a range of values from -1 to 1 where 1 indicates a correct binary classifier and -1 indicates the wrong classifier. Proposed classifier recorded MCC values as 0.944 from the MCC value; it is evident that proposed model is well suited for the classification of breast cancer data.

Accuracy comparisons of the proposed approach with other existing approaches were shown in Table 4. Accuracy obtained by various machine-learning approaches was shown in Table 4 and Figure 4 and Figure 5. From the comparison table, we claim that the proposed approach records reasonable accuracy compared with other methods. Table 5 compares the accuracy of recent machine learning techniques applied on breast cancer data. It is evident from the Figure 5 result that the proposed approach recorded good accuracy. It implies that this model performs well for classification of breast cancer data. Table 6 gives the comparison of various approaches for breast cancer detection Time taken by each function is given in Table 7 and Figure 6.

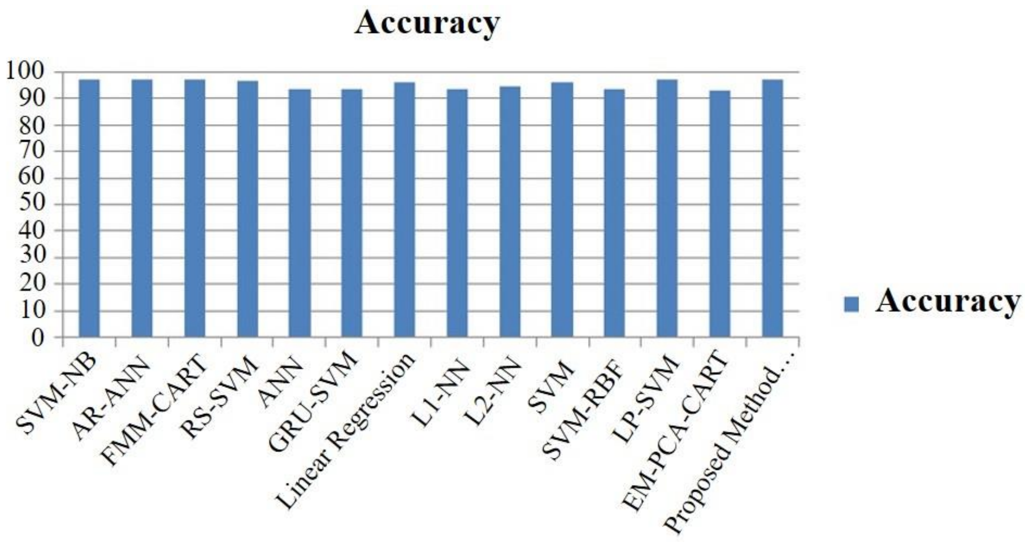

Figure 4 Accuracy obtained by various approaches

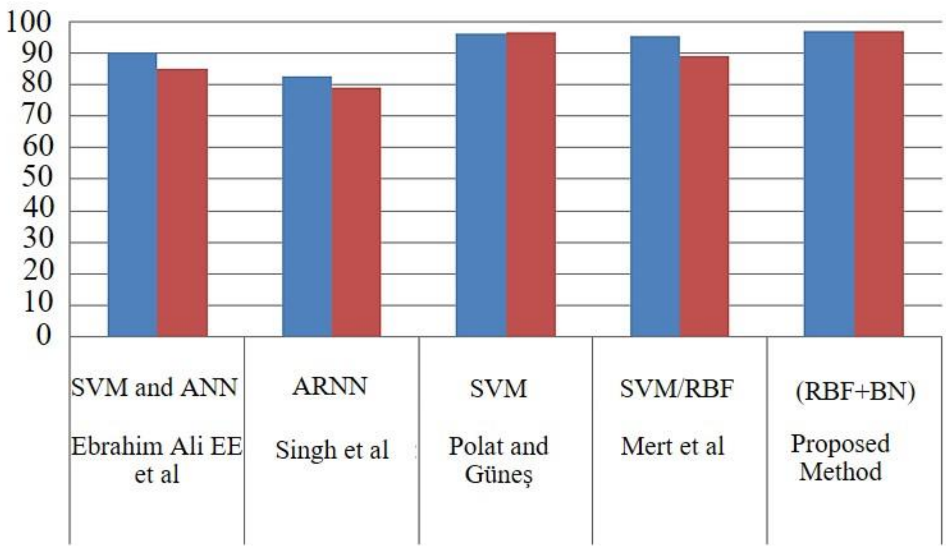

Figure 5 Accuracy obtained by various approaches

Table 6 Comparison of accuracy for various approaches for breast cancer prediction (Experiments carried out on WBCD)

SI. no	Author	Method	Accuracy	Precision
1	Ebrahim Ali et al [22]	SVM and ANN	90%	85%
2	Singh et al [23]	ARNN	82.64%	79%
3	Bayrak et al [33]	SVM(SMO)	96.9	95.4
4	Chaurasia et al [34]	RBF	96	96.23
5	Chaurasia et al [34]	J48	93.41	90.37
6	Polat and Güneş [35]	SVM	95.89	96.4%
7	Mert et al [36]	SVM/RBF	87.17 for RBF 95.25 for SVM	89% for RBF 88.9 for SVM
8	Amrane et al [37]	Multi-Layer Perceptron (MLP)	95.4	95.4
9	Proposed Method	(RBF+BN)	97%	96.72%

Table 7 Time taken by each function in predicting the model

Radial basis functions	Time taken to build the model(min)
The Gaussian radial basis function	2.15
Quadratic	3
Inverse quadratic	3.45
Thin plate spline	5

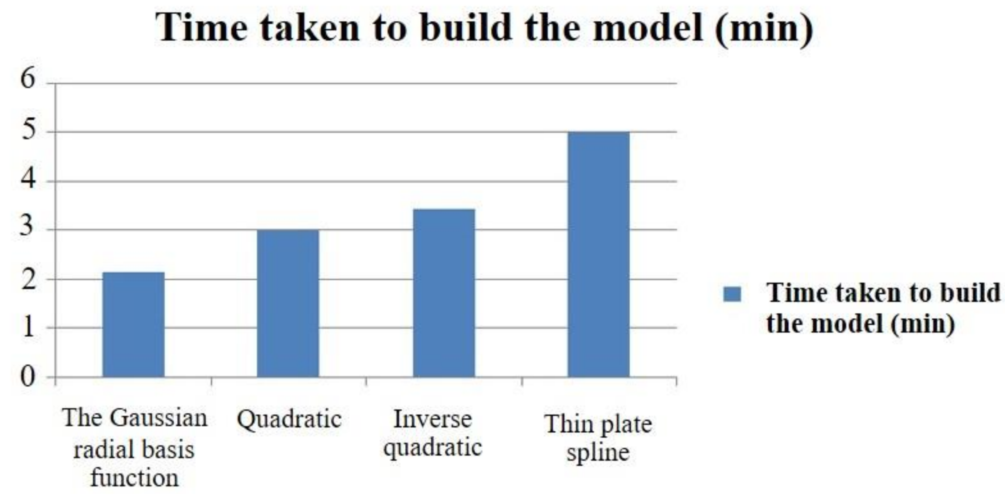

Figure 6 Time Recordings for various methods

5. Conclusion

Accurate and effective classification of breast cancer data is critical in medical diagnosis. Machine learning methods are widely used and are instrumental in classifying health care data. Although many techniques have been developed to classify breast cancer data, still face various challenges like accuracy. To address this, we proposed a model for classification of breast cancer data. In this work Ensemble model is build using Bayesian network and RBF to classify the breast cancer data. Bayesian network and RBF provides good accuracy in classification. Various metrics are used to measure the performance of the model. The proposed model is compared with other existing methods in classifying the breast cancer data. By combining these two heterogeneous classifiers, we obtained an optimal and remarkable accuracy of 97.42%. From the results it has been proved that proposed techniqueis, in each perspective, is superior to single classifier. The proposed ensemble learning system would be viable in helping cancer specialists in recognizing cancer.

6. References

[1] Qasem A, Sheikh Abdullah SNH, Sahran S, Iqbal Hussain R, Ismail F. An accurate rejection model for false positive reduction of mass localisation in mammogram. Pertanika J Sci Technol. 2017;25(S6):49-62.

[2] DeSantis C, Siegel R, Bandi P, Jemal A. Breast cancer statistics. CA Cancer J clin. 2011;61:408-18.

[3] Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China 2015. CA Cancer J Clin. 2016;66(2):115-32.

[4] Siegel RL, Miller KD, Jemal A. A cancer statistics 2016. CA Cancer J Clin.2016;66(1):7-30.

[5] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. A global cancer statistics 2012. CA Cancer J Clin.2015;65(2):87-108.

[6] Yue W, Wang Z, Chen H, Payne A, Liu X. Machine learning with application in breast cancer diagnosis and prognosis. Designs. 2018;2:1-17.

[7] MayoClinic. Breast cancer [Internet]. 2020 [cited 2020 Jan 1]. Available from: https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470.

[8] Medindia. Breast Cancer [Internet]. 2020 [cited 2020 Jan 1]. Available form: https://www.Medindia.net.

[9] Al-Hadidi MR, Alarabeyyat A, Alhanahnah M. Breast cancer detection using K-nearest neighbormachine learning algorithm. 9th International Conference on Developments in eSystems Engineering; 2016 Aug 31 – Sep 2; Liverpool, UK. USA: IEEE; 2016. p. 35-9.

[10] Khuriwal N, Mishra N. Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm. 2018 IEEMA Engineer Infinite Conference (eTechNxT); 2018 Mar 13-14; New Delhi, India. USA: IEEE; 2018. p. 1-5.

[11] Turgut S, Dağtekin M, Ensari T. Microarray breast cancer data classification using machine learning methods.2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT); 2018 Apr 18-19;Istanbul, Turkey. USA: IEEE; 2018. p. 1-4.

[12] Asri H, Mousannif H, Moatassime HA, Noel T. Using machine learning algorithms for Breast cancer risk Prediction and Diagnosis. Procedia Comput Sci. 2016;83:1064-9.

[13] Shailaja K, Seetharamulu B, Jabbar MA. Machine learning in healthcare: a review.2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA); 2018 Mar 29-31; Coimbatore, India. USA: IEEE; 2018. p. 910-4.

[14] Shailaja K, Seetharamulu B, Jabbar MA. Prediction of breast cancer using big data analytics. Int J Eng Tech. 2018;7(4.6):223-6.

[15] Jabbar MA, Samreen S, Aluvalu R. The future of healthcare: machine learning. Int J Eng Tech. 2018;7(4.6):23-5.

[16] Douangnoulack P, Boonjing V. Building minimal classification rules for breast cancer diagnosis. 2018 10th International Conference on Knowledge and Smart Technology (KST); 2018 Jan 31-Feb 3; Chiang Mai, Thailand. USA: IEEE; 2018. p. 278-81.

[17] Fu MR, Wang Y, Li C, Qiu Z, Axelrod D, Guth AA, et al. Machine learning for detection of lymphedema among breast cancer survivors. MHealth. 2018;4:17.

[18] Ragab DA, Sharkas M, Marshall S, Ren J. Breast cancer detection using deep convolutional neural networks and support vector machines. Peer J. 2019;7:e6201.

[19] Agarap AFM. On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. Proceedings of the 2nd International Conference on Machine Learning and Soft Computing; 2018 Feb 2-4; Phu Quoc Island, Viet Nam. New York: ACM; 2018. p. 5-9.

[20] Pritom AI, Munshi MAR, Sabab SA, Shihab S. Predicting breast cancer recurrence using effective classification and feature selection technique. 19th International Conference on Computer and Information Technology (ICCIT); 2016

Dec 18-20; Dhaka, Bangladesh. USA: IEEE; 2016. p. 310-4.

[21] Karthik S, Srinivasa Perumal R, Chandra Mouli PVSSR. Breast cancer classification using deep neural networks. In: Margret Anouncia S, Wiil U, editors. Knowledge Computing and Its Applications. Singapore: Springer; 2018. P. 227-41.

[22] Ebrahim Ali EE, Feng WZ. Breast cancer classification using support vector machine and neural network. Int J Sci Res. 2016;5(3):1-6.

[23] Singh S, Saini S, Singh M. Cancer detection using adaptive neural network. Int J Adv Res Tech. 2012;1(4):93-7.

[24] Liu L. Research onlogisticregression algorithm of breast cancer diagnosis data by machine learning. 2018 International Conference on Robots & Intelligent System (ICRIS); 2018 May 26-27; Changsha, China. USA: IEEE; 2018. p 157-60.

[25] Wadkar K, Pathak P, Wagh N. Breast cancer detection using ANN network and performance analysis with SVM. Int J Comput Eng Tech. 2019;10(3):75-86.

[26] Cowsik A, Clark JW. Breast cancer diagnosis by higher order probabilistic perceptrons. arXiv: 1912.06969. 2019:1-17.

[27] Murugan S, Kumar BM, Amudha S. Classification and prediction of breast cancer using linear regression, decision tree and random forest. 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC); 2017 Sep 8-9; Mysore, India. USA: IEEE; 2017. p. 763-6.

[28] Kumar UK, Nikhil MBS, Sumangali K. Prediction of breast cancer using voting classifier technique. 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM); 2017 Aug 2-4; Chennai, India. USA: IEEE; 2017. p. 108-14.

[29] Wolberg WH, Street WN, Mangasarian OL. Breast Cancer Wisconsin (Diagnostic) Data Set [Internet]. 2020 [cited 2020 Jan 2] Available from: https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic).

[30] Karabatak M, Ince MC. An expert system for detection of breast cancer based on association rules and neural network. Expert Syst Appl. 2009;36(2):3465-9.

[31] Seera M, Lim CP. A hybrid intelligent system for medical data classification. Expert Syst Appl. 2014;41(5):2239-49.

[32] Chen HL, Yang B, Liu J, Liu DY. A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. Expert Syst Appl. 2011;38(7):9014-22.

[33] Bayrak EA, Kirci P, Ensari T. Comparison of machine learning methods for breast cancer diagnosis.2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT); 2019 Apr 24-26; Istanbul, Turkey. USA: IEEE; 2019. p. 1-3.

[34] Chaurasia V, Pal S, Tiwari B. Prediction of benign and malignant breast cancer using data mining techniques. J Algorithm Comput Tech. 2018;12(2):119-26.

[35] Polat K, Güneş S. Breast cancer diagnosis using least square support vector machine. Digit Signal Process. 2007;17(4):694-701.

[36] Mert A, Kilic N, Akan A. Breast cancer classification by using support vector machines with reduced dimension. Proceedings ELMAR-2011; 2011 Sep 14-16; Zadar, Croatia. USA: IEEE; 2011. p. 37-40.

[37] Amrane M, Oukid S, Gagaoua I, Ensarİ T. Breast cancer classification using machine learning. 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT); 2018 Apr 18-19; Istanbul, Turkey. USA: IEEE; 2018. p. 1-4.