

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/323210602>

Breast Cancer Classification Using Deep Neural Networks

Chapter · February 2018

DOI: 10.1007/978-981-10-6680-1_12

CITATIONS

12

READS

2,800

3 authors:



Karthik Sekaran

VIT University

7 PUBLICATIONS 56 CITATIONS

SEE PROFILE



Srinivasa Perumal Ramalingam

VIT University

31 PUBLICATIONS 85 CITATIONS

SEE PROFILE



Chandra Mouli P.V.S.S.R.

VIT University

70 PUBLICATIONS 251 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



on Face Analysis [View project](#)



Pedestrian detection in image [View project](#)

Breast Cancer Classification Using Deep Neural Networks

S. Karthik, R. Srinivasa Perumal and P. V. S. S. R. Chandra Mouli

Abstract Early diagnosis of any disease can be curable with a little amount of human effort. Most of the people fail to detect their disease before it becomes chronic. It leads to increase in death rate around the world. Breast cancer is one of the diseases that could be cured when the disease identified at earlier stages before it is spreading across all the parts of the body. The medical practitioner may diagnose the diseases mistakenly due to misinterpretation. The computer-aided diagnosis (CAD) is an automated assistance for practitioners that will produce accurate results to analyze the criticality of the diseases. This chapter presents a CAD system to perform automated diagnosis for breast cancer. This method employed deep neural network (DNN) as classifier model and recursive feature elimination (RFE) for feature selection. DNN with multiple layers of processing attained higher classification rate than SVM. So, the researchers used deep learning method for hyper-spectral data classification. This chapter used DNN to learn deep features of data. The DNN with multiple layers of processing is applied to classify the breast cancer data. The system was experimented on Wisconsin Breast Cancer Dataset (WBCD) from UCI repository. The dataset partitioned into different sets of train-test split. The performance of the system is measured based on accuracy, sensitivity, specificity, precision, and recall. From the results, the accuracy obtained 98.62%, which is better than other state-of-the-art methods. The results show that the system is comparatively outperformed than the existing system.

Keywords Classification · Deep learning · Healthcare system · Feature selection
Breast cancer diagnosis

S. Karthik (✉) · R. Srinivasa Perumal

School of Information Technology and Engineering, VIT University, Vellore, India
e-mail: karthik.s2012@vit.ac.in

R. Srinivasa Perumal

e-mail: r.srinivasaperumal@vit.ac.in

P. V. S. S. R. Chandra Mouli

School of Computer Science and Engineering, VIT University, Vellore, India
e-mail: chandramouli@vit.ac.in

1 Introduction

Nowadays, there is a drastic level of increase in the amount of women around the world, getting affected by breast cancer and it is raising over time. Earlier detection of breast cancer reduces death rate and avoids it till reaches the chronic level. In 2012, almost 1.7 million women are affected by breast cancer [20]. The impact of breast cancer is increasing day by day, due to that the healthcare professionals are not in a condition to state the people affected at different stages earlier to save their lives. Digital mammography is a major diagnosis model used throughout the world for breast cancer detection. Computer-aided diagnosis (CAD) is widely used in detecting numerous diseases with accurate decision. It assists the healthcare professionals to analyze and conclude the stages of the various diseases. CAD systems are developed in a way to provide promising results and perfect decisions on patient's condition that helps medical practitioners to diagnose the stages of the diseases. It supports radiologists to avoid misconceptions and wrong diagnosis due to inaccurate data, lack of focus, or inexperience, who uses visually screening mammogram of patients. The aim of the system is to develop a novel CAD model to diagnose a breast cancer in earlier stages with more accurate results to save their precious lives by using DNN.

DNN followed by a similar artificial neural network with a complex network structure which has 'n' hidden layers, which can process the input data from the previous layer. The error rate of the input data will be consistently reduced by adjusting the weights of every node, which leads to achieve an accurate result. It helps to create a model and define its complex hierarchies in a simple form. It supports all kinds of algorithms, namely supervised, unsupervised, semi-supervised, and reinforcement. So, the system did not define any specific algorithm. The DNN generates a better model themselves to train the given data. The architecture of the DNN is shown in Fig. 1.

The rest of the chapter is summarized as follows: Section 2 describes the related work, Sect. 3 explains the proposed method, Sect. 4 demonstrates the experimental results, and Sect. 5 concludes the work.

2 The Literature Review

Neural Network is inspired by the working principle of biological neural networks, which has its own input and output channels called as dendrites and axons, respectively. A typical ANN will have millions of processing units or elements, which forms a highly interconnected network that processes a huge amount of information based on the response, given by the external input of a computing system. Every single neuron in a typical neural network is called as unit. A layer in a neural network is considered as a set of neurons in a stack. A layer may have n number of nodes in it. A typical neural network system has single input layer and

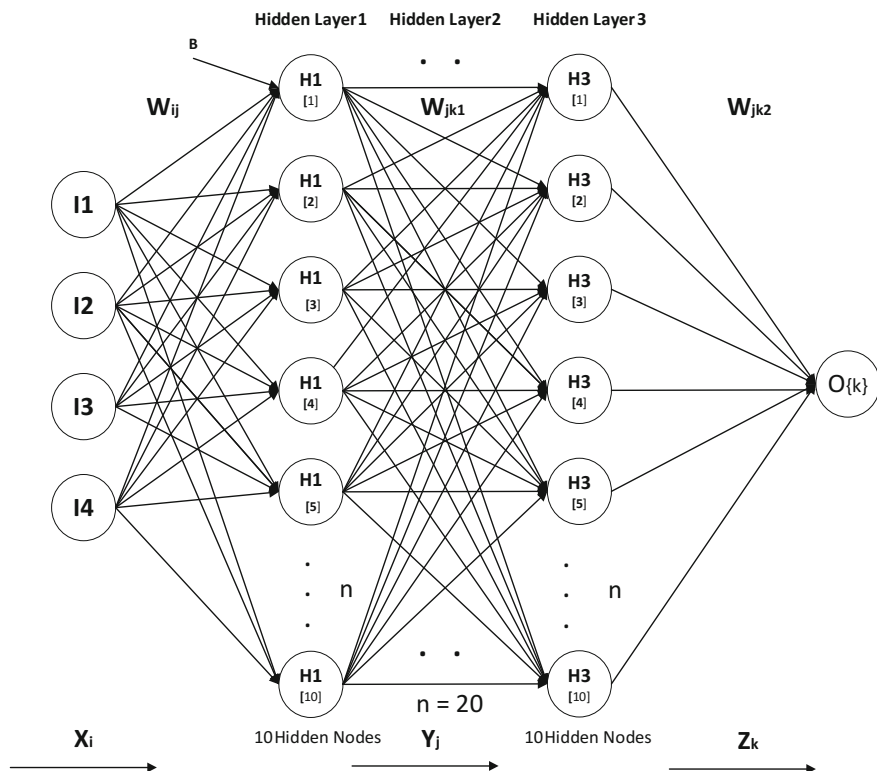


Fig. 1 Deep neural network architecture

may have a single or two hidden layers, which is directly connected to the output layer, which receives input from the input layer, i.e., previous layer of the current node.

Abbass developed a system with pareto-differential evaluation algorithm with local search scheme, called memetic pareto-artificial neural network (MPANN) [1]. MPANN analyzes the data effectively than other models. The method achieved 98.1% accuracy on random split. Tuba and Tulay proposed the statistical neural network-based breast cancer diagnosis system [13]. In the diagnosis system, they used RBF, general regression neural network (GRNN), and statistical neural network structures on WDBC dataset. The system obtained 98.8% on 50–50 partitioning split.

Paulin and Santhakumaran [19] developed a system with back-propagation neural network (BPNN) and obtained 99.28% accuracy with Levenberg–Marquardt algorithm. They used median filter for preprocessing and normalized the data using min–max technique. None of the features are eliminated from the dataset. The accurate result attained from 80:20 partition scheme. Karabatak and Ince [12] developed an expert system for breast cancer detection. Association rules (AR) are

used to reduce the dimensions of the dataset. In the system, AR1 and AR2 are developed to reduce the features. AR1 reduces one feature from 9, and AR2 reduces 5 features out of 9. The conventional neural network is used for classification in both AR1 and AR2. The method attained 95.6% accuracy on AR1, 97.4% on AR2, and 95.2% on all 9 features with threefold cross-validation scheme.

Mert et al. [14] used radial basis function neural network (RBFNN) for medical data classification and independent component analysis for feature selection. The method selects the one feature vector randomly from 30 features. The method obtained the accuracy in the average of 86%. Bhattacharjee et al. [6] used BPNN for classification. The method achieved 99.27% accuracy. An intelligent medical decision model was developed based on evolutionary strategy [8]. They validated the performance of the method by testing on different datasets. Neural network (NN), genetic algorithm (GA), support vector machine (SVM), K-nearest neighbor (KNN), multilayer perceptron (MLP), radial basis function (RBF), probabilistic neural network (PNN), self-organizing map (SOM), and Naive Bayes (NB) are used as classifiers. Crossover and mutation techniques are applied between different algorithms. The method proves that the SVM classifier on WBC dataset attained better recognition rate than other classifiers.

Ahmad and Ahmed investigated three classification methods, namely RBF, MLP, and PNN and applied to WBC dataset [4]. Among these, PNN shows better result of 97.66% and outcome of MLP is 96.34%. RBF outperformed than MLP, but validation error is little bit higher. Jhaharia et al. [10] proposed a model for breast cancer classification using PCA, to extract features from the dataset and a feed-forward neural network to classify the data. The result is obtained by dividing data by percentage split for training and testing data. Kemal Polat and Salih Gne examined WBCD dataset with least square support vector machine (LSSVM). A set of linear equations is used for training process in LSSVM, but quadratic optimization problem is used in SVM for the same process. This system achieved 98.53% accuracy with k -fold cross-validation method.

Jouni et al. [11] proposed a model based on artificial neural network with multilayered perceptron networks and BPNN. This model learns to classify whether the result of the simulation will be malignant or benign. It also includes weight adjustment factors and bias values. Bewal et al. [5] used multilayered perceptron network with four back-propagation training algorithms such as quasi-Newton, gradient descent with momentum and adaptive learning, Levenberg–Marquardt, and resilient back propagation to train the network. Steepest descent back propagation is used to measure the performance of other neural networks. Levenberg–Marquardt algorithm with MLP achieved best accuracy rate of 94.11%.

SVM with recursive feature elimination (RFE) applied on Wisconsin Diagnostic Breast Cancer (WDBC) Dataset. Principal component analysis (PCA) applied separately for the same dataset for dimensionality reduction process, and SVM is used to classify the dataset. After PCA applied, 25 features are selected. It achieved 98.58% which outperforms SVM and SVM-RFE techniques [22]. Huang et al. [9] investigated a new classification system to develop a robust system using SVM ensembles. Two ensemble learning methods are called bagging and boosting.

Any kernel function in SVM applied to dataset without feature selection process will increase computational time of the system. GA is used to select best features from dataset. The results show that, for small scale datasets, linear kernel with bagging ensembles and RBF with boosting ensembles outperforms than other classifiers. The dataset is divided into 90–10% splits based on k -fold cross-validation. GA+RBFSVM achieved accuracy of 98.00 and 99.52%, respectively, for small and large datasets. Nayak et al. [16] proposed a system which includes adaptive resonance theory (ART-1) network for classification, and it is compared with PSO-MLP and PSO-BBO algorithms which prove that ART is best than other two classifiers. They split the dataset into 70–30 for training and testing the dataset.

Anooj [3] used weighed fuzzy rules to develop a clinical decision support system (CDSS). This system mainly involves two processes, namely generation of fuzzy rules and developing a fuzzy-rule-based decision support system. To ensure better predictions, fuzzy-rule-based decision support system is taken decisions based on the fuzzy rules generated with decision support system. Fuzzy rules are generated based on historical data for better learning. It applied weighted fuzzy rule based on the importance of attributes. Onan [18] proposed a novel classification model based on fuzzy-rough nearest neighbor method. This system consists of three phases, namely instance selection, feature selection, and classification. Fuzzy-rough instance selection method is used for instance selection with weak gamma evaluator to remove erroneous and ambiguous instances. Consistency-based feature selection method is used in conjunction with re-ranking algorithm to efficiently search for possible enumerations in search space. Fuzzy-rough nearest neighbor method is used for the classification process. The method outperformed than other fuzzy approaches.

Ghosh et al. [7] introduced a neuro-fuzzy-based breast cancer classification system. They used WBC, WDBC, and mammographic mass (MM) datasets to apply and evaluate the method. The dataset fuzzified using sigmoidal membership functions and computes degree of membership for individual patterns to various classes. Multilayer perceptron model is used for classification process. At last, defuzzification is applied to generate the results. This method achieved 97.8% accuracy rate. Nilashi et al. [17] proposed knowledge-based system with fuzzy logic for breast cancer. Maximization technique is applied for grouping (clustering) of data. To overcome the problem of multicollinearity, PCA was used. Classification and regression technique (CART) is applied to dataset to generate fuzzy rules for knowledge-based system. Finally, fuzzy-rule-based reasoning system is used the fuzzy rules for classification. Kindie developed breast CDSS with rough set and BPNN-based knowledge mining process [15]. Rough set indiscernibility relation method is applied on dataset to extract minimal set of attributes. Also missing of data is handled in this process. Then, BPNN is used for classification of the dataset. The dataset was divided into 80–20% splits, and it attained 98.6% accuracy.

Schmidhuber [21] provides an overview of deep learning in neural networks. The method proves that the deep learning algorithms reduced the error rate and increase the accuracy with respect to training of algorithm. Abdel and Eldeib [2]

applied deep belief network (DBN) for WBC dataset and achieved 99.68% accuracy. The dataset was divided into train-test split of 54.945.1%. DBN follows unsupervised path and back-propagation network to follow supervised path. This system was constructed by BPNN with Levenberg–Marquardt learning function. Here, the weights are initialized with DBN path. This system provides the promising result and outperforms than other classifiers. This motivated to use deep learning concepts for medical data classification. Deep learning reduces the error rate, and it will improve the accuracy rate.

3 Proposed System

The proposed method is used DNN model for classification process and RFE system for selecting a subset of features from all given features. The steps involved in the proposed method are described below, and it is represented in Fig. 2.

1. The WBC dataset is preprocessed to remove the noise from the instances.
2. Select the best four features from the dataset by applying logistic regression model.
3. For iteration process, recursive feature elimination is used.
4. RFE ranked and extracted the best feature and eliminates other features.
5. Classify the dataset using deep neural network.

3.1 Preprocessing

In a machine learning model, to normalize and eliminate redundant, ambiguous data from the dataset, preprocessing techniques are applied. In breast cancer dataset, it consists of 699 instances with 9 feature variables. This dataset supports binary classification models since it has only two class labels called benign and malignant.

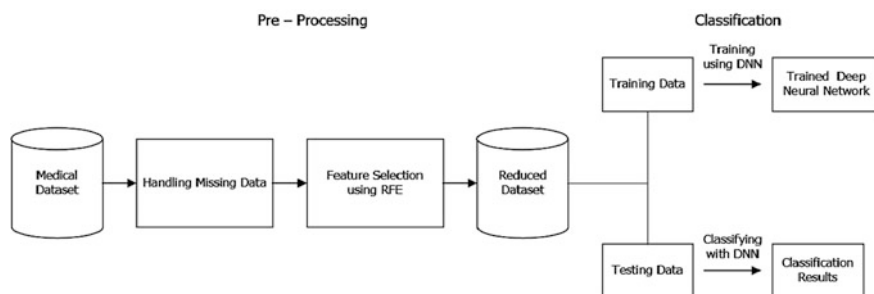


Fig. 2 Architecture of the system

Benign is to identify patients without cancer, and malignant is to identify patients with cancerous tumors. In the 699 instances, 16 of them have missing values. To handle those values, the system removed all the 16 instances from the dataset before feature selection to improve the stability of the system. This system selects the best features from the feature variables to improve the performance and accuracy of the system.

3.2 Feature Selection

The importance of feature selection in a machine learning model is inevitable. It turns the data to be free from ambiguity and reduces the complexity of the data. Also, it reduces the size of the data, so it is easy to train the model and reduces the training time. It avoids over fitting of data. Selecting the best feature subset from all the features increases the accuracy. Some feature selection methods are wrapper methods, filter methods, and embedded methods.

3.2.1 Recursive Feature Elimination

RFE is one of the best feature selection techniques, which adopts greedy optimization technique that comes under wrapper methods. It selects a feature subset from all the given features. In every iteration, the subset is selected, using logistic regression model to train the features. It will be decided whether to keep the feature or to remove it from the subset of features. The model will be constructed by removing unnecessary features. This process continues until all the best features are selected as a subset. Finally, those selected features are ranked based on their elimination order that can calculate every iteration. The RFE process is illustrated in Fig. 3.

In this system, RFE is applied on the pre-processed data to rank the features for next level. These features are chosen based on the outcome of the result from the logistic regression model. At the end of all the iterations, remaining features will be available for selecting those specific features for reducing the dataset. The system

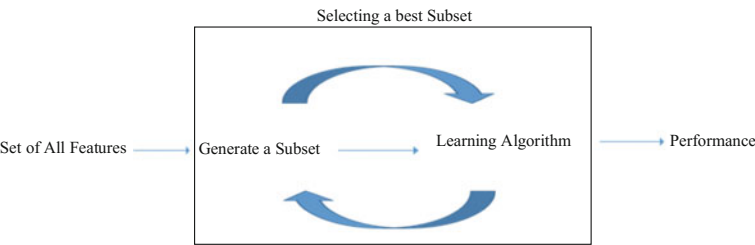


Fig. 3 Process of recursive feature elimination

Table 1 Ranking of attributes using RFE

Attributes	Ranking
Clump thickness	1
Uniformity of cell size	1
Uniformity of cell shape	1
Marginal adhesion	5
Single epithelial cell size	6
Bare nucleoli	1
Bland chromatin	2
Normal nucleoli	4
Mitosis	3

selected the four best ranked features to provide input for DNN classifier. Those features are clump thickness, uniformity of cell size, uniformity of cell shape, and bare nucleoli. The ranking of attributes are shown in Table 1.

Algorithm of Recursive Feature Elimination

1. Initially, select all the features for feature selection

$$f_n = f_1, f_2, f_3 \dots f_{n-1}$$

2. Define a model to train the selected features during every iteration
Logistic regression: $y = f(x) = \begin{cases} 1 & \beta_0 + \beta_1 x + \varepsilon > 0 \\ 0 & \text{Otherwise} \end{cases}$
3. At every iteration, some features will be removed from the set of all features, based on the inference of the training model.
4. A final subset of features will be generated based on the limitation of the model or selections of best features are done.
5. Ranking of selected features is done by following the sequence order of elimination of features during every iteration.

3.3 Classification

The system divides the dataset into 80–20, 70–30, and 60–40% train-test split for experimental purpose. Splitting of the dataset is done randomly without following any sequences. After partitioning, a training set of data is initially applied to the classifier. This deep neural network classifier has an input layer with four input nodes, three hidden layers with 10, 20, and 10 hidden nodes, and an output layer with a single node. Since this network has multiple layers with a huge amount of inner nodes, computationally it is expensive but provides promising results after training the model.

3.3.1 Deep Neural Network

Deep neural networks follow the structure of a typical artificial neural network with a complex network model. It helps us to create a model and define its complex hierarchies in a simple form. It has 'n' hidden layers and processes the data from the previous layer called as the input layer, and after every epoch, error rate of the input data will be gradually reduced by adjusting the weights of every node, back propagating the network and continues till reaches better results. Any number of inputs can be assigned as input nodes in input layer. Normally, number of nodes in DNN will be more than the input layer to increase the learning process intensively. Number of outputs can be defined individually as unique output nodes in output layer. The parameters used in DNN are said to be number of nodes in input and output layer, bias, learning rate, initial weights for adjustment, number of hidden layers, number of nodes in every hidden layers, and stop condition for terminating the epochs while execution. In this model, bias value is assigned as 1, which is usually assigned to be 1 in any neural network to avoid nullified network results. Also, learning rate is assigned as 0.15 by default and randomly changed later by trial and error for obtaining varying outcomes from the model. The initial weight of the nodes can be assigned randomly and changed by the network during back propagation by calculating error rate and updated periodically after every epoch. Number of hidden layers and number of nodes in every hidden layer are decided based on the number of inputs and size of the data. Termination condition of the network is said to be either the number of epochs are reached or the expected result is achieved from the learning model. If the number of layers and nodes in a network is more, it will take more time and resource to train the model.

Algorithm of Deep Neural Network

1. Define a neural network with an input layer having n input nodes.
2. Initialize the number of hidden layers needed to train the data.
3. Define the learning rate and bias value for every node. The weight will be randomly selected in initial forward propagation.
4. Define the activation function

Rectified linear unit (ReLU): $f(x) = \max(0, x)$.

5. Define the number of epochs to back propagate the value from the output node.
6. Train the network for given set of training data.
7. After the network is trained, pass the test data to the trained network to find the classification rate of the model.
8. Train the network until the number of epochs is completed (or) expected output is achieved.
9. Calculate the accuracy of the model using evaluation metrics.

4 Experimental Results

WBC dataset consists of 699 instances and 9 feature variables. The dataset supports for binary classification models, since it has only binary values as class label values, i.e., 0 for benign and 1 for malignant. But, the actual values given in the dataset for benign and malignant would be 2 and 4, respectively. To keep the system more stable, convert all the values of the class label from 2 and 4 to 0 and 1. Out of 699 instances, 16 instances contain missing value. The missing instances are removed to reduce the error rate of the system. Finally, 683 instances are used for feature selection. The description of WBC dataset is shown in Table 2.

In recursive feature elimination process, initially 9 features are given as an input to logistic regression model, which is used as learning algorithm. The objective of RFE is to select best possible subset of features from all given inputs. Initially, all the features are trained using learning algorithm and the performance of each feature is separately maintained. Then, the features with least coefficient will be omitted during every epoch and retrained after removing each feature from the input set and continued until required number of features are retained. As an outcome of RFE, single epithelial cell size, marginal adhesion, normal nucleoli, mitosis, and bland chromatin are eliminated individually during every iteration of training process. These features are removed from dataset, and selected four features are used as inputs to DNN for classification process.

The performance of a model is estimated through confusion matrix. The confusion matrix helps to find the classified and misclassified rate of the system. Effectiveness and performance of a system can be measured by calculating the accuracy. The accuracy is defined in Eq. (1).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Table 2 Description of Wisconsin Breast Cancer Dataset

Number	Attribute name	Domain	Missing values
1	Clump thickness	1–10	0
2	Uniformity of cell size	1–10	0
3	Uniformity of cell shape	1–10	0
4	Marginal adhesion	1–10	0
5	Epithelial cell size	1–10	16
6	Bare nucleoli	1–10	0
7	Bland chromatin	1–10	0
8	Normal nucleoli	1–10	0
9	Mitosis	1–10	0
10	Class	2, 4	0

where TP represents true positive, TN represents true negative, FP represents false positive, and FN represents false negative.

Sensitivity can be measured to calculate the proportion of correctly identified instances with actual positives.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

Specificity is a measure to find the proportion of correctly identified instances with actual negatives.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (3)$$

F-score is calculated to find the test accuracy of the model. To compute *f*-score, precision and recall are also calculated.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

$$F - \text{Score} = \frac{(\beta^2 + 1) * \text{precision} * \text{recall}}{\beta^2 * \text{precision} + \text{recall}} \quad (6)$$

While calculating *F*-score, it is balanced with $\beta = 1$ and here, bias value is mentioned as β . Also, when $\beta < 1$, it is favor for precision and when $\beta > 1$, it favor for recall.

Once trained the network model with training set, then test set with 20% of instances is applied to explore the accuracy rate of the classifier. As expected, this model outperforms and gives promising result of 98.62% for 80–20% partition split. Also, this system produces a better result of 97.66 and 96.88% for 70–30% and 60–40% splits, respectively. Table 3 shows the accuracy rate, and it is compared with other existing methods.

Table 3 Analysis of result with existing methods

Classification model	Accuracy	Model evaluation
EM-PCA-CART	93.2	10-fold
ART-1	93.2	70–30
LP-SVM	97.33	75–25
SVM-RBF	93.47	10-fold
Proposed method	98.62	80–20
Proposed method	97.66	70–30
Proposed method	96.88	60–40

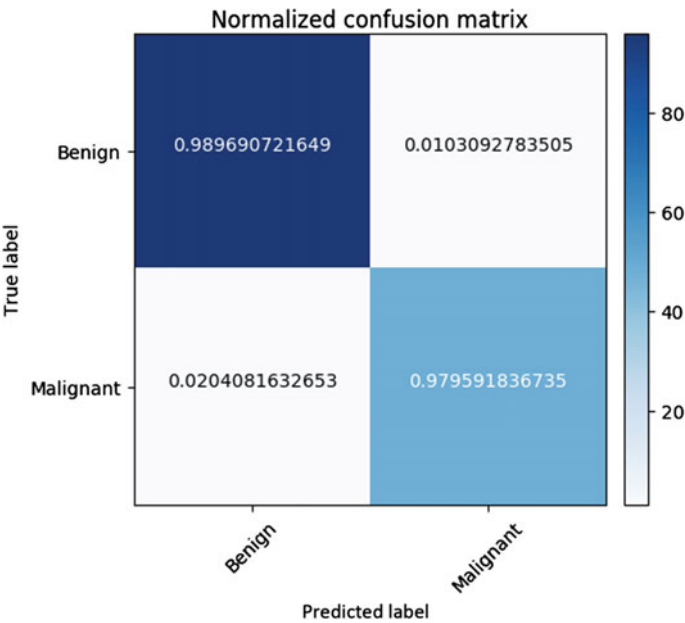
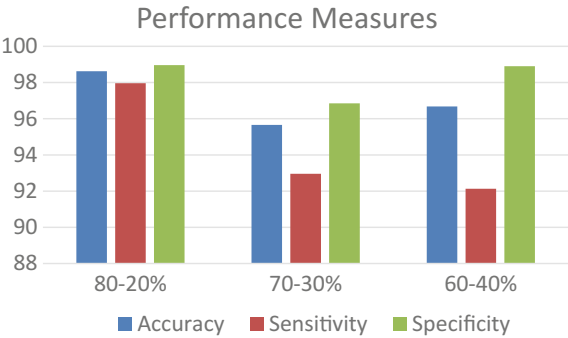


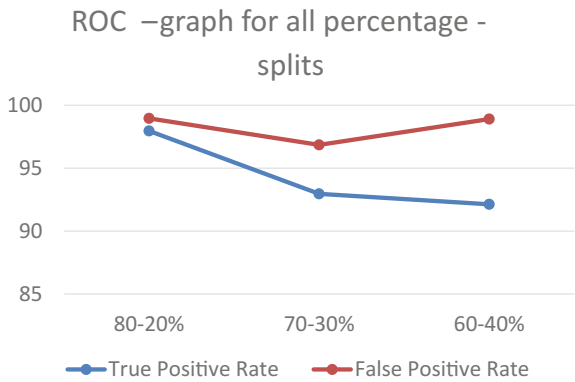
Fig. 4 Confusion matrix for the proposed system

Fig. 5 Performance for the proposed system



This system achieved better accuracy of 98.62% for 80–20% partitioning. Figure 4 helps to visualize the result of the proposed method. Figure 5 shows the performance of the method in terms of accuracy, sensitivity, and specificity with different train-test split partitioning. Figure 6 displays the ROC curve to analyze the performance of the system.

Fig. 6 Receiver operational characteristic curve



5 Conclusion

In this modern era, lots of people are facing many problems with modern age diseases. Breast cancer is one of the most common types of deadliest disease raising over time among different countries. Lack of awareness and post-identification of disease will be the major reason for more death rates. Computer-aided diagnosis will be a perfect solution for all kind of peoples to diagnose with accurate results. CAD system will not be a perfect replacement for professional doctors, but this aid will help them a lot, by assisting practitioners, to make a perfect decision by analyzing patient reports. Sometimes, practitioners may do some mistake due to lack of experience or poor analysis of reports. So, it will act as a better remedy for the current medical environment. More accurate decisions are taken, only if the model used to train the system will be unambiguous.

This system provides better results, compared with previous models, and needs a little improvement. The limitation of this system is training time of the algorithm since it has deeply trained the neural network. In GPU-incorporated systems, this system will take less amount of time compared with commercial hardware. So it is expected that the user of this system will have a computationally better system to test and process their data.

6 Future Work

The proposed system used RFE for feature selection and deep neural network system for classifying the data. It provides 98.62% of accuracy in DNN-RFE model. This system has complex multiple hidden layers with a lot of neurons in every layer, which can propagate the input and parse through multiple layers. So, in every epoch, during back-propagation process, the error rate of the system is gradually reduced by adjusting the weights of nodes and network fine tuning the

values of nodes in every layer. Since this network architecture has very complex structure, increase in training time is inevitable. In future, to improve the classification rate the researchers can perform particle swarm optimization technique or genetic algorithm for feature selection, which may increase the accuracy of the overall model, by selecting the features based on its fitness values. Deep learning algorithms are more powerful and need high-end computing resources to run for both training and testing phases of learning models, which may lack in its performance when executed in local machines due to less amount of computational power. Cloud-assisted virtual machines or parallel processing systems may be used to optimize the computational efficiency of the model. It will reduce the time to train the system and makes the system to be computationally inexpensive.

References

1. Abbass, H. A. (2002). An evolutionary artificial neural networks approach for breast cancer diagnosis. *Artificial Intelligence in Medicine*, 25(3), 265–281.
2. Abdel-Zaher, A. M., & Eldeib, A. M. (2016). Breast cancer classification using deep belief networks. *Expert Systems with Applications*, 46, 139–144.
3. Anooj, P. (2012). Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. *Journal of King Saud University-Computer and Information Sciences*, 24(1), 27–40.
4. Azar, A. T., & El-Said, S. A. (2014). Performance analysis of support vector machines classifiers in breast cancer mammography recognition. *Neural Computing and Applications*, 24(5), 1163–1177.
5. Bewal, R., Ghosh, A., & Chaudhary, A. (2015). *Journal of Clinical and Biomedical Sciences*, 5(4), 143–148.
6. Bhattacharjee, A., Roy, S., Paul, S., Roy, P., Kausar, N. & Dey, N. (2015). Classification approach for breast cancer detection using back propagation neural network: a study. *Biomedical Image Analysis and Mining Techniques for Improved Health Outcomes*, p. 210.
7. Ghosh, S., Biswas, S., Sarkar, D. C., & Sarkar, P. P. (2016). Breast cancer detection using a neuro-fuzzy based classification method. *Indian Journal of Science and Technology*, 9(14).
8. Gorunescu, F., & Belciug, S. (2014). Evolutionary strategy to develop learning-based decision systems. Application to breast cancer and liver fibrosis stadialization. *Journal of Biomedical Informatics*, 49, 112–118.
9. Huang, M. W., Chen, C. W., Lin, W. C., Ke, S. W., & Tsai, C. F. (2017). Svm and svm ensembles in breast cancer prediction. *PloS one*, 12(1), e0161501.
10. Jhajharia, S., Varshney, H. K., Verma, S., & Kumar, R. (2016). A neural network based breast cancer prognosis model with pca processed features. In: *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1896–1901.
11. Jouni, H., Issa, M., Harb, A., Jacquemod, G., & Leduc, Y. (2016). Neural network architecture for breast cancer detection and classification. In: *IEEE International Multidisciplinary Conference on Engineering Technology (IMCET)*, pp. 37–41.
12. Karabatak, M., & Ince, M. C. (2009). An expert system for detection of breast cancer based on association rules and neural network. *Expert Systems with Applications*, 36(2), 3465–3469.
13. Kiyani, T., & Yildirim, T. (2004). Breast cancer diagnosis using statistical neural networks. *Journal of Electrical and Electronics Engineering*, 4(2), 1149–1153.
14. Mert, A., Kilic, N., Bilgili, E., & Akan, A. (2015). Breast cancer detection with reduced feature set. *Computational and Mathematical Methods in Medicine*.

15. Nahato, K. B., Harichandran, K. N., & Arputharaj, K. (2015). Knowledge mining from clinical datasets using rough sets and backpropagation neural network. *Computational and Mathematical Methods in Medicine*.
16. Nayak, T., Dash, T., Rao, D. C., & Sahu, P. K. (2016). Evolutionary neural networks versus adaptive resonance theory net for breast cancer diagnosis. In: *Proceedings of the International Conference on Informatics and Analytics (ACM)*, p. 97.
17. Nilashi, M., Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). A knowledge-based system for breast cancer classification using fuzzy logic method. *Telematics and Informatics*, 34(4), 133–144.
18. Onan, A. (2015). A fuzzy-rough nearest neighbor classifier combined with consistency-based subset evaluation and instance selection for automated diagnosis of breast cancer. *Expert Systems with Applications*, 42(20), 6844–6852.
19. Paulin, F., & Santhakumaran, A. (2011). Classification of breast cancer by comparing back propagation training algorithms. *International Journal on Computer Science and Engineering*, 3(1), 327–332.
20. Prevention Control: Center for Diseases Control and Prevention (2014). URL <https://www.cdc.gov/cancer/breast/index.htm>.
21. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117.
22. Yin, Z., Fei, Z., Yang, C., & Chen, A. (2016). A novel svm-rfe based biomedical data processing approach: Basic and beyond. In: *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*, pp. 7143–7148.