Early Breast Cancer Detection Utilizing Artificial Neural Network

ZAKIA SULTANA1, 2,*, MD. ASHIKUR RAHMAN KHAN1, NUSRAT JAHAN1
1Department of Information and Communication Engineering
Noakhali Science and Technology University
Noakhali-3814, BANGLADESH

2Department of Computer Science and Engineering
Daffodil International University, Dhaka
BANGLADESH

Abstract: - Breast cancer is one of the most dangerous cancer diseases for women in worldwide. A Computer-aided diagnosis system is very helpful for radiologist for diagnosing micro calcification patterns earlier and faster than typical screening techniques. Maximum breast cancer cells are eventually form a lump or mass called a tumor. Moreover, some tumors are cancerous and some are not cancerous. The cancerous tumors are called malignant and non-cancerous tumors are called benign. The benign tumors are not dangerous to health. But the unchecked malignant tumors have the ability to spread in other organs of the body. For that early detection of benign and malignant tumor is important for confining the death of breast cancer. In these research study different neural networks such as, Multilayer Perceptron (MLP) Neural Network, Jordan/Elman Neural Network, Modular Neural Network (MNN), Generalized Feed-Forward Neural Network (GFFNN), Self-Organizing Feature Map (SOFM) Neural Network, Support Vector Machine (SVM) Neural Network, Probabilistic Neural Network (PNN) and Recurrent Neural Network (RNN) are used for classifying breast cancer tumor. And compare the results of these networks to find the best neural network for detecting breast cancer. The networks are tested on Wisconsin breast cancer (WBC) database. Finally, the comparing result showed that Probabilistic Neural Network shows the best detection result than other networks.

Key-Words: - Breast cancer, classification, support vector machine, multilayer perceptron, self-organizing map, Jordan neural network, Probabilistic Neural Network, Recurrent Neural Network.

Received: January 28, 2021. Revised: March 10, 2021. Accepted: March 12, 2021. Published: March 18, 2021.

1. Introduction
Breast cancer is the most common cancer in women. Breast cancer is the second leading cause of death among women. Breast cancer also developed in men, breast cancer in men is rare, an accounting for less than 1% of all breast cancers, only about 1 out of 1,000 men will get it at some point in their lives. The breast is made up of different breast tissue, ranging from very fatty tissue to very dense tissue. This tissue consists of a network of lobes. Each lobe is made up of tiny, tube-like structures called lobules that contain milk glands. Tiny ducts connect the glands, lobules, and lobes, carrying milk from the lobes to the nipple. The nipple is located in the middle of the areola, which is the darker area that surrounds the nipple. Blood and lymph vessels also run throughout the breast. Blood nourishes the cells. The lymph system drains bodily waste products. The lymph vessels connect to lymph nodes. Cancer that forms in the tissues of breast, usually in the ducts (tubes that carry milk to the nipple) and in the lobules (glands that make milk) is called, the breast cancer [1].

Breast cancer can develop when the genetic material in cells changes and causes them to start multiplying in an uncontrolled way. Lumps and nodules then form after some time. Why those changes come about, and how the cells develop, depends on many factors that can influence each other. The main factors include the woman’s age, hormones, and whether there is a higher risk of breast cancer in her family.
Breast cancer accounts for 1 in 6 female cancer deaths. The prevention method of the breast is not found out until now. Early detection is the main prevention method of breast cancer. The mortality rate for breast cancer will be reduce if we detect is the early stage of the cancer before spreading into the other organs of the body. But unfortunately, about 90 present breast cancer is detected in advance stage such as stages III and IV [2]. It is a very bad news for breast cancer patient.

It is important to detect breast cancer early to reduce the mortality rate, and this requires accurate and reliable diagnoses. When several tests are involved, the ultimate diagnosis may be difficult to obtain, even for a medical expert [3]. Diagnosis of breast cancer has improved during the last decades for the development of more effective diagnosis system. Several Artificial Intelligence (AI) techniques have been intensively applied to radiological assessments to predict the biopsy outcome in breast cancer [4]). These techniques include many artificial neural networks. Moreover Computer-Aided Diagnosis (CAD) could be beneficial to help radiologists in the US based detection of breast cancer, minimizing the effect of the operator-dependent nature of US imaging [5]). There are several other classification techniques to predict and classify breast cancer pattern. Classification systems, helping possible errors that can be done because of fatigued or inexperienced expert to be minimized, provide medical data to be examined in shorter time and in more detail [6].

In this study, different artificial neural networks such as multilayer Perceptron, Jordan/elman network, modular neural network, generalized feed-forward neural network, radial basis function with probabilistic neural network, self-organizing feature map, support vector machine, recurrent network are used to diagnosis the beaming and malignant cancer cell of breast. And after comparing the diagnosis result of these networks the best networks will be introduced for the early detection of the breast cancer.

In this research work, apply different neural network models for the classification of benign and malignant cell for the classification of breast cancer. This thesis study main objective is the early detection of breast cancer disease through different neural networks. This thesis classifies benign and malignant cell using different neural network accurately and make a comparison among the networks to find the best network, which classify more accurately than other networks.

2. Literature Survey

Computer aided diagnosis of breast cancer getting famous day by day. Computer aided diagnosis helps the radiologists to detect abnormalities earlier than traditional procedures. The first step for CAD is the ability to identify the abnormal masses in the breast, while the second step is to diagnosis the masses detected in the first step. There are many steps in diagnosis of breast cancer, start from the segmentation and ended by tumor classification. The diagnosis requires precise and reliable diagnosis to ensure that doctors can distinguish between benign and malignant tumors. According to American Cancer Society, some studies have shown that CAD can help find cancers that radiologist otherwise might have missed. The use of classifier systems in medical diagnosis is increasing gradually. There is no doubt that evaluation of data taken from patients and decisions of experts are the most important factors in diagnosis. However, expert systems and different artificial intelligence techniques for classification also help experts in a great deal.

Automatic diagnosis of breast cancer is an important real world medical problem. Researchers made many attempts to efficiently use different neural networks to improve the diagnosis efficiency in breast cancer detection.

A method was implemented using Back Propagation Algorithm (BAP), Radial Basis Function Networks (RBFN), and Competitive Learning Network (CLN) and Learning Vector Quantization (LVQ) classifier network to detect breast cancer disease on the Wisconsin Breast Cancer Dataset (WBCD). After evaluating these four networks they found that LVQ is the best classifier neural network to detect breast cancer. And the order of the accuracy CLN, MLP, RBFN. The accuracy of the LVQ is 95.82% [7].

A system was proposed for breast cancer detection (classification of benign and malignant) using back-propagation neural network on the Wisconsin breast cancer dataset (WBCD). This proposed method results were compared with radial basis function network and found that for detecting breast cancer back-propagation neural network is best. By increasing the number of neurons in hidden layer the accuracy is improved, neural network with 9 neurons in hidden layer provide 99% accuracy [8]. A comparative study of support vector machine (SVM), Bayesian classifier and other artificial neural network classifier (Back-propagation, learning vector quantization, and K nearest neighbourhood) was provided on the Wisconsin breast cancer Fine Needle Aspiration (FNA) biopsy data set. The use of machine learning technique cancer cells is detected.
early with previously. By comparing the result K nearest neighbourhood provide prediction accuracy 100%, SVM using the RBF kernel function provide 98.24% and K2 Bayesian network provide poor accuracy by comparing all the networks evaluated. And the Bayesian network shows a low prediction time of 0.07 seconds than other networks evaluated [9].

A method was proposed for detecting breast cancer using modified imperialist competitive algorithm (MICA), K-means algorithm and different neural network and working on the Wisconsin Breast Cancer database (WBCD). This method includes clustering module and classification module. In the clustering module the input data are clustered by the combination of modified imperialist competitive algorithm (MICA), K-means algorithm and the Euclidean distance of each pattern is examined. In the classifier method the multilayer perceptron (MLP), probabilistic neural networks (PNN) and the radial basis function neural networks (RBFNN) determines the membership of the patterns using examined distance. Among various neural networks MLP with back-propagation algorithm provide 97.44% accuracy of system using only WBC database and with the Euclidean distance of patterns from cluster center which are find from the K-MICA, RBFNN’s provide the accuracy of system 99.11%. The performance of classifier also compared by K-mean clustering, subtractive clustering and genetic algorithm clustering. From the comparison result shows that K-MICA clustering based system achieves higher recognition accuracy of 99.11% [10].

A more cost effective and easy to use system was developed for classifying breast cancer disease tumor using neural network with feed-forward back-propagation algorithm to classify the tumor that causes cancer by working on the Wisconsin breast cancer database (WBCD). Many neural networks model were created and trained using different number of neurons in hidden layer and by the investigation of the networks they find that neural network model with 7 node in the hidden layer acquire the highest accuracy than others. The result of the feed-forward back-propagation algorithm was compared with support vector machine (SVM) and decision tree classifier and showed that feed-forward back-propagation algorithm is the best classifier than others [11].

A hybrid evolutionary neural network classifier (HENC) combining the evolutionary algorithm, which has a powerful global exploration capability, with gradient-based local search method, which can exploit the optimum offspring to develop a diagnostic aid that accurately differentiates malignant from benign pattern using WBCD dataset was implemented. After examination the HENC approach can obtain better generalization and much lower computational cost than the existing methods. This proposed approach had an average accuracy more than 98%. A CAD method also developed using the HENIC approach to make the classifier interesting and easy to apply in the real world [4].

A hybrid machine learning method using fuzzy-artificial immune system with k-nearest neighbour algorithm for diagnosing breast cancer was proposed by classifying Wisconsin Breast Cancer Dataset (WBCD). After using this method, the highest classification accuracy of 99.14%. This classification accuracy was obtained by 10-fold cross validation [6].

A Computer Aided Diagnosis algorithm was developed for identifying breast nodule malignancy using multiple ultrasonography (US) feature and artificial neural network (ANN) classifier was developed from a database of 584 histologically confirmed cases containing 300 benign and 284 malignant breast nodules. The features determining whether a breast nodule is benign or malignant were extracted from US images through digital image processing with a relatively simple segmentation algorithm applied to the manually preselected region of interest. An ANN distinguished malignant nodules in US images based on five morphological features representing the shape, edge characteristics, and darkness of a nodule. The structure of ANN was selected using fold cross-validation method with = 10. The ANN trained with randomly selected half of breast nodule images showed the normalized area under the receiver operating characteristic curve of 0.95. With the trained ANN, 53.3% of biopsies on benign nodules can be avoided with 99.3% sensitivity. Performance of the developed classifier was re-examined with new US mass images in the generalized patient population of total 266 (167 benign and 99 malignant) cases [13].

A hybrid system for diagnosis breast cancer tumours using Wisconsin breast cancer database (WBCD) was developed. The system has three main modules, such as feature extraction module, training module and classifier module. In training module, they used a hybrid bees’ algorithm (BA) - back-propagation (BP) algorithm is proposed to train the classifier and multi-layer perceptron (MLP) neural network is used for train classification. After analysing this proposed system has high diagnosing accuracy than other method and the accuracy 95.86%. And for increasing the accuracy of recommended system, fuzzy feature used as the input
of optimized classifier that significantly improves the accuracy to 97.83% [14].

A method was proposed using Support Vector Machine (SVM) combine with feature section for diagnosis breast cancer using Wisconsin breast cancer database (WBCD). The performance of this model is evaluated using classification accuracy, sensitivity, specificity, positive and negative values, receiver operating characteristic (ROC). The result showed that the highest classification accuracy is 99.51% is obtained using SVM with these five features [15].

The use of deep learning approaches was proposed for breast ultrasound lesion detection and investigates three different methods: a Patch-based LeNet, a U-Net, and a transfer learning approach with a pretrained FCN-AlexNet. Their performance is compared against four state-of-the-art lesion detection algorithms (i.e. Radial Gradient Index, Multifractal Filtering, Rule-based Region Ranking and Deformable Part Models). A comparing and contrasting two conventional ultrasound image datasets acquired from two different ultrasound systems. Dataset A comprises 306 (60 malignant and 246 benign) images and Dataset B comprises 163 (53 malignant and 110 benign) images. To overcome the lack of public datasets in this domain, Dataset B will be made available for research purposes. After analyzing the Transfer Learning FCN-AlexNet achieved the best results for Dataset A and the proposed Patch-based LeNet obtained the best results for Dataset B in terms of FPs/image and F-measure [5].

The statistical neural network structures, radial basis network (RBF), general regression neural network (GRNN) and probabilistic neural network (PNN) was used on the Wisconsin breast cancer data (WBCD). Then compare the statistical neural networks result with the result of Multilayer Perceptron (MLP) neural network on the WBCD database. Finally they found that for the test set GRNN gives the best classification accuracy [3].

A new method for detecting the breast cancer more accurately was proposed and in this methods first use the image processing techniques to prepare the mammography images for the feature and pattern extraction process then used this features as input for the Back Propagation Neural Network (BPNN) and the Logistic Regression (LR) machine learning supervised algorithm. After analyzing the result of this two algorithm they find that, the number of features utilized in LR model was much higher than BPNN. A good regression value using BPNN that exceeded 93% with 204 features [12].

It is apparent from previous studies that complex neural networks approaches are applied for detecting breast cancer and more research is needed to improve the accuracy of early detection of breast cancer. This research works for finding early detection of breast cancer with more accuracy.

3. Methodology

The aim of this research is to define the best artificial neural network (ANN) methods to classify the breast cancer cell for detecting breast cancer in the early stage. Breast cancer microscopic and clinical tests reports are collected. Based on the tests reports the diagnostic system will predict either Benign (not cancer) or Malignant (cancer). There are different networks like multilayer perceptron (MLP), Jordan/Elman network, modular neural network, generalized feed forward network, self-organizing feature map, support vector machine, recurrent network.

3.1 Parameters Considered

All the networks were trained, tested and validated using a variety of parameters. Each of the networks was tested using many combinations of hidden node and epoch number. But the hidden layer, activation function and learning rule of all the networks were same. After a number of experimentation and parameter alterations using theoretical understandings and best practices, the optimal classifier was identified.

The dataset in this research was obtained from the UCI Irvine machine learning repository [16]. This dataset was originally created by Dr. Wolberg, Street and Mangasarian all from University of Wisconsin. Data items in the dataset are composed of ID number, the diagnosis which will either be classified as malignant (M) or benign (B) and numeric shape features of extract cellular nuclei such as radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, and symmetry and fractal dimension. These data were extracted from preprocessed Fine Needle Aspiration biopsy image of cell slides. The dataset was composed of a total of 569 observations with benign and malignant cases and the number of benign 357 and the number of malignant 212 observations. And benign will be detected by “0” and malignant will be detected by “1”

Attribute Information:
1. ID number
2. Diagnosis (M = malignant, B = benign)

Ten real-valued features are computed for each cell nucleus:
a. radius (mean of distances from center to points on the perimeter)  
b. texture (standard deviation of gray-scale values)  
c. perimeter  
d. area  
e. smoothness (local variation in radius lengths)  
f. compactness (perimeter^2 / area - 1.0)  
g. concavity (severity of concave portions of the contour)  
h. concave points (number of concave portions of the contour)  
i. symmetry  
j. fractal dimension (“coastline approximation” - 1)

In the network structures there are ten nodes in the input layer, on the basis of breast cancer dataset, j nodes in the hidden layer, and one nodes in the output layer or diagnosis result benign (B) or malignant (M). Thus, the network structure can be defined as 10-j-1.

![Figure 1: Structure of neural network](image)

<table>
<thead>
<tr>
<th>Artificial Neural Network Model</th>
<th>Number of hidden layer and elements / Configuration</th>
<th>Epoch number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multilayer Perceptron (MLP)</td>
<td>one hidden layer with seven processing elements</td>
<td>40,000</td>
</tr>
<tr>
<td>Jordan/Elman network</td>
<td>one hidden layer with four processing element</td>
<td>20,000</td>
</tr>
<tr>
<td>Modular neural network</td>
<td>two hidden layer, one hidden layer with six processing element another hidden layer with four processing element</td>
<td>40,000</td>
</tr>
<tr>
<td>Generalized feed forward network</td>
<td>one hidden layer with seven processing element</td>
<td>30,000</td>
</tr>
<tr>
<td>Self-organizing feature map</td>
<td>one hidden layer with four processing element. Square kohonen with radius of the cluster is 1. For unsupervised learning is 20,000 and for supervised learning the epoch number is 40,000</td>
<td></td>
</tr>
<tr>
<td>Support vector machine</td>
<td>Compose without regression</td>
<td>20,000</td>
</tr>
<tr>
<td>Recurrent network</td>
<td>one hidden layer with four processing element</td>
<td>50,000</td>
</tr>
<tr>
<td>Probabilistic neural network</td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

From the above table we can say that the network structure of modular neural network is 10-6-4-1 & the network structure is 10-4-1 for jordan/elman, self-organizing feature map, and recurrent neural network. The network structure can be defined as 10-7-1 for both multilayer perceptron, generalized feed forward neural network.
3.3 Performance Analysis (Training, Testing & Validation)

Once a network has been structured for a particular study, that network has been trained. After the appropriate training, the selected network has the ability to interconnect one value of output to a given input [18]. For starting this process, the initial weights are chosen randomly and then, the training or learning begins. For developing an appropriate network, the networks had been trained many times.

The study of detecting breast cancer used 60% of total dataset for training approach, 30% of total dataset for testing approach, 10% of total dataset for validation approach.

Based on the training testing process on the data set, networks will predict the target value or classify the benign or malignant. But the prediction of these network may contain error. The error is computed with the equation

\[ \text{Error} = E = \frac{1}{2} (t - y_0)^2 \]  

where,

- \(t\) = target value
- \(y_0\) = the neural network output
- \(E\) = error

The actual error of the network is calculated by running all the training cases and testing cases. The performance measure parameters of all networks are mean square error (MSE), the \(r\) value and the diagnosis accuracy of the networks output. Other parameter minimum absolute error (min abs error), maximum absolute error (max abs error) will be showed in the result section.

4. Result Analysis

The aim of the project is to explore the best network among different neural networks for nearest detection of breast cancer. The input data can be used to train, test and validate the different neural networks with the uniform activation function and learning rate. These yielding 10 input data features have different impact for detecting cancer. Some features are most useful in predicting malignant or benign cancer. Figure 2 shows the impact of important features with important scores by applying the random forest model [17]. All the networks were run only once. These networks will be analysed in each training testing and validation part and finally the networks will be discussed and compared for finding the best network, which accurately detect the breast cancer.

Training, testing and validation result for all used networks are shown in Table 2. It shows the best epoch with minimum MSE and the final MSE of the training error at last epoch, testing result shows the mean square error (MSE), minimum absolute error (Min Abs Error), maximum absolute error (Max Abs Error) and the linear correlation coefficient (\(r\)) value. The \(r\) value explained the variation between the neural network outcomes and desired outcomes. The perfect value of \(r\) is 1. The accuracy result for detecting best cancer depends on the validation results of the networks.

From Table 2, it is noticed that probabilistic neural network shows the minimum training MSE 0.002376483. Generalized feed-forward neural network produce the second minimum training MSE 0.011683897. Multilayer perceptron produces the third minimum training MSE 0.016001718 and self-organizing feature map produce the fourth minimum training MSE 0.017059825. From the table 2 it is also noticed that recurrent neural network shows the minimum testing MSE 0.053405378 and shows the height linear correlation coefficient (\(r\)) 0.889811535. Modular neural network produce the second minimum testing MSE 0.075174462 and second highest correlation coefficient (\(r\)) 0.838853385. Probabilistic neural network shows the third lowest testing MSE 0.081914184 and third height liner coefficient 0.830980024. MLP Testing MSE 0.099679093 and linear correlation coefficient is 0.80010834. Self-organizing Feature map produce testing MSE 0.103715541 and correlation coefficient is 0.78119486.
Table 2. Training, testing and validation result of all used neural networks

<table>
<thead>
<tr>
<th>Neural Network</th>
<th>Training result</th>
<th>Testing result</th>
<th>Validation Result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Epoch</td>
<td>Min. MSE</td>
<td>Final MSE</td>
</tr>
<tr>
<td>Multilayer Perceptron Neural Network</td>
<td>10930</td>
<td>0.00149437</td>
<td>0.01600171</td>
</tr>
<tr>
<td>Jordan and Elman Neural Network</td>
<td>3344</td>
<td>0.01229998</td>
<td>0.03598988</td>
</tr>
<tr>
<td>Modular Neural Network</td>
<td>19651</td>
<td>0.02186522</td>
<td>0.037517146</td>
</tr>
<tr>
<td>Generalized Feed-Forward Network</td>
<td>30000</td>
<td>0.01168549</td>
<td>0.0168549</td>
</tr>
<tr>
<td>Probabilistic Neural Network</td>
<td>2</td>
<td>0.002376483</td>
<td>0.01636263</td>
</tr>
<tr>
<td>Self-Organizing Feature Map</td>
<td>37</td>
<td>0.01229998</td>
<td>0.03598988</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>2933</td>
<td>0.033648998</td>
<td>0.033648126</td>
</tr>
<tr>
<td>Recurrent Neural Network</td>
<td>757</td>
<td>0.24878252</td>
<td>0.119278819</td>
</tr>
</tbody>
</table>

The detection or diagnosis accuracy of all used networks shown in table 2 based on the validation result, probabilistic neural network accuracy for detecting breast cancer is 98.24%. SOMF network accuracy is also 98.24%. The second height accuracy 96.5% of support vector machine. MLP network accuracy for detection is 91.23%. Jordan/Elman and recurrent neural network accuracy is 93%.

To investigate accuracy of the networks use the line pattern of the data between the target values and neural networks outputs. The target values and the neural networks outputs are generated on the same graph. In this way the line pattern of all neural networks are shown in figure 3 and figure 4. Figure 3 and figure 4 shows that the predicted and experimental values are how close to each other for all networks in testing and validation step respectively.
Figure 3: Diagnosis result of testing with respect to exemplar in different neural network (a) MLP Network, (b) Jordan/Elman Network, (c) modular network, (d) GFFNN, (e) PNN, (f) SOFM network, (g) SVM network, and (h) Recurrent Neural Network.
Figure 4: Accuracy result with respect to exemplar in different neural network (a) MLP Network, (b) Jordan/Elman Network, (c) modular network, (d) GFFNN, (e) PNN, (f) SOFM network, (g) SVM network, and (h) RNN

Based on the result, probabilistic neural network produce more accurate result than others. The accuracy of this network is 98.24% and the r value of this neural network is 0.830980024. SOMF network accuracy is also 98.24% but the r value is 0.78119486. Recurrent neural network shows the highest r value 0.889811535 but the accuracy of the network is 93%.

5. Conclusion

Breast cancer is the second leading cause of cancer deaths worldwide. According to the statistical data of breast cancer in the world, this cancer disease is among the most comprehensive cancer types. But the cancer type is also among the most curable ones if it
can be diagnosed early. It is important to detect breast cancer early to reduce the mortality rate, and this requires accurate and reliable diagnoses.

In this paper, an automatic diagnosis system for early detection of breast cancer based on artificial neural network is proposed. In this paper a number of artificial neural networks was presented for the classification of benign and malignant. This paper presented multilayer neural network, Jordan/Elman network, modular neural network, generalized feed-forward network, self-organizing feature map network, support vector machine neural network, recurrent network, and probabilistic neural network for the classification. Therefore from the overall discussion it showed that the probabilistic neural network with 98.24% accuracy and 0.830980024 r value emerged the best prediction result for detecting breast cancer than other networks and second best is the Self organizing feature map.

5.1 Future Work
Considering the initial, empirical nature of the work done for this dissertation, the results are also informative with respect to potential directions for future work that are likely to yield valuable results. The tests focused on the accuracy of disease detection. However, additional tests on other suitably annotated data sets can reveal the accuracy of the method in detecting each type of tissue. Thus, reliable information on the distribution of such tissue and perhaps even a technique to further investigate such tissue more thoroughly could offer additional useful diagnosis help. Another direction for future work is, focus on the examination of other type of cancer for the early detection of the cancer using neural networks.

References:


[17] Raul Eulogio, “predicting Breast Cancer (Wisconsin Data Set) using R,” January 26, 2018


Creative Commons Attribution License 4.0
(Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US