Weight Optimization Model based on Neural Network for Breast Cancer Prediction

1K. Preetha and 2S.K. Jayanthi

1Department of Computer Science,
Vellalar College for Women,
Erode, India.
kpreetha.msc@gmail.com

2Department of Computer Science,
Vellalar College for Women,
Erode, India.
jayanthiskp@gmail.com

Abstract

Mammography is the process of using low-energy x-rays to examine the human breast for diagnosis and screening. The goal of mammography is to save lives by finding breast cancer as early as possible when they are more treatable. The aim of this paper is to classify the mammogram images as either normal or abnormal patterns, which helps the radiologists in reducing the false positive and false negative rate in mammography, hence biopsy is avoided. Mammogram images are pre-processed using MAX_AVM filter. Texture features are extracted and used as inputs to the classifier. Weight optimization model based on Neural Network helps to classify the mammogram images as normal or abnormal. Experiments were conducted on MIAS database. The result shows that the weight optimization model based on Neural Network achieved the maximum classification accuracy when compared to SVM and KNN classifier.

Key Words: Breast cancer, mammogram, neural network, genetic algorithm, SVM, KNN.
1. Introduction

Cancer is the abnormal, uncontrollable, continuous replication of cells which will inevitably lead to the formation of a tumor. Breast Cancer is a malignant tumor that starts in the cells of the breast. It mostly occurs in women, but men can get breast cancer too. Diagnostic tests and procedures for the detection of breast cancer include breast self-exam, mammograms, ultrasound, breast MRI scan and biopsy. According to 2017, breast cancer is the most common form of cancer in India, having overtaken cervical cancer. Almost 50% of all cases are in the age group of 25 to 50 and more than 70% of the cases in advanced stage had poor survival and high mortality. Breast cancer is a treatable disease and chances of survival are higher if it is detected in time.

Mammography is the preferred screening examination for breast cancer. During Mammography the breast is exposed to a small dose of ionizing radiation that produces an image of the breast tissue. Mammogram shows tiny clusters of calcium called micro-calculifications. Each pixel of the mammogram corresponds to the class either normal or abnormal. The main objective of this research is to classify the mammogram image as normal or abnormal pattern. Various image processing steps like pre-processing, feature extractions are carried out along with machine learning techniques to classify the mammogram images.

This paper is organized as follows: Section 2 briefly discusses the existing classification techniques, Section 3 describes the methodology, Section 4 discloses the experimental results and performance evaluation and section 5 states the conclusion of the work.

2. Literature Review

Currently, breast cancer detection is a challenging issue for women. Breast cancer is curable if it is detected at an initial stage. Many researchers have contributed their ideas for early detection of breast cancer.

Y.Ireaneus Anna Rejani et.al [8], have stated that mammogram images are classified as normal or abnormal based on morphological features using support vector machine. This method is tested on 75 mammogram images from MIAS database and achieved 88.75% of sensitivity.

Mohamed J.Islam et.al [6], presented an efficient computer-aided mass classification method for digitized mammogram using Artificial Neural Network. Statistical features are extracted to increase the effectiveness and efficiency of the classification process. The results proved that 90.91% of sensitivity, 83.87% of specificity and 77% of accuracy achieved is better when compared to the results given by radiologists.

Naveen Chandra Yadav et al [14], stated that feed forward neural network with back propagation learning algorithm can be used for training the classifier.
Experimental results using MIAS database of the research achieved maximum classification accuracy of 96.34% to predict the breast cancer.

Sertan Kaymak et al [15], have proposed a method for automatic classification of breast cancer using Back Propagation Neural Network (BPPN) and Radial Basis Neural Networks (RBFN). The classification performance of mammogram images using Back Propagation Neural Network achieved is better when compared to Radial Basis Neural Networks. The accuracies of BPPN and RBFN are 70.4% and 59.0% respectively.

The overall literature survey indicates that various classification techniques are applied to classify the images as normal or abnormal. Based on the study, an efficient classifier has been proposed to improve the accuracy of classification using a weight optimization model based on Neural Network.

3. Proposed Methodology

The proposed system consists of three phases namely pre-processing, feature extraction and classification to classify the mammogram images as normal or abnormal. The overview of the proposed methodology is depicted in Fig.1.

Fig 1: Diagrammatic Representation of the Proposed Method
**Pre-Processing**

The mammogram images may contain noises such as gaussian, speckle, salt and pepper and poisson which reduce the classification accuracy of the cancer detection. Hence de-noising is performed to improve the accuracy of classification and also to enhance the quality of the image. The aim of pre-processing is to suppress unwanted distortions. The image quality is enhanced by using MAX_AVM filter [5]. Resultant images obtained after pre-processing is used for feature extraction.

**Feature Extraction**

Feature Extraction plays an important role in the classification of mammogram images. It is the process by which certain features of interest within an image are detected and used for further processing of the mammogram image.

Texture feature is used to identify the cancer region in the mammogram image. The texture is an entity consisting of a group of mutually related pixels in an image. Texture features are extracted using Gray-level co-occurrence matrix (GLCM) and Gray level run length matrix (GLRLM).

GLCM is the second order statistical method of examining the texture of an image that considers the spatial relationship between two pixels. The GLCM functions characterize the texture of an image by calculating how often a pair of the pixel with gray-level value $i$ occur horizontally, vertically, or diagonally to adjacent pixels with the value $j$ [7]. After creating the GLCM, several texture features derived from the images like contrast, correlation, homogeneity and energy are calculated on the co-occurrence matrix. Contrast returns a measure of the intensity between a pixel and its neighbor over an image. Homogeneity is the measure of closeness of the distribution of elements in the GLCM. Correlation is the measure of an interrelated pixel to its neighbor in an image. Energy is the sum of squared elements in the GLCM. It is also known as uniformity.

GLRLM is a way of extracting higher order statistical features in mammogram image. GLRLM is the set of continuous pixels having same gray level. The run length is the number of neighboring gray levels in particular direction. It is calculated by counting the number of times the corresponding run occurs in the image. In the gray level run length matrix $P(i,j \mid \theta)$, the $(i,j)^{th}$ element describes the number of runs with gray level $i$ and length $j$ occur in the image along angle $\theta$ [7]. The gray level run length can be defined as,

$$
\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j \mid \theta) \quad \text{and} \quad 1 \leq N_Z(\theta) \leq N_P
$$

Where, $N_g$ be the number of intensity values in the image

$N_r$ be the number of run lengths in the image
\( P(i, j | \theta) \) run length matrix for an arbitrary direction \( \theta \)

\( N_Z(\theta) \) be the number of runs in the image along angle \( \theta \)

\( N_P \) be the number of pixels in the image

Once the features are extracted, they are given as input to the classification.

**Classification**

Classification is performed by a wide variety of supervised machine learning algorithms like SVM, KNN and weight optimization model based on Neural Network have been used to classify the mammogram images as normal or abnormal.

**Support Vector Machine (SVM)**

SVM classifier is trained and tested using extracted features and target values as an input to classify the mammogram images. It is the process of segregating the normal and abnormal classes with a hyper plane. The goal is to maximize the distance between the hyper plane to the nearest data point of either class. SVM model predicts the class as normal or abnormal.

**K-Nearest Neighbor (KNN)**

K-Nearest Neighbor is proposed for performing the pattern classification task by Fix and Hodges in 1951. KNN is a non-parametric supervised learning technique, it captures information of all training cases and classifies new cases based on similarity. The step by step procedure to compute the K-Nearest Neighbor algorithm:

- Extract features and target values of N number of training images are given as input
- Calculate Euclidean distance (E) between the feature values of training image \( (x_i) \) and testing images \( (y_j) \) by using, and arrange it in ascending order.

\[
E = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2}
\]

- Identify the class which occurs maximum number of times in first k rows (k=2) in sorted array.

**Weight Optimization based on Neural Network**

A Genetic algorithm is a search-based optimization technique based on the principles of genetics and natural science. It has been used to solve optimization problems in machine learning to classify the breast cancer. The neural network is trained by adjusting the weights, to predict the correct class. In this method, the weights are optimized using a genetic algorithm to improve the accuracy of classification.

Weight Optimization is performed as follows:
Step 1: Extract features and set target values

Step 2: Assign the value for number of hidden neurons

Step 3: Optimize the weight by using following steps:
- Generate random population of size $N$ based on hidden neurons and size of feature
- Evaluate the fitness value using

$$
E = \sum_{i=1}^{N} (x_i - y_i)^2
$$

where $x_i$ can be represented as,

$$
x_i = \frac{\text{sum}(F) \times 1}{1 + \exp(-S)}
$$

$$
F = \text{sum(chromosome_value \times weight)}
$$

$$
S = \text{sum(Extracted features \times weight)}
$$

$$
\text{weight} = \text{size(Extracted features)} + 1
$$

$$
y_i = \text{target value}
$$

- Perform crossover and mutation operations to generate new population

Step 4: Repeat the step 3 until the population of size $N$ to identify the optimum weight

Step 5: Calculate the feature values of the test image and predict the class.

From the result, it proves that the weight optimization model based on Neural Network shows better performance compare with other classifiers.

4. Experimental Result Analysis and Discussion

The research uses the data set obtained from MIAS [3]. The set consists of 322 images that fall into one of the following classes: 67 benign, 54 malignant and 201 normal images. GLCM and GLRLM features are extracted from the pre-processed mammogram images. The extracted feature values are trained by weight optimization model based on Neural Network with twenty numbers of hidden neurons to classify the mammogram image as normal or abnormal.

The metrics like Sensitivity or True Positive Rate (TPR), Specificity or True Negative Rate (TNR) and Accuracy (AC) are used to evaluate the performance of the classifier. The formulas for the above values are given in Table 1. Sensitivity and specificity values indicate the proportion of positive and negative cases respectively. Classification accuracy rate shows the number of samples that are correctly classified.
Table 1: Formula for Measures

<table>
<thead>
<tr>
<th>Measures</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>( SE = \frac{TP}{TP+FN} )</td>
</tr>
<tr>
<td>Specificity</td>
<td>( SP = \frac{TN}{TN+FP} )</td>
</tr>
<tr>
<td>Accuracy</td>
<td>( AC = \frac{(TP+TN)}{(TP+FP+TN+FN)} )</td>
</tr>
</tbody>
</table>

where TP is the number of true positives; FP is the number of false positives; TN is the number of true negatives; FN is the number of false negatives. The Confusion matrix is shown in Table 2.

Table 2: Confusion Matrix

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
</tr>
<tr>
<td>Positive</td>
<td>TP</td>
</tr>
<tr>
<td>Negative</td>
<td>FN</td>
</tr>
</tbody>
</table>

TP - Predicts abnormal as abnormal.
FP - Predicts abnormal as normal.
TN - Predicts normal as normal.
FN - Predicts normal as abnormal.

The extracted feature values are tested by using a weight optimization model based on Neural Network, SVM and KNN classifier.

The SVM and KNN classifier are implemented using classification Learner APP in Matlab 2015b.

The Confusion Matrix for SVM, KNN and weight optimization model based on Neural Network is depicted in Table 3 to 5.

Table 3: Confusion matrix for SVM Classifier

<table>
<thead>
<tr>
<th>Method (SVM)</th>
<th>Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abnormal</td>
<td>Normal</td>
</tr>
<tr>
<td>Abnormal</td>
<td>46(TP)</td>
<td>0(FP)</td>
</tr>
<tr>
<td>Normal</td>
<td>8(FN)</td>
<td>177(TN)</td>
</tr>
</tbody>
</table>

Table 4: Confusion matrix for KNN Classifier

<table>
<thead>
<tr>
<th>Method (KNN)</th>
<th>Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abnormal</td>
<td>Normal</td>
</tr>
<tr>
<td>Abnormal</td>
<td>44(TP)</td>
<td>2(FP)</td>
</tr>
<tr>
<td>Normal</td>
<td>3(FN)</td>
<td>182(TN)</td>
</tr>
</tbody>
</table>
Table 5: Confusion matrix for Weight optimization model based on Neural Network

<table>
<thead>
<tr>
<th>Method</th>
<th>Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abnormal</td>
<td>Normal</td>
</tr>
<tr>
<td>Weight Optimization Model based on Neural Network</td>
<td>46(TP)</td>
<td>0(FP)</td>
</tr>
<tr>
<td>Normal</td>
<td>3(FN)</td>
<td>182(TN)</td>
</tr>
</tbody>
</table>

Table 6: Performance comparison between different Classification techniques

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>85.1%</td>
<td>100%</td>
<td>96.5%</td>
</tr>
<tr>
<td>KNN</td>
<td>93.6%</td>
<td>98.3%</td>
<td>97.8%</td>
</tr>
<tr>
<td>Weight Optimization Model based on Neural Network</td>
<td>93.9%</td>
<td>100%</td>
<td>98.7%</td>
</tr>
</tbody>
</table>

To evaluate this research work, SVM, KNN and weight optimization model based on Neural Network classification techniques are implemented and trained with 70% of mammogram images and tested with 30% of mammogram images (46 malignant and 185 normal). Results shown in Table 6 prove that the weight optimization model based on Neural Network yields high classification accuracy compared to SVM and KNN classifiers.

5. Conclusion and Future Work

In this research, machine learning classification techniques are implemented to provide more reliable and consistent performance accuracy. From the experimental results, it is proved that SVM and weight optimization model based on Neural Network classifier achieves 100% specificity when compared to KNN and also 98.7% accuracy is achieved by weight optimization model based on Neural Network other than SVM and KNN classifier. Weight Optimization model based on Neural Network is used to classify the mammogram image as normal or abnormal pattern. In future work, the severity level of an abnormal mammogram image can be determined for early diagnosis of breast cancer, which helps the doctor to take the necessary steps for treatment.
References


