Classification of Breast Lesions using Modified Masood Score and Neural Network

Sonali Nandish Manoli¹, Anand Raj Ulle¹, N.M. Nandini² and T.S. Rekha²

¹Department of Information Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, India.
²Department of Pathology, JSS Medical College, Affiliated to JSS University Mysuru, India.
*Corresponding author E-mail: sonalinandish@gmail.com

http://dx.doi.org/10.13005/bpj/1544

(Received: 31 May 2018; accepted: 30 July 2018)

In this paper, we propose a novel method to classify Breast Lesions based on minute changes in the cell and nuclear features of the cell. It is important to note these changes as they play a significant role in diagnosis and the line of treatment by an oncologist. To overcome the problem of inter-observer variability the method of scoring is used to grade the lesions considered for the study. We have used the Modified Masood Score and designed an algorithm which classifies a given breast lesion into 6 classes namely Benign, Intermediate class-1, Intermediate class-2, Malignant class-1, Malignant class-2 and Malignant class-3. We have developed a sensitive model using the feed-forward neural network and Pattern Network to achieve the above objective. The Rank of the features is observed using ReliefF Algorithm.

Keywords: Computer Aided Diagnosis, Machine Learning, Neural Networks, Performance Function, Training Function, Modified Masood Score, Feed-Forward Neural Network, Pattern Network, ReliefF Algorithm.

Breast lesion is an extra growth or lump in the tissues of the Breast¹. There is a necessity to diagnose the condition of the breast lesion because it is estimated that 13.4% of the women born today will be diagnosed with cancer at some stage in their life². For better diagnosis the intervention of machines is necessary as they help in removing human errors caused by fatigue, oversight and inter-observer variability. To develop such a system computer-aided diagnosis plays a major role. For breast cancer detection the earliest systems where developed using supervised machine learning approaches by classifying the lesion into Benign(Non-Cancerous) and Malignant (cancerous) condition³–⁴. To improve the detection of breast lesion it is also important to look into samples which progress from the benign lesion towards malignancy. It gives rise to intermediate condition which is neither Benign nor Malignant.

The detection of such lesions can be achieved under cytology by a cost effective method called as the Fine Needle Aspiration Cytology⁵. Under Medical domain, grading system has been introduced to give a range for the observation of the characteristic features and if a given sample falls under this range it is classified into one condition of the breast lesion.

Shahla Masood in⁶ has proposed multi-class classification proposing the Masood score. The score has been modified and validated to improve the grading system for more accurate diagnosis⁶–⁸. Among the two scoring systems namely Masood Score and the Modified Masood Score it is proved that Modified Masood Score is...
better for more accurate classification of breast lesions\textsuperscript{9}. Hence Modified Masood Score has been considered in the study.

From\textsuperscript{10} we can observe that the Malignant condition can be classified further into grade -I , grade-II, grade-III. This change under malignant condition occurs to due to minor changes in the characteristic features. Each condition has a different line of treatment but if untreated it may cause progress of the disease, reoccurrence or even death. The machine which is sensitive cannot be effectively obtained using conventional classifiers when the data is divided into train and test data set.

To handle this problem, sensitive machine is required for classification. Hence we have considered a Feed Forward Neural Network and Pattern Network where both have one hidden layer in our study. The network can be trained to classify the input features to a particular class by setting the targets as outputs\textsuperscript{11-12}. In our system we have used linear regression to perform the classification of the breast lesion samples. To observe the classification accuracy various algorithms to train the neural network are used namely: Levenberg-Marquardt, Bayesian Regularisation, BFGS Quasi newton, Reselient Back-Propagation, Scaled-conjugate Gradient, Conjugate Gradient with powell restart, Conjugate Gradient with bealle restart, Fletcher-Powell Conjugate Gradient, Polak-Ribière Conjugate Gradient, One Step Secant, Variable Learning Rate Gradient Descent, Gradient Descent with Momentum, Gradient Descent with \textsuperscript{13-15} Each of the above methods are further trained using cross entropy, sum of absolute error, Mean of absolute error, sum of squared error, mean of squared error\textsuperscript{16}. The Rank of the features is obtained using ReliefF Algorithm\textsuperscript{17}

**Implementation methodology**

**Dataset**

For doing the data analysis using the Modified Masood Score, the samples were obtained from the Department of Pathology at JSS Hospital, JSS Medical College, Mysore. The study uses 321 such samples in which 122 samples belong to Benign or Non-Proliferative Breast Disease class, 64 samples belong to Intermediate class-1 or Proliferative Breast Disease without atypia class, 25 samples belong to Intermediate class -2 or Proliferative Breast Disease with Atypia class, 55 samples belong to Carcinoma class-1, 40 samples belong to Carcinoma class-2, 15 samples belong to Carcinoma class-3. Here both the intermediate classes indicate a condition where the breast lesion is moving towards cancerous condition and all the carcinoma classes indicate malignant condition or cancer which requires accurate treatment based on the severity of the disease indicated by class.

**Experimentation and Results**

The classification of the samples has been performed by using the feed-forward neural network (FNN). To test the accuracy of classification the samples have been considered in the ratio 70:30 as train : test. The feed-forward neural network is built using the input layer, the hidden layer and the output layer. In the input layer the six characteristic features with the scores are given as input. The weight and bias perform the activation function based on the sigmoid operation using neurons in the hidden layer. During experimentation it is observed that the neural network performed the best fit using 30 neurons for training in the hidden layer. The system performed under-fitting when the neurons used for training were lesser than 30 and it performed over-fitting when the neurons used for training were above 30. Hence the overall accuracy would decrease in both under-fitting and over-fitting. So 30 neurons were fixed to train the network using the feed-forward neural network. To enhance the accuracy of performance of the network, Pattern-Neural Network with the same features, train: test ratio and neurons was used to perform classification. Under this network the performance of the network was measured by considering five performance measures.

The results obtained when these various training algorithms and performance measures were used is as shown in Figure 1.

In the above Figure, Y Axis indicates the Classification Accuracy and the X-Axis indicates the training functions denoted as LM ,BR, BFG, RP, SCG, CGB, CGP, OSS, GDX, GDM, GD to indicate twelve training functions Levenberg-Marquardt, Bayesian Regularisation, BFGS Quasi newton, Reselient Back-Propagation, Scaled-conjugate Gradient, Conjugate Gradient with
Fig. 1. Classification Accuracy of various training and performance functions in FNN and PNN

**Table 1.** Six features used in the study and rank obtained to show the most and least significant feature

<table>
<thead>
<tr>
<th>Features 1 to 6 initially according to Rank</th>
<th>Chromatin Clumping</th>
<th>Myoepithelial cells</th>
<th>Anisonucleosis</th>
<th>Nucleoli</th>
<th>Chromatin Clumping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

Conjugate Gradient with bealle restart, Fletcher-Powell Conjugate Gradient, Polak-Ribiére Conjugate Gradient, One Step Secant, Variable Learning Rate Gradient Descent, Gradient Descent with Momentum and Gradient Descent respectively.

The five performance measures used to observe the performance of each training function are Cross-Entropy(CE), Sum of Absolute Error(SAE) ,Sum of Squared Error(SSE), Mean Of Absolute Error(MAE), Mean of Squared Error(MSE).

Here FNN Indicates Feed-Forward Neural Network and PNN Indicates Pattern Network.

Finally based on the above target functions and performance parameters used for activation, the neuron is fired to the target in the output layer based on which is the most probable class, the sample belongs to. The difference between the expected output target and the observed output target gives the error rate in performance of the network or in otherwords the accuracy of the system. The highest accuracy obtained is using Bayesian regularisation with accuracy of 87.53 % in the feed forward neural network. An accuracy of 88.44% is obtained using pattern network with Bayesian Regularisation as the training function and with cross-entropy as the performance parameter.

The rank importance of each feature is obtained using ReliefF Algorithm. It is as shown in the table below:

So from the above table we can understand that chromatin clumping is the most significant feature and Myoepithelial cells are the least significant feature to classify breast lesions.

**CONCLUSION**

In this paper, an automated method of classifying breast lesions into six classes based on Modified Masood Score is presented. This system overcomes the problem of classification in conventional classifiers when samples in a particular class are very less, It is efficient to classify samples even though the number of samples present in each class varies greatly with respect to another class. The system is chosen because it is simple and cost-effective to categorize
the breast lesions. The system is also sensitive to minor changes in the scores of characteristic features which help an oncologist to give the most effective dosage of treatment for early recovery. The rank of features has been approved by the pathologists involved in the study. It is an efficient step towards precise treatment for the cancer patient. A system which classifies the image of a breast lesions with better accuracy and sensitivity is being developed.

REFERENCES