

Copyright © 2010 Institute of Electrical and electronics Engineers, Inc.

All Rights reserved.

Personal use of this material, including one hard copy reproduction, is permitted.

Permission to reprint, republish and/or distribute this material in whole or in part for any other purposes must be obtained from the IEEE.

For information on obtaining permission, send an e-mail message to stds-igr@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

Individual documents posted on this site may carry slightly different copyright restrictions.

For specific document information, check the copyright notice at the beginning of each document.

A Novel Min-Max Feature Value Based Neural Architecture And Learning Algorithm For Classification of Microcalcifications

Brijesh Verma
School of Information Technology
Griffith University, Australia
b.verma@griffith.edu.au

Rinku Panchal
School of Information Technology
Griffith University, Australia
rinkoopanchal@yahoo.com

Kuldeep Kumar
School of Information Technology
Bond University, Australia
kuldeep_kumar@bond.edu.au

Abstract

The paper proposes a novel min-max feature value based neural architecture and learning algorithm for classification of microcalcification patterns in digital mammograms. The neural architecture has a single hidden layer and it has a fixed number of hidden units and outputs. One class is represented by three hidden units and an output. The suspicious areas represented by chain code, are extracted from digital mammograms. The feature values are extracted for benign and malignant microcalcifications. A set of min, average and max values for every input feature is defined and assigned to the weights between input and hidden layer. The weights of the output layer are calculated using least squares methods or assigned in such a way that it maximizes the output value for only one class. Many experiments were conducted on a benchmark database of digital mammograms and comparative results are included in this paper.

1 Introduction

Breast Cancer is a leading cause of non-preventable cancer deaths among women, with 1 in 11 women developing this disease during their lifetime¹. Early detection and diagnosis of breast cancer gives good chance of survival. While late detection and diagnosis often leads patient to unrecoverable stage of cancer ending in casualty. Digital Mammography currently offers the best control strategy for early detection of breast cancer. However, it is very difficult to distinguish benign and malignant microcalcification in digital mammogram. Sometimes microcalcification are very small and random in appearance. It makes very hard for doctor to detect and diagnose them. Recent revolution/development in image processing techniques, make microcalcification detection easy, however classification of malignant and benign microcalcifications is still very challenging and difficult

problem for radiologists. In last two decades, many papers have been published with novel techniques [1-14] to improve the detection and classification of benign and malignant microcalcifications. Researchers have proposed and investigated both traditional as well as intelligent techniques such as nearest neighbor, bayes rule, neural networks, etc. Some of the existing techniques are reviewed below.

Mascio et al [2] developed a microcalcification detection technique, which operates on original digitised mammograms by combining morphological image processing with arithmetic processing. Yoshida et al [3] used wavelet transform for their CAD application. They employed *Least Asymmetric Daubechies'* wavelets in conjunction with different image processing techniques to detect microcalcifications. The technique is very effective in separating microcalcification from normal background tissue and they achieved a detection rate of approximately 90%. Zheang et al [4] used a Multistage Neural Network. In proposed network architecture the first stage called as the "detailed network" where all pixel values of an original image is used as its input, while the second is known as a "feature network". Feature network used backpropagation learning algorithm and extracted features with output for detailed network set as an input. Qain et al [5] used Multichannel ($M = 4$) Wavelet Transform for segmentation and feature extraction purposes. They used M-channel ($M = 4$) Wavelet analysis in three ways: with a polyphase quadrature mirror filter (QMF) structure, with a tree structure, and with a lattice structure. Barman et al [6] used a low-pass filter to detect microcalcification by analysing digital mammogram. Woods et al [7, 8] used backpropagation algorithm and reported that a long training time was required for neural network training. It was difficult to determine the learning rate by which the weights were changed and updated. A learning rate that was too small needed more time to converge on a solution while a large learning rate jumped over the solution, this did not let to network learn properly resulting with poor classification rate. Chitre et al [9] used backpropagation neural network for image structure microcalcification classification and compared results with statistical classifiers. Though results were not promising but

¹ NSW Breast Cancer Institute, NSW, Australia
This research was funded by Griffith University Research Development (GURD) grant.

they were better than the statistical classifiers. Verma et al [11] investigated two neural network based techniques such as an error backpropagation (EBP) and a direct solution method (DSM). Zakos et al [12] proposed and investigated a neuro-fuzzy technique of detecting and classifying microcalcification patterns in Digital Mammogram. They have investigated 10 standard features and 4 modified features. Main idea behind the modified feature was to exploit characteristic of area more efficiently. They achieved promising results with the use of modified features. Normal features were – {skew, number of pixels, histogram, standard deviation}, Modified features were – {entropy, number of pixels, histogram, standard deviation} which produced good results.

As it can be seen from literature review above, that a lot of research has been conducted with promising results especially neural based intelligent techniques. Radiologists (at least from our local hospital) are very impressed with classification rates such as 88% on Nigmegean database, however they have reservations in using such a technique because it doesn't give any reason why it is classifying a particular area into benign or malignant. Other problems are inconsistency with classification rates, variable neural architecture, long training, etc. This paper proposes a novel architecture and a weight adjustment algorithm based on set of min, average and max values extracted from feature values which will be able to avoid a number of problems discussed above with current techniques. There are a number of advantages of our novel approach over the existing approaches. Some of them are as follows:

1. The architecture especially the #of hidden units is fixed therefore there will be no searching of a proper architecture using a very inefficient time consuming methods such as a trial and error based method.
2. The learning is based on a set of <min, max> values therefore it is possible to explain the reason for any output (benign, malignant, rubbish) provided by the network.
3. There is a simple rule based weight adjustment without any iterative time consuming learning process, therefore the learning process will be very fast and guaranteed.

The remainder of the paper is broken down into 4 sections. Section 2 describes the proposed approach, Section 3 provides experimental results and analysis, and a conclusion is drawn in Section 4.

2 Proposed Approach

2.1 An Overview of the Proposed Approach

The proposed approach first extracts the suspicious area from the image (digital mammogram) using the *chain*

code already annotated (provided by radiologists) in the database. Features are then extracted from this area and most significant features are fed to the neural network for classification into benign, malignant and rubbish. A novel *learning algorithm* is proposed in this paper to train the classifier. An overview of the proposed approach is presented below in Figure 1.

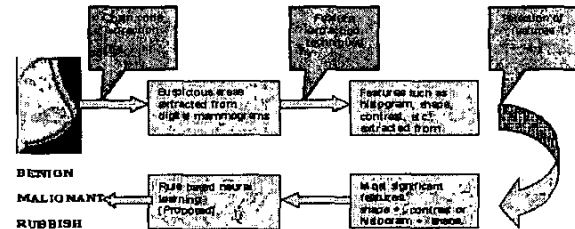


Figure 1. An Overview of Proposed Technique

The proposed approach contains the following stages: Mammographic database, Area extraction, Feature extraction, Selection of most significant features, Classification of features into benign/malignant/rubbish (Neural Network).

Stage 1: Mammographic database

The benchmark database (DDSM) of digital mammograms from the University of South Florida is currently being used. Previously, the database from the University of Nijmegen (The Netherlands), was used. All images (mammograms) were/are in raw format and were of size 2048x2048. They used 12 bits (2 bytes) per pixel of grey-level information.

Stages 2-3: Area extraction/Feature extraction

The suspicious areas are extracted as described in Section 3 from every mammogram using a chain code supplied with digital mammograms. The features such as average histogram, average grey level, energy, modified energy, entropy, modified entropy, number of pixels, standard deviation, modified standard deviation, skew, modified skew, average boundary grey level, difference and contrast were investigated. The formulae for entropy, energy, skew, and standard deviation were modified so that the iterations started with the first pixel of the pattern and ended at the final pixel. Traditionally, the formulae for these features have iterations starting with the lowest grey level and ranging to the highest grey level. This modification was done in an attempt to achieve a better classification rate than its traditional version. All 14 features are described below in detail. The following symbols are used in feature extraction formulae (1)-(12).

- T is the total number of pixels,
- g is an index value of image I,
- K is the total number of grey levels (i.e 4096),
- j is the grey level value (i.e. 0-4095),
- I(g) is the grey level value of pixel g in image I,

$N(j)$ is the number of pixels with grey level j in image I ,
 $P(I(g))$ is the probability of grey level value $I(g)$ occurring in image I ,
 $P(j)$ is the probability of grey level value j occurring in image I

Number of Pixels

Number of pixels is a count of the pixels in the microcalcification area. It gives an indication of the size of area.

Average Histogram

A histogram value gives an indication of the grey-level distribution for the mammogram. First generating the histogram for microcalcification and then calculating the average frequency across all grey-level intervals calculate average histogram.

$$\text{Average Histogram} = \frac{1}{k} \sum_{j=0}^{k-1} N(j) / T \quad (1)$$

Average Grey Level

Average grey level gives an indication of the grey level value associated with the entire microcalcification area.

$$\text{Average Grey} = \frac{1}{T} \sum_{g=0}^{T-1} I(g) \quad (2)$$

Energy and Modified Energy

Energy gives an indication of how the grey levels are distributed. If the energy value is high there is high grey level values. If energy value is low than there is low grey level values.

$$\text{Energy} = \sum_{j=0}^{k-1} [P(j)]^2 \quad (3)$$

$$\text{Modified Energy} = \sum_{g=0}^{T-1} [P(I(g))]^2 \quad (4)$$

Entropy and Modified Entropy

Entropy provides a measure of non-uniformity. As the pixel values in the image are distributed among more grey levels, the entropy increases.

$$\text{Entropy} = - \sum_{j=0}^{k-1} P(j) \log_2[P(j)] \quad (5)$$

$$\text{Modified Entropy} = - \sum_{g=0}^{T-1} P(g) \log_2[P(I(g))] \quad (6)$$

Standard Deviation and Modified Standard Deviation

Standard deviation is the spread of the data in the area.

$$\text{Standard Deviation} (\sigma) = \sqrt{\sum_{j=0}^{T-1} (j - \text{AvgGrey})^2 P(j)} \quad (7)$$

Modified Standard Deviation (σ) =

$$\sqrt{\sum_{g=0}^{T-1} (I(g) - \text{AvgGrey})^2 P(I(g))} \quad (8)$$

Skew and Modified Skew

Skew is a measure of the asymmetry about the mean grey level.

$$\text{Skew} = \frac{1}{\sigma^3} \sum_{j=0}^{k-1} (j - \text{AvgGrey})^3 P(j) \quad (9)$$

$$\text{Modified Skew} = \frac{1}{\sigma^3} \sum_{g=0}^{T-1} (I(g) - \text{AvgGrey})^3 P(I(g)) \quad (10)$$

Average Boundary Grey

The average boundary grey is the average grey level value around the outside of the microcalcification. Define some width of strip around microcalcification area, and calculate average grey level value of that strip.

Difference

Difference gives an indication of the change of intensity across the boundary of the microcalcification.

$$\text{Difference} = \text{Average Grey} - \text{Average Boundary Grey} \quad (11)$$

Contrast

Contrast gives an indication of how sharp the microcalcification structural features are.

$$\text{Contrast} = \frac{\text{Difference}}{\text{AvgGrey} + \text{AvgBoundaryGrey}} \quad (12)$$

Stage 4: Selection of most significant features

Initially, we determine the ranking of single features from best to worst by using each feature as a single input to the neural network. After this is completed, a combination of features is tested and a best feature or a combination of features is determined.

First feature vector [10 features]: average histogram, average grey level, number of pixels, average boundary grey, difference, contrast, energy, entropy, standard deviation and skew.

Second feature vector [14 features]: average histogram, average grey level, number of pixels, average boundary grey, difference, contrast, modified energy, modified entropy, modified standard deviation and modified skew.

The most significant feature or combination of features are selected based on neural network classification. It is done as follows. We start with a single feature by feeding it to the

neural network and analysing the classification rate. If it is increased or unchanged by adding a particular feature then we include this feature to the input vector. Otherwise we remove this feature and add another feature to the existing input vector and repeat the whole process again. In our other research project, an evolutionary approach is being investigated to find the most significant feature.

Stage 5: Classification of features into benign/malignant/rubbish

A novel learning architecture described in the next Section (2.2) is used for classification of features into benign/malignant/rubbish.

2.2 Novel Learning Algorithm

2.2.1 Neural Architecture

A novel neural architecture is proposed which is shown below in Figure 2. It has inputs, outputs and a fixed number of hidden units. There are 3 hidden neurons for each class. The number of input and output nodes must be equal to the number of features and number of classes respectively. Possible input feature vector values include one feature or combination of features described in the previous section.

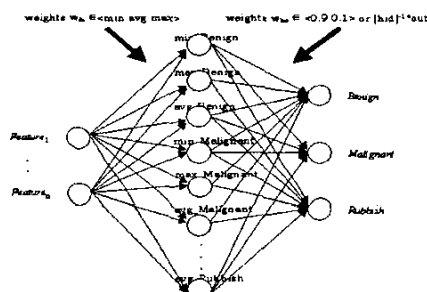


Figure 2. Proposed Neural Architecture

There are two novel things introduced in the above architecture, first that it has a fixed number of hidden units in the hidden layer. The total number of hidden units depends on the total number of classes. The *hidden unit* represents the part of (lower, middle or upper boundary) class. Depending on the input feature value (s), one of the hidden units will be fired by minimising/maximising the hidden neuron output, which will have influence in firing output layer neuron by maximising one of the outputs of the neural network. The second novel thing is that in the output layer, the *rubbish neuron* represents the areas, which don't contain benign and malignant microcalcifications. This is useful when radiologist select random areas on digital mammograms. The system should be able to provide suggestions even for random non-microcalcification areas.

2.2.2 Stepwise Learning Algorithm

Step 1: Calculate Min, Max and Average values of every input feature for all classes

Step 2: Initialise network architecture

#of inputs = #of features

#of outputs = #of classes + an extra class for rubbish

#hidden units = 3*#of classes (min, max, avg for ben/mal/rubbish)

#activation function = Gaussian type function

Step 3: Assign values to weights between input and hidden layer

$W_{ih}[i][j] = \{\text{min, max, average}\} \text{value}_{ik}$, Where k is the class {ben/mal/rub}

Step 4: Calculate the output of the hidden layer

Out {benign, malignant, rubbish} = $e^{-|(\text{wih} \cdot \text{feature}) \cdot \text{whol}|}$

Step 5: Calculate the weights between hidden and output layer

$W_{ho}[i][k] = \{0.9/0.1\}$,

Where, i is the hidden unit class {3 classes for each} and k is the output class

or $W_{ho}[i][k] = \text{HID}^{-1} * \text{OUT}$ {using least squares method}

3 Experimental Results And Analysis

3.1 Database (Digital Mammograms) Preparation

Digital mammogram database "DDSM" is taken from the university of South Florida. This is a benchmark database, available free of cost. It can be downloaded from the following web site

<http://marathon.csee.usf.edu/Mammography/>

DDSM has four cases for each patient. Each case contains four mammograms from a screening exam. Each case consist of three files as follows:

- **"ics" file**, this file provides information about a case as a whole. Including other relevant information, it provides the size of each image file, number of bits per pixel, and the scanning resolution (in microns).
- **"OVERLAY" file**, this file gives *no of abnormalities* a particular digital mammogram has and the most important, it gives *starting position (x, y coordinates) for suspicious area with chain code value* that is very important for area extraction.
- **"LJPEG" file**, this file contains a *compressed digital mammogram*. The images were scanned on a HOWTEK 960 digitizer with a sample rate of 43.5 microns at 12 bits per pixel. The images were preprocessed to crop out much of the image that did not contain breast tissues. Each image was compressed using truly loss less compression algorithm. Software is also available for decompression. Once uncompressed, each image file contains only raw pixel values (0-4095).

3.2 Area Resizing and Extraction

Once digital mammogram is uncompressed, then suspicious area is resized and extracted. Suspicious area is already marked in all digital mammograms of DDSM by three expert radiologists. Starting position of suspicious area and a chain code for area extraction is available from "OVERLAY" file. With the help of chain code we find the exact width and length of each suspicious area from each mammogram. The suspicious areas from different mammograms are different in size. To make all the suspicious area of same size, we need average size of each width and length. Using chain code values we define the boundary of area, by defining upper most side, lower most side, left outer side and right outer side of area. With the help of boundary, we define width and length of area. Using the length and width of all suspicious areas from all mammograms, we calculate the average width and average length. Finally we extract the 14 features as described in Section 2. Sample feature values for modified features are shown below in Table 1.

3.3 Feature Extraction Results

Table 1. Feature Values for 4 Modified Features

Malignant	Benign	mod No of Pixels	mod Histogram	mod Entropy	mod Stand Dev
0.9	0.1	0.09239	0.09239	0.107923	0.131301
0.9	0.1	0.16983	0.16983	0.121968	0.205651
0.9	0.1	0.094304	0.094304	0.121569	0.084525
0.9	0.1	0.06046	0.06046	0.117162	0.059979
0.9	0.1	0.057288	0.057288	0.123136	0.047323
0.9	0.1	0.376842	0.376842	0.388009	0.214422
0.9	0.1	0.099289	0.099289	0.102017	0.117806
0.9	0.1	0.180932	0.180932	0.215848	0.129321
0.1	0.9	0.031011	0.031011	0.090777	0.03305
0.1	0.9	0.032935	0.032935	0.050893	0.045669
0.1	0.9	0.039426	0.039426	0.062295	0.047891
0.1	0.9	0.067667	0.067667	0.089175	0.070222
0.1	0.9	0.000436	0.000436	0.001854	0.002644
0.1	0.9	0.01711	0.01711	0.039887	0.02785
0.1	0.9	0.082332	0.082332	0.081155	0.218348
0.1	0.9	0.008716	0.008716	0.011071	0.034072
0.1	0.9	0.016763	0.016763	0.015076	0.069995

Min-Avg-Max Values

The min, max and average values have been extracted for every feature. It is not easy to always find a clear boundary, so if there is a conflicting value, it is removed. For example, min, max and average values for the mod Entropy features can be described as follows:

Benign microcalcification:

<min, max> = <0.001854, 0.090777>

<avg>=0.0491

Malignant microcalcification

<min, max> = <0.102017, 0.388009>

<avg>=0.151

3.5 Feature Selection Results -Best 4 Features

The top four features from many initially tested experiments are shown in Table 2. The genetic selection of best features is still in progress. The results shown in Table 2 were obtained from experiments conducted on

neural network based selection without using genetic algorithm.

Table 2 Ranking of normal and modified features

Ranking	Features	
	Normal	Modified
1	Skew	Entropy
2	Number of pixels	Number of pixels
3	Histogram	Histogram
4	Standard Deviation	Standard Deviation

3.6 Classification Results

We have implemented our proposed novel neural learning approach and MLP based approach and we have tested them using just one feature, which was mod Entropy. The preliminary experiments are listed below in Tables 3 and 4.

Table 3. Classification rates for benign & malignant using proposed algorithm

#	# Hidden Units (fixed size)	# Iterations (no iteration)	Classification Rate [%]	
			Training Set	Test Set
1	6	0	100	81.80

Table 4. Classification rates for benign & malignant using EBP-MLP

#	# Hidden Units	# Iterations	Classification Rate [%]	
			Training Set	Test Set
1	8	20000	88.90	63.60
2	12	20000	100	81.80
3	12	50000	84.40	54.50

As it can be seen above in Tables 3 and 4, the classification rate on training and testing data sets for both techniques, is the same. However, first thing is to note that the proposed approach needs to be run only once because there is no problem in finding the number of iterations or hidden units, etc as in our traditional neural network based technique such as an error back propagation algorithm based MLP. The second most important thing is to note that our proposed approach can provide a reason for misclassification and an exact rule can be extracted. The third and final thing is to note that the classification rate on test set for both techniques is not very high. During the analysis of misclassification, we found that the reason of misclassification is incorrect (conflicting) feature value. Some miss-classified classes in test set have the following feature values: 0.06 - malignant, 0.1 - benign, etc. As you can see above that the value 0.06 is

labeled as malignant and 0.1 as benign in our test set, which is incorrect based on our min-max values. The results so far are in favor of the proposed approach.

4 Conclusions

We have proposed and investigated a novel min-max feature value based neural architecture and learning algorithm for classification of microcalcification patterns. The proposed approach was implemented in C/C++ on UNIX platform and many experiments were conducted. The preliminary experiments are very promising. As discussed in previous section and shown in Tables 1-4, our proposed approach has many advantages over traditional EBP-MLP based approach. The most important advantages are (1) it can provide the reason for misclassification or correct classification, which is very important in convincing radiologists (2) it is very quick, there are no unsuccessful training, trial & error based training to find hidden units, iterations, best classifications, etc. We are currently investigating and testing our approach with other features, combination of features, etc. and we will be able to include more results and a detailed analysis in our next paper.

References

1. Karessemeijer N., Computer-Assisted Reading of Mammograms, *European Radiology*, 1997, vol. 7, pp. 743-748.
2. Mascio L. N., Hernandez J. M. and Logan C. M., Automated Analysis for Microcalcifications in High Resolution Digital Mammograms, can be found at <http://canopus.llnl.gov/docum/documents/imaging/jmhspie93.html> ..
3. Yoshida H., Doi K. and Nishikawa R. M., Application of The Wavelet Transform To Automated Detection of Clustered Microcalcifications In Digital Mammogram., 1994, *Academic Reports of Tokyo Institute of Polytechnics*, vol.16, pp.24-37.
4. Zheng B., Qian W. and Clarke L. P., Multistage Neural Network for Pattern Recognition in Mammogram Screening, *IEEE International Conference on Neural Networks (ICNN)*, 1994, pp. 3437-3447.
5. Qian W. and Xuejun S., Digital mammography: Wavelet Transform and Kalman-Filtering Neural Network in Mass Segmentation And Detection, *IEEE Joint International Conference on Neural Networks (IJCNN)*, 2001, pp. 234-239.
6. Barman H., Granlund G. and Haglund L., Feature Extraction for Computer-Aided Analysis of Mammograms in state in *The Art of Digital Monographic Image Analysis*, World Scientific Publishing, 1994, vol. 7, no 6, pp. 128-147.
7. Woods K., Doss C. C. and Bowyer K. W., Comparative Evaluation of Pattern Recognition Technique for Detection of Microcalcifications in Mammography, *International Journal of Pattern Recognition and Artificial Intelligence*, 1993, vol. 7, no. 6, 1417-1436.
8. Woods K., Automated Image Analysis techniques for Digital Mammography, PhD Thesis, University of South Florida, 1994.
9. Chitre Y., Dhawan A. P. and Moskowitz M., Artificial Neural Network Based Classification of Mammographic Microcalcifications using Image Structure Features, *International Journal of Pattern Recognition and Artificial Intelligence*, 1993, vol. 7, no 6, pp.1377-1401
10. Shen L., and Rangayyan R. M., Detection and Classification of Mammographic Classifications, *International Journal of Pattern Recognition and Artificial Intelligence* 1993, vol. 7, no 6, pp.1403-1416.
11. Verma B.K., Comparative Evaluation of Two Neural Network based Techniques for Classification of Microcalcifications in Digital Mammograms, *Knowledge and Information Systems: An International Journal*, Springer-Verlag, 1999, vol. 1, no. 1, pp. 107-117.
12. Zakos J., A Computer-Aided Diagnosis System For Digital Mammograms using Computational Intelligence Techniques, 1998, pp.1-110, Honours Thesis, Griffith University, Australia.
13. Sutton M., Bezdek J., Neumann R., and Goar B., Enhancement and Analysis of Digital Mammograms Using Fuzzy Models, *SPIE Proceedings 3240 on Exploiting New Image Sources and Sensors*, 1997, pp. 345-352.
14. Solka J. and Poston W., The Detection of Micro-Calcification in Mammographic Images Using High Dimensional Features, *Seventh Annual IEEE Symposium on Computer-Based Medical Systems*, 1993, pp. 139-145.