MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

An Evaluation Study of Three Different Classification Methods on Mammogram Images

Levent E. Akman, Assoc. Prof. Dr. Nazife Baykal

Informatics Institute, Middle East Technical University, Turkey

Abstract

This study focuses on comparing three different types of classification methods –discriminant analysis, back propagation neural network and support vector machines (SVM) - on a special type of information -numerical data extracted from digitized mammography images-. Data extracted from mammography images show high variations among each other and, have an inhomogeneous nature. Therefore classification of this type of data becomes a challenging task. Neural networks have been known to be successful in analyzing data extracted from mammography images. Various examples of SVM analysis are also applied on mammogram classification in the literature. On the other hand, there are not many studies on discriminant analysis. Furthermore these three methods are not evaluated on the same data set. This study aims to compare and evaluate these methods by using the same data. The output of this work would provide valuable information about the comparative effectiveness of the three different classification methods on mammography image analysis for the researchers.

Keywords:

mammogram, microcalcification, classification,
neural networks, discriminant analysis,
support vector machines, evaluation.

Introduction

It is well known that screening mammography is the best tool available for detecting breast cancer before clinical symptoms appear because it can show changes in the breast up to two years before a patient or physician can feel them [1].

The main abnormality that the radiologists go through to detect breast cancer are the presence of microcalcifications [2]. The detected abnormalities are evaluated by the radiologists. This evaluation includes the decision task of whether the detected masses are malignant or benign by visually inspecting the detected microcalcifications [3].

Since mammograms have a high inhomogeneous nature, the data acquired from them have several drawbacks. First the data may be incomplete, inconsistent and very sensitive. These facts make the classification of this specific type of data a very complex problem.

As a first step of this study a detection application is implemented. The application uses Lee et. al.'s algorithm as the detection algorithm [4]. Using the detection application, selected features of the abnormalities are extracted for the classification.

In this study three different methods will be compared. A statistical method –discriminant analysis- and artificial intelligence methods –back propagation neural network and SVM- are selected for the analysis. The main idea behind this selection is to make a meaningful comparison and evaluation since the methods are based on different theories and mathematical backgrounds. Discriminant analysis is not a widely used method in the literature. On the other hand there are man examples on back propagation neural network classifier. SVM is also another analyzed method for the classification of microcalcification clusters by researchers. Our aim is to determine whether discriminant analysis produce worse results then the other two methods and furthermore determine the best classifier for microcalcification classification.

Finally a research study is accomplished for selecting the most discriminative features for the classification.

Materials and methods

Data

A total number of 117 mammograms were used in this study. Mammograms were acquired from two different sources: 68 mammograms from Cardiology service of Medical School of Hacettepe University and 49 mammograms from the mammographic image analysis society (MIAS) mini-mammographic database [5]. All of these mammograms are visually inspected by a faculty member at the Mammography service of Medical School of Hacettepe University. 117 of the mammograms contain microcalcification clusters where 60 of them are malignant and 57 of them are benign clusters.

The mammograms acquired from Hacettepe University were film mammograms. These mammograms are digitized with 300 dots per inch (DPI) using a Fuji's film scanner. The mammograms taken from the MIAS [5] mini-mammographic database were also digitized with 300 DPI.

All of the digitized mammograms are converted to an 8 bit image (i.e. the mammogram pixel intensities are mapped in to a 0-255 grey level range) before starting the applications since it is easier to work with an 8 bit image.

Detection of the abnormalities

The first step for a proper classification analysis on mammography images, is implementing an effective detection algorithm. In this study, two detection algorithms are developed and evaluated. Lee et. al.'s algorithm is selected as for the detection process since it produced better results. It should be mentioned that this detection algorithm is proven to give good results in many different environments.

The detection system is implemented by using Java-2 platform standard edition, software development kit version 1.4.2 (J2SE v 1.4.2 SDK) taken from Sun Microsystems Inc. [6]. The hardware used is a 1.0 GHz Pentium 3 processor with 256 Mb Read Access Memory (RAM).

The algorithm has 4 main steps: *Gradient Enhancement, Contrast Enhancement, Noise Elimination and Segmentation.*

The output of the segmentation sub-module is the final output of the entire detection system. The microcalcifications are detected satisfactorily by this system. Moreover the shapes of the microcalcifications are preserved which is a must for a microcalcification detection system since the classification is based on the shape and intensity features of the detected calcifications. The result of the detection system is depicted in Figure 1.

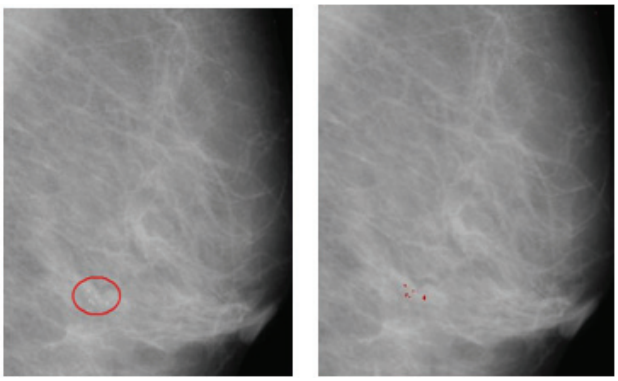

Figure 1 - Original mammogram (left) and segmented mammogram (right) with the detection system

The microcalcifications are marked in a circle in the original view mammogram. The system paints the detected calcifications as depicted in Figure.

The detection system is operated for 153 different mammograms and the detected abnormalities are cropped for the classification analysis. It should be mentioned that the detection results are evaluated by a professional and experienced radiologist and only the successful results (117 samples) are taken for classification.

Classification of the detected abnormalities

After detecting the abnormalities, the radiologists evaluate these results. This evaluation includes the decision task of whether the detected masses are malignant or benign and moreover whether to biopsy the suspicious findings [3].

A false positive diagnosis (i.e. diagnosing a benign tumor as malign) may cause an unnecessary biopsy while a false negative diagnosis (i.e. diagnosing a malign tumor as benign) may lead to cost a human life. .It has been reported that only 15 to 34% of calcifications requiring biopsy are cancerous which indicates a very high False Positive (FP) rate [7].Therefore it is important to correctly identify problematic fields as malignant or benign.

Feature extraction

Feature extraction is a preliminary step for starting classification. It is very important since the selected features will dominate the classification results. In order to define the correct features, the properties of the calcifications that point to breast cancer should be defined. Malignant calcifications are typically very numerous, clustered, small, dot-like or elongated, variable in size, shape and density. Benign calcifications are generally larger, more rounded, smaller in number, more diffusely distributed, and more homogeneous in size and shape [3].

The calcifications are analyzed in two different categories: individual calcifications and clustered calcifications [8]. In a computer aided diagnosis (CAD) system, these two are implemented separately. In other words, it is meaningless to examine only individual calcifications or examine only the groups. Size, shape and radiographic density of the calcifications are the main issues that define the level of malignancy when analyzing individual microcalcifications while their number, orientation, size, position are taken into account when examining microcalcification clusters.

Considering the above definitions about the malignancy indicators the features that point to the breast cancer can be generalized as shown in Table 1 [9].

Table 1

Features relating to descriptions of individual calcification:	**Features relating to the description and the distribution of clustered calcifications:**
Area of calcification	Average distance between calcifications
Average grey level	Orientation of calcifications with respect to its neighbors
Density	Variation of parameters of individual calcifications

Perimeter	Number of calcifications in the cluster
Contrast	Shape of the cluster
Smoothness	Size of the cluster
Approximate horizontal and vertical lengths	Compactness of the cluster
Shape descriptors- compactness, higher moments, Fourier descriptors	Cluster location in the breast
Effective thickness	

Furthermore, additional information can be included in the diagnostic procedure; such as patient history, clinical or demographic data. This study does not include such demographic data but concentrates on microcalcification features only.

Considering the indicators of malignancy described in Table 1,eight different attributes (representing the indicators) are extracted from the regions of interests as follows:

- **Attribute1** is the total number of detected abnormalities in the processed mammogram.
- **Attribute2** is the total area of the detected abnormalities in terms of pixels divided by the total area of the current mammogram.
- **Attribute3** is the mean intensity value of the detected abnormalities divided by the mean intensity value of the mammogram.
- **Attribute4** is the standard deviation of intensity values of the detected abnormalities.
- **Attribute5** is a measure of irregularity of detected abnormalities in shape.
- **Attribute6** is the standard deviation of Attribute 5.
- **Attribute7** is the mean distance between the abnormalities
- **Attribute8** is the standard deviation of the distances between the abnormalities.

Classification

The classification is implemented by extracting the selected features from the detection results. This data set then analyzed in Matlab 6.5 software package for three different classification techniques.

Discriminant analysis

Discriminant analysis is a technique for classifying a set of observations into predefined groups [10]. The classification model is built based on observations, using a set of variables (known as predictors or input variables), for which the groups are known. Discriminant analysis may be used for either measuring the success of classification, given the groups of objects, or assigning objects to one of the groups of objects.

60 of the data will be used for constructing the discriminant formula and the training of the back propagation neural network. And 57 of the data will be used for testing. We have utilized discriminant analysis to classify the mammographic image data.

In this analysis *linear* discriminant analysis and *quadratic* discriminant analysis are used.

For discriminant analysis, we have implemented the test in two different schemes. First we have developed the discriminant function with the training data and then tested it using the test data. “0” value is assigned for benign calcifications and “1” is assigned for malignant calcifications.

When the test data gives a value higher then “0,85” it is accepted in the malignant calcification class. Test values lower then “0,15” is included in the benign calcifications class. This scheme applies for all the three methods implemented in this study.

Before starting the analysis we should define the statistical measures for evaluating the classification methods. “Overall Success Ratio” defines the ratio of the “Total Number of Correct Classifications” to the “Total Number of Observations”. The other descriptive statistics are “Sensitivity” and “Specificity” measures.

Here it should also be mentioned that the same training and testing data which is utilized in discriminant analysis is also used for other classification techniques. The results of the linear discriminant analysis are given in Table 2:

Table 2

Overall Success Rate	Sensitivity	Specificity
84%	0,88	0,80

For quadratic discriminant analysis the results are as follows:

Table 3

Overall Success Rate	Sensitivity	Specificity
82%	0,88	0,76

Linear discriminant analysis gave slightly better results than quadratic analysis for the selected features. However it was expected to gain more benefit from quadratic discriminant analysis since the problem domain is ambiguous and fuzzy in nature.

Neural networks

Neural networks try to solve problems with a structure roughly similar to the neurons in the human brain. McCulloch and Pitts [11] introduced the first neural model. Back Propagation Neural Networks (BPNN) are one of the most common neural network structures, as they are simple and effective, and have a wide variety of application areas.

Neural networks are suitable for applications where only a few decisions are required from a massive amount of data, and also for the applications where a complex nonlinear relation needs to be learned. This property fits with the mammographic data where there is a huge amount of non-linear and fuzzy data which has to be classified into two classes namely benign and malignant.

We started the analysis by utilizing back-propagation neural network. Various back-propagation nets are implemented to get the best results. However due to space constraints we will give the results of the best network structure. The best net structure is determined as a three layered – one input, one hidden and one output- network. Since the data has 8 dimensions there are 8 neurons in the input layer. For this first design 3 neurons for the hidden layer gave the best results. And there is one output neuron. The best results are achieved with 500 epochs of training. The activation function is the tangent function and a standard back-propagation learning algorithm (gradient descent) is utilized. The net is depicted in Figure 2.

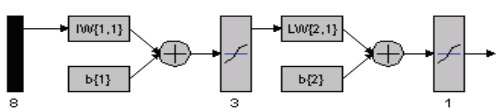

Figure 2 - Back propagation net constructed for mammogram data

The same data set selection with the discriminant analysis is used in the back propagation neural network technique as mentioned earlier. According to this schema the network is trained using 60 pre-selected data. And then the structured scheme is tested with the remaining 57 (27 malign,30benign) data. Below in Figure 3 the results are depicted for 500 epochs which yields the best results among many different trials.

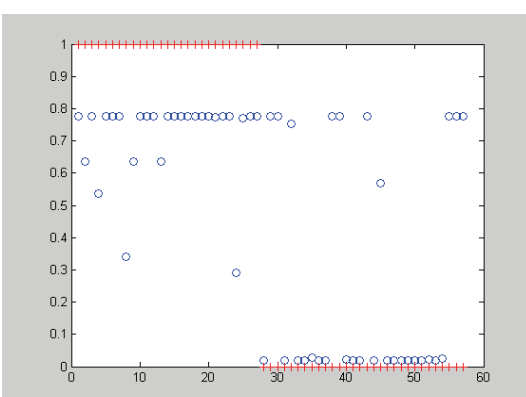

Figure 3 - Test results

The net could not classify 5 samples. The results are shown in Table 4.

Table 4

Overall Success Rate	Sensitivity	Specificity
72%	0,91	0,69

The results of the back propagation neural network can not said to be satisfactory from a clinical point of view. However our main aim is to evaluate this result with the other classification techniques. We can conclude that BPP network is less satisfactory then discriminant analysis.

The less number of epochs resulted in a lower success rate. On the other, when we tried higher number of epochs (i.e. 1000 epochs) the network tends to summarize the data and therefore gives ambiguous results.

Support Vector Machines (SVM)

The attention on this subject is increasing in the last few years. SVMs provide a new approach to the problem of data classification. This approach has relationships with statistical learning theory. SVM is characterized according to its Kernel functin. There are four basic and frequently used kernels. These are:

- linear
- polynomial
- radial basis function (RBF)
- sigmoid

SVM has major differences with comparable approaches such as neural networks: SVM training always finds a global minimum, and their simple geometric interpretation makes further investigation easy. An SVM is largely characterized by the choice of its kernel.

After implementing back-propagation neural networks and discriminant analysis we have analyzed this specific classification problem with SVM. For this analysis, linear SVM and non-linear SVM methods are utilized. Linear SVM uses linear kernel and we utilized polynomial kernel for non-linear SVM in our study. The results of these two methods are discussed below respectively.

Linear SVM results are depicted in Figure 4 and Table 5. In Figure 4, the red pluses represent the original output values, blue circles represent the output values of the test data after toe SVM application and green stars represent the decision values utilized to get the test data output values.

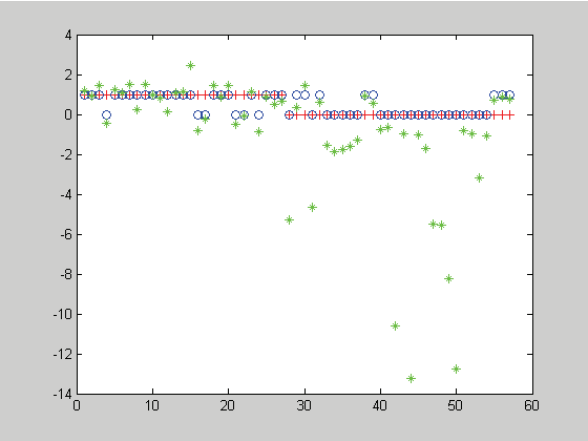

Figure 4 - The linear SVM results

Table 5

Overall Success Rate	Sensitivity	Specificity
75%	0,77	0,73

Linear SVM result in a higher success rate then neural net however the sensitivity and specificity results are not satisfactory. The linear SVM considered being unsuccessful from a clinical point of view.

The non-linear SVM method is implemented finally. We were expecting the most satisfactory results from this method because SVM has a high performance on non-linear and ambiguous data. The results are summarized in Figure 5 and Table 6.

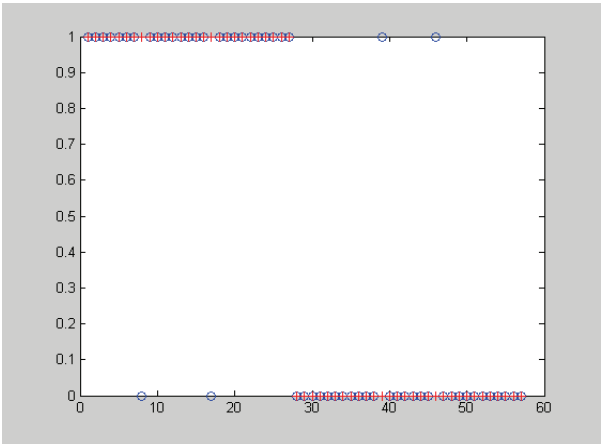

Figure 5 - The results of non-linear SVM

Table 6

Overall Success Rate	Sensitivity	Specificity
93%	0,93	0,93

It is obvious that non-linear SVM is very successful for the classification of mammographic data both from the clinical point of view and respectively to other methods. Only 4 entries are misclassified which is a high satisfactory result.

Discussions and conclusion

Three different classification methods, namely discriminant analysis, back propagation neural networks and support vector machines, are evaluated on a data set that is extracted from mammogram films. Discriminant analysis is not a widely used technique for classifying microcalcifications on mammograms. Using this fact, we comparatively assessed the performance of discriminant analysis method with respect to BPN and SVM techniques on the same training and testing data sets. This study also measures the results of the classifications from a clinical success perspective. In other words, we evaluated the overall success of a method for real life scenarios. Moreover two different detection algorithms are evaluated and the superior one is selected for the detection of microcalcifications. Finally a research study is made to determine the best features for microcalcification classification.

Discriminant analysis produced an overall success rate of 84% which is higher than the BPN success result of 72%. BPN result is also unsatisfactory from a clinical point of view. The most effective method is non-linear SVM for our analysis. Non-linear SVM produced 93% success rate. The sensitivity and specificity measures of SVM are also very satisfactory. SVM is also satisfactory from a clinical usage perspective.

As a result discriminant analysis should be considered during microcalcification classification on mammograms since it produced competent results. The high success of SVM on ambiguous data is again proved after this study.

Acknowledgments

We would like to thank to Prof. Dr. Figen Baaran Demirkazk for her support for her valuable support on providing and evaluating the mammogram images.

References

[1] Cheng, H. D. & Cui, M. (2004). Mass lesion detection with a fuzzy neural network. Pattern Recognition, 37(6), 1189-1200.

[2] Lanyi, M. (1986). Diagnosis and differential diagnosis of breast calcifications. Berlin: Springer-Verlag.

[3] Cheng, H. D., Cai, X., Chen, X., Hu, L. & Lou, X. (2003). Computer-aided detection and classification of microcalcifications in mammograms: a survey. Pattern Recognition, 36(12), 2967–2991.

[4] Lee, S. K., Lo, C. S., Wang, C., Chung, P., Chang, C., Yang, C., Hsu, P. (2000). A computer-aided design mammography screening system for detection and classification of

microcalcifications. International Journal of Medical Informatics, 60(1), 29-57

[5] MIAS. (2004). The Mammographic Image Analysis Sc.,URL http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html. Accessed in: 2006.

[6] Sun Microsystems Inc. (2004). Sun Microsystems Inc. URL: http://www.sun.com/. Accessed in: 2004.

[7] Knutzen, A. M. & Gisvold, J. J. (1993). Likelihood of malignant disease for various categories of mammographically detected, nonpalpable breast lesions. Mayo Clinic Proceedings, 68, 454-460.

[8] Qian, W., Mao, F., Sun, X., Zhang, Y., Song, D. & Clarke, R. A. (2002). An improved method of region grouping for microcalcification detection in digital mammograms. Computerized Medical Imaging and Graphics, 26(6), 361-368.

[9] Markopoulos, C., Kouskos, E., Koufopoulos, K., Kyriakou, V. & Gogas, J. (2001). Use of artificial neural networks (computer analysis) in the diagnosis of microcalcifications on mammography. European Journal of Radiology, 39, 60-65.

[10] McLachlan, G. J. (1992). Discriminant analysis and statistical pattern recognition. New York: Wiley.

[11] McCulloch, W., and Pitts, W. (1943). A logical calculus of the ideas imminent in nervous activity. Bulletin of Mathematical Biophysics 5: 115-33.

[12] Dhawan A.P., Chitre Y., Moskowitz M. (1993). Artificial neural network based classification of mammographic microcalcifications using image structure features, Proc. SPIE 1905 820–831.

[13] Kramer D., Aghdasi F. (1998). Classification of microcalcifications in digitised mammograms using multiscale statistical texture analysis, Proceedings of the South African Symposium on Communications and Signal Processing, September 7–8, pp. 121–126.

[14] Kramer D., Aghdasi F., (1999). Texture analysis techniques for the classification of microcalcifications in digitized mammograms, Proceedings of the 1999 Fifth IEEE AFRICON Conference Electrotechnical Service for Africa, September 28–October 1, pp. 395–400.

[15] Jensen F.V. (1996). An Introduction to Bayesian Network, Springer, New York, NY, 1996

[16] Woods K.S., Doss C.C., Bowyer K.W., Solka J.L., Priebe C.E., Kegelmeyer W.P. (1993). Comparative evaluation of pattern recognition techniques for detection of microcalcifications in mammography, Int. J. Pattern Recognition Artif. Intell. ,7 1417–1436.