

Ensemble Supervised Classification Method Using the Regions of Interest and Grey Level Co-Occurrence Matrices Features for Mammograms Data

Hossein Yousefi Banaem,¹ Alireza Mehri Dehnavi,^{1,2,*} and Makhtum Shahnazi³

¹Department of Biomedical Engineering, Faculty of Advanced Medical Technology, Isfahan University of Medical Sciences, Isfahan, Iran

²School of Optometry and Visual Science, University of Waterloo, Waterloo, Canada

³Department of Radiology, Faculty of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

*Corresponding author: Alireza Mehri Dehnavi, Department of Biomedical Engineering, Faculty of Advanced Medical Technology, Isfahan University of Medical Sciences, Isfahan, Iran. Tel: +98-31 95016497, E-mail: mehri@med.mui.ac.ir

Received: April 20, 2013; Revised: July 21, 2013; Accepted: December 15, 2013

Background: Breast cancer is one of the most encountered cancers in women. Detection and classification of the cancer into malignant or benign is one of the challenging fields of the pathology.

Objectives: Our aim was to classify the mammogram data into normal and abnormal by ensemble classification method.

Patients and Methods: In this method, we first extract texture features from cancerous and normal breasts, using the Gray-Level Co-occurrence Matrices (GLCM) method. To obtain better results, we select a region of breast with high probability of cancer occurrence before feature extraction. After features extraction, we use the maximum difference method to select the features that have predominant difference between normal and abnormal data sets. Six selected features served as the classifying tool for classification purpose by the proposed ensemble supervised algorithm. For classification, the data were first classified by three supervised classifiers, and then by simple voting policy, we finalized the classification process.

Results: After classification with the ensemble supervised algorithm, the performance of the proposed method was evaluated by perfect test method, which gave the sensitivity and specificity of 96.66% and 97.50%, respectively.

Conclusions: In this study, we proposed a new computer aided diagnostic tool for the detection and classification of breast cancer. The obtained results showed that the proposed method is more reliable in diagnostic to assist the radiologists in the detection of abnormal data and to improve the diagnostic accuracy.

Keywords: Classification; Mammogram; Breast Cancer

1. Background

Breast cancer is one of the most frequent cancers among women throughout the world. One in every 1000 women has been suffering from breast cancer during 1974 - 1978. However, nowadays it occurs in one in every 10 women. This means that effective preventive actions must be taken to reduce the rate of this dangerous cancer (1). The commonly used diagnostic techniques for breast cancer screening is mammography, thermography and ultrasound imaging. Among these techniques, mammography is the gold standard approach for early detection. In early stage, the detection of microcalcifications appears in the breast tissue. Microcalcifications are small calcium deposits and appear as groups of radioopaque spots in most cancerous mammograms. Detection and classification of mammogram abnormalities is the challenging field of breast cancer diagnosis. There are different techniques for breast cancer detection, such as neural network, fuzzy logic and wavelet based

algorithms (2, 3). Mammography is the best screening tool for the detection of breast cancer in early stages, before appearance in physical examination. There are several features in mammography that help physicians to detect abnormalities in early stage, and these features can be directly extracted by image processing methods (4). The cancerous breast symptoms comprise of mass, changes in shape, color and dimension of breast. If the cancer is detected in earlier stage, a better treatment can be provided. Recently, computer aided diagnosis (CAD) systems have been developed to detect breast cancer automatically. Normal tissues typically have smooth boundary and surface, whereas abnormal tissue presents rough surfaces and jagged boundaries (5). The goal of diagnosis is distinguishing between normal and abnormal images. For this purpose, there are several methods available that we can use for features extraction from the digital image, such as: region-based

features, shape-based features, texture based features and position based features. In digital mammography, the most common used feature for classifying normal and abnormal pattern is texture feature. In this paper, we used the texture based Gray-Level Co-occurrence Matrices (GLCM) features for this purpose. For breast cancer classification, we should select several features with special criteria.

For breast cancer detection and classification, there are numerous research methods and algorithms. Yachoub et al. (6) used a hypothesis test to determine if the feature can discriminate or not. Verma et al. (7) developed a diagnosis algorithm based on a neural-genetic algorithm feature selection method for digital mammograms and the obtained accuracy was 85%. Alolfe et al. (8) used the filter model and wrapper model for feature selection. Chen et al. (9) proposed rough set-based feature selection. Vasantha et al. (10) proposed the hybrid feature selection method for mammogram classification. The highest classification accuracy obtained by this approach was 96%. Huang et al. (11) used a support vector machine based feature selection and obtained accuracy was of 86%. Prathibha et al. (12) used the Sequential Floating Forward Selection (SFFS) to reduce the feature dimensionality. Luo et al. (13) used two well-known feature selection techniques, including forward selection and backward selection, and two classifiers for ensemble classification. They have used a decision tree and support vector machine, as an initial classifier. Wei et al. (14) used a sequential backward selection method for the purpose of selecting the most relevant features. In their work, 18 features were extracted, out of which 12 features were finally selected for the classification of benign and malignant pattern.

2. Objectives

To overcome the problem of overfitting and underfitting encountered in other studies, we present the ensemble supervised classification method with simple voting policy for the detection of normal and abnormal pattern in the mammogram data, with reasonable accuracy.

3. Patients and Methods

The proposed method consists of three main steps, as follows: 1- feature extraction, 2- feature selection and 3- the classification process using ensemble supervised classification technique. In brief our methodology has presented in Figure 1.

3.1. Data

We obtained the required data from the Digital Database for Screening Mammography (DDSM). The resolution of obtained images was 42 microns with 4964×2900

pixels and breast density rating was up to 3 in the Breast Imaging-Reporting and Data System (BI-RADS). study instead of paper, we have used 300 mammograms for classification. More than 2500 DDSM data sets were available at <http://marathon.csee.usf.edu/Mammography/DDSM> (15). The original image that was downloaded from DDSM has been shown in Figure 2.

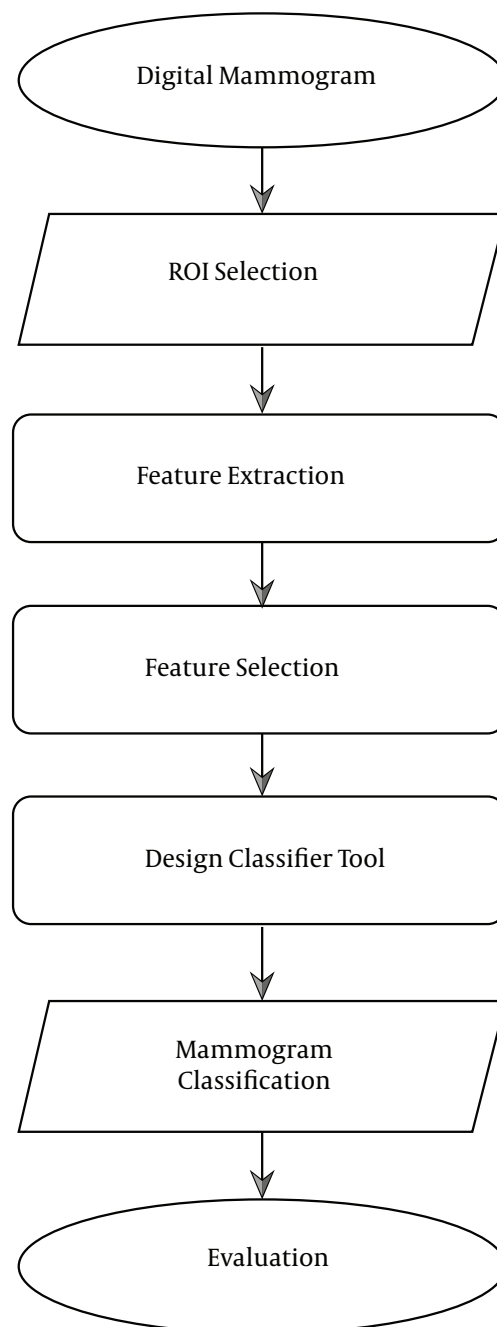


Figure 1. Block diagram of the proposed method.

3.2. Feature Extraction

Feature extraction is a crucial step in the mammograms classification. If the extracted features are not proper, overfitting, underfitting and misclassification occurs. For obtaining relevant features, after reading the image, we restricted ourselves to the region of mammograms with highest probability of cancer occurrence. We selected a region of interest (ROI) rectangular window with 512×512 pixels in size and then we extracted the features from this region. The obtained image has been shown in Figure 3.

In this approach, the obtained parameters are more reliable for classification procedures. Here we have used GLCM for feature extraction method. The GLCM features are calculated in four directions, which are 0° , 45° , 90° , 145° , and four distances (1, 2, 3, 4). The 20 expressions of GLCM descriptors are listed in Table 1. Extracted features from data have been shown in Table 2. These features are more redundant for classification and several of them are unnecessary; therefore, we applied the feature selection method.

Where x and y are the coordinates of an entry in the co-occurrence matrix, μ_x , μ_y , σ_x , and σ_y are the mean and standard deviation, and the partial probability function, $P_{x+y(i)}$ is the probability of co-occurrence matrix coordinating summing to $x + y$. The HX and HY are the entropies of p_x and p_y :

$$HXY = -\sum_i \sum_j p(i, j) \log(p(i, j))$$

$$HXY1 = -\sum_i \sum_j p(i, j) \log(p_x(i) p_y(j))$$

and

$$HXY2 = -\sum_i \sum_j p_x(i) p_y(j) \log(p_x(i) p_y(j))$$

3.3. Feature Selection

Feature selection is an important step for feature dimension reduction in the classification procedure. After feature extraction, feature selection method was applied to select the best features. The maximum difference method was used as a feature selection method in this paper. This method selects the features that have maximum difference between two groups of data. Therefore, the selected features show more differences between normal and abnormal data. At the end, six dominant features of 20 features were selected, as shown in Table 3. In the next step, the selected features were used for classification.

3.4. Classification

The classification process includes two steps: 1) initial classification and 2) ensemble. In the initial step, the K-nearest neighbors (KNN), naive Bayes and support vector machine (SVM) algorithms have been used as supervised classifiers for classification of normal and abnormal data.

The KNN algorithm is a method for classifying objects on closed training data in the feature space. In KNN algorithm, classification of an object enrolled by a majority vote of its neighbors is performed. In this paper, we used the KNN algorithm with $K = 5$ (16).

Naive Bayes classifier can handle an arbitrary number of independent variables, whether continuous or categorical. Given a set of variables, $= \{x_1, x_2, x_3, \dots, x_d\}$, we want to construct the posterior probability for the given C_j among a set of possible outcomes $C = \{c_1, c_2, c_3, \dots, c_d\}$. Using Bayes' rule (17):

$$(1) \quad p(C_j | x_1, x_2, x_3, \dots, x_d) \propto p(x_1, x_2, x_3, \dots, x_d | C_j) p(C_j)$$

Where $P(C_j | x_1, x_2, x_3, \dots, x_d)$ is the posterior probability of class membership. Since Naive Bayes assumes that the conditional probabilities of the independent variables are statistically independent, we can decompose the likelihood to a product of terms:

$$(2) \quad p(X | C_j) \propto \prod_{k=1}^d p(x_k | C_j)$$

And rewrite the posterior as:

Equation 3.

$$(3) \quad p(C_j | X) \propto p(C_j) \prod_{k=1}^d p(x_k | C_j)$$

Using the Bayes rule above, we label a new case X with a class level C_j , that achieves the highest posterior probability.

The SVMs construct a decision surface in the feature space that bisects the two categories and maximizes the margin of separation between two classes of points. This decision surface can then be used as a basis for classifying points of unknown class (18).

Suppose we have N training data points $f(x_1; y_1); (x_2; y_2); (x_N; y_N)$ g, where $x_i \in R^d$ and $y_i \in \{-1; +1\}$. The problem of finding a maximal margin separating hyperplane plane can be written as:

$$(4) \quad \min_{w, b} \frac{1}{2} w^T w \text{ subjected to } y_i (w^T x_i - b) \geq 1, i = 1, \dots, N$$

This is a convex quadratic programming problem. Introducing Lagrange multipliers and solving to get the Wolfe dual, we obtain:

$$(5) \quad \max_{\alpha} \text{imize } \xi_D \equiv \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i x_j$$

Subject to:

$$(6) \quad \sum_i \alpha_i y_i = 0, \alpha \geq 0$$

The solution of the primal problem is given by:

$$(7) \quad w = \sum_{i=1}^N \alpha_i y_i x_i$$

To train the SVM, we search through the feasible region of the dual problem and maximize the objective function.

To classify the mammograms, the first 200 data features have been used for classifier training, and the remaining data was used for classifier evaluation. To obtain acceptable accuracy, we have used ensemble classifier approach. In this approach, we initially classify the data by three classification algorithms. In the next step, we applied the simple voting policy for finalizing the classification.

This policy is done in two steps. At the first step, we assign a label to data, as temporary label. After assigning temporary labels to all data, the second step commences. In this step, the final label for each pixel will be the one that obtains the maximum number of votes between the temporary labels of its surrounding neighbors. Equation 8 presents this process for each data set.

$$(8) \quad \text{Final Class}(i) = \operatorname{argmax}_{\lambda_m \in \Gamma} \left(\sum_{s=1}^{i+1} \beta(s) \right)$$

Where the Final_class (i) is the final label allocated to data (i), $L = (\lambda_1, \lambda_2)$ is a set of associated labels to two normal and abnormal, s is variable for defining neighborhood around data (i) and is $\beta(s)$ defined by Equation 9.

$$(9) \quad \beta(s) = \begin{cases} 1 \rightarrow \text{if } \operatorname{argmax}_{\lambda_q \in \Gamma} \left(\sum_{k=1}^3 \alpha_k(s) \right) = \lambda_m \\ 0 \rightarrow \text{otherwise} \end{cases}$$

In this Equation, $\operatorname{argmax}_{\lambda_q \in \Gamma} \left(\sum_{k=1}^3 \alpha_k(s) \right)$ determines temporary label for data (i) and $\alpha_k(s)$ is defined through Equation 10.

$$(10) \quad \alpha_k(s) = \begin{cases} 1 \rightarrow \text{lable } (s, k) = \lambda_q \\ 0 \rightarrow \text{otherwise} \end{cases}$$

In Equation 10, the label is a matrix with the size of M Multiply 3 (M×3), where M is the length of data. It contains all labels that different classifiers assign to the data. For example, label can be defined for data (M) as follow:

- Label (M, 1) = KNN_Label (M)
- Label (M, 2) = Bayes_Label (M)
- Label (M, 3) = SVM_Label (M).

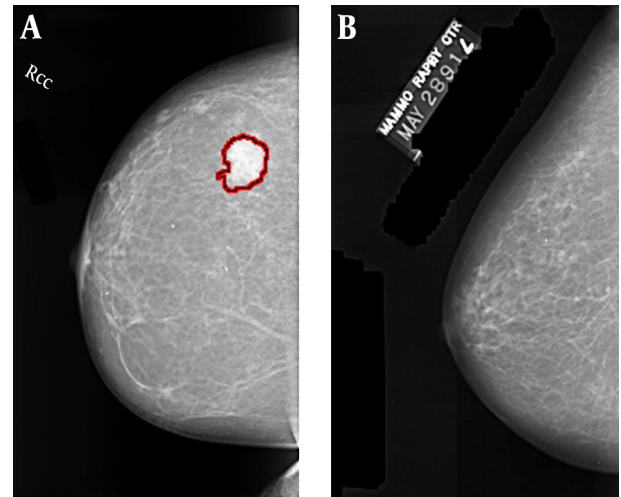


Figure 2. Original image that was obtained from the digital database for screening mammography database: A, Cancerous image; B, Normal image.

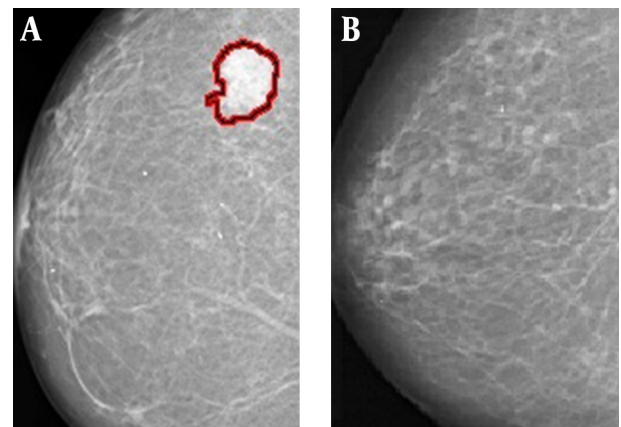


Figure 3. Region of interest selected image: A, cancerous image; B, normal image.

Table 1. Expression of Gray-Level Co-Occurrence Matrices Descriptors

| Features Computed | Formulation |
|---|--|
| Autocorrelation | $= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (p_x - \mu_x) (p_y - \mu_y) / \sigma_x \sigma_y$ |
| Contrast | $= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) (i - j)^2$ |
| Correlationw | $= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) (ij) - (\mu_x \mu_y) / \sigma_x \sigma_y$ |
| Cluster prominence | $= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) (i + j - \mu_x - \mu_y)^4$ |
| Cluster shade | $= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) (i + j - \mu_x - \mu_y)^3$ |
| Dissimilarity | $\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} i - j P(i, j)$ |
| Energy | $\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j)^2$ |
| Entropy | $-\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \log(P(i, j))$ |
| Homogeneity | $-\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{P(i, j)}{1 + i - j }$ |
| Maximum probability | $\max(i, j) P i, j $ |
| Sum of squares variance | $= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) (i - \mu)^2$ |
| Sum average | $= \sum_{i=0}^{2G-2} i P_{x+y}$ |
| Sum variance | $= \sum_{i=0}^{2G-2} (i - \sum_{i=0}^{2G-2} i P_{x+y})^2 P_{x+y}$ |
| Sum entropy | $-\sum_{i=0}^{2G-2} P_{x+y}(i) \log(P_{x+y}(i))$ |
| Difference variance | $\sum_{i=0}^{G-1} i^2 P_{x-y}(i)$ |
| Difference entropy | $-\sum_{i=0}^{G-1} P_{x+y}(i) \log(P_{x+y}(i))$ |
| Information measure of correlation 1 | $\frac{HXY - HXY1}{\max(HX, HY)}$ |
| Information measure of correlation 2 | $\sqrt{(1 - \exp[-2.0(HXY2 - HXY)])}$ |
| Inverse difference normalized (INN) | $\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{p(i, j)}{1 + i - j }$ |
| Inverse difference moment normalized | $\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{p(i, j)}{1 + i - j ^2}$ |

Table 2. List of Gray-Level Co-occurrence Matrices Features Sets that Were Obtained From Selected Regions of Interest

| Features Computed | Normal | | Abnormal | |
|--------------------------------------|----------|----------|----------|----------|
| | Sample 1 | Sample 2 | Sample 3 | Sample 4 |
| Autocorrelation | 1.20143 | 1.20115 | 7.28408 | 7.28177 |
| Contrast | 1.72836 | 1.67006 | 1.83504 | 1.73961 |
| Correlation | 9.70845 | 9.71811 | 9.42761 | 9.45678 |
| Cluster prominence | 3.52433 | 3.52576 | 1.01281 | 1.01238 |
| Cluster shade | 2.69382 | 2.69782 | 8.70367 | 8.72303 |
| Dissimilarity | 1.71778 | 1.66048 | 1.80913 | 1.72401 |
| Energy | 1.36384 | 1.37367 | 1.81605 | 1.83811 |
| Entropy | 2.31528 | 2.30567 | 2.05505 | 2.03751 |
| Homogeneity | 9.14278 | 9.17125 | 9.09961 | 9.14058 |
| Maximum probability | 2.20597 | 2.20477 | 3.08118 | 3.09041 |
| Sum of squares variance | 1.20107 | 1.19866 | 7.30073 | 7.28414 |
| Sum average | 6.04537 | 6.04403 | 4.80535 | 4.80318 |
| Sum variance | 2.65242 | 2.65528 | 1.45384 | 1.45814 |
| Sum entropy | 2.19301 | 2.18787 | 1.92265 | 1.91338 |
| Difference variance | 1.72836 | 1.67137 | 1.83509 | 1.73961 |
| Difference entropy | 4.61256 | 4.51838 | 4.78272 | 4.63405 |
| Information measure of correlation 1 | -7.2285 | -7.27798 | -6.5625 | -6.66892 |
| Information measure of correlation 2 | 9.62953 | 9.63631 | 9.30402 | 9.32727 |
| Inverse difference normalized (INN) | 9.80925 | 9.81561 | 9.79926 | 9.88614 |
| Inverse difference moment normalized | 9.97342 | 9.97431 | 9.97181 | 9.97341 |

Table 3. Selected Features for Classification From 20 Features

| Features Computed | Normal | | Abnormal | | Mathematical Equation |
|-------------------------|----------|----------|----------|----------|---|
| | Sample 1 | Sample 2 | Sample 3 | Sample 4 | |
| Autocorrelation | 1.20143 | 1.20115 | 7.28408 | 7.28177 | $= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (p_x - \mu_x) (p_y - \mu_y) / \sigma_x \sigma_y$ |
| Sum of squares variance | 1.20107 | 1.19866 | 7.30073 | 7.28414 | $= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i,j)(i - \mu)^2$ |
| Sum average | 6.04537 | 6.04403 | 4.80535 | 4.80318 | $\sum_{i=0}^{2G-2} iP_{x+y}$ |
| Sum variance | 2.65242 | 2.65528 | 1.45384 | 1.45814 | $= \sum_{i=0}^{2G-2} (i - \sum_{i=0}^{2G-2} iP_{x+y})^2 P_{x+y}$ |
| Cluster prominence | 3.52433 | 3.52576 | 1.01281 | 1.01238 | $= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i,j) (i+j - \mu_x - \mu_y)^4$ |
| Cluster shade | 2.69382 | 2.69782 | 8.70367 | 8.72303 | $= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i,j) (i+j - \mu_x - \mu_y)^3$ |

4. Results

In this section, the performance of our ensemble-supervised classifier is investigated using the DDSM dataset provided by Massachusetts General Hospital, Boston, MA, USA, the University of South Florida, Tampa, FL, USA, and Sandia National Laboratories, Albuquerque, NM, USA.

Here, we used 200 data sets for training process, of which 100 data sets are chosen randomly for evaluation of classifier. Finally, the evaluation of data consists of sixty abnormal and forty normal data sets. By applying feature extraction and selection method on ROI in training data, six salience features were selected, which lead to appropriate accuracy. In the evaluation step, we only extracted these features from ROI in the test data and fed them to classifier. Finally, the obtained results were compared with the gold standards that were labeled as normal or abnormal by an expert. Sensitivity and specificity were used to investigate classifier performance;

$$(11) \quad \text{sensitivity} = \frac{TP}{TP+FN}$$

$$(12) \quad \text{specificity} = \frac{TN}{TN+FP}$$

Where TP = True positive, TN = True negative, FP = False positive, and FN = False negative, a 100% sensitivity is the theoretical desired prediction for the cancerous data. Also, a 100% specificity is the theoretical desired prediction for the non-cancerous data. The sensitivity and specificity of the proposed system are shown in Table 4.

To verify that our selected features are robust and our feature selection method is acceptable, we compared the obtained results from the proposed method with the results of random feature selection method. In random selection method, we assume that the data distribution is normal and therefore, by using "randm" we selected the random features. Obtained results are shown in Table 5.

Table 4. Measured Sensitivity and Specificity of the Proposed System With Maximum Difference Feature Selection ^a

| Test Outcome | Condition Positive | Condition Negative | PV |
|-----------------------------|--------------------|--------------------|--------------------------|
| Positive | 58 (TP) | 1 (FP) | 98.30 (PPV) ^b |
| Negative | 2 (FN) | 39 (TN) | 95.12 (NPV) ^c |
| Sensitivity/ Specificity | 96.66 | 97.50 | |

^a Abbreviations: PV, Predictive Value; TP, True Positive; FP, False Positive; FN, False Negative; TN, True Negative; PPV, Positive Predictive Value; NPV, Negative Predictive Value.

^b $PPV = TP/(TP+FP)$.

^c $NPV = TN/(TN+FN)$.

Table 5. Comparison of Proposed Method With Difference Feature Selection Method

| Breast Cancer Classifier Performance | Percent |
|--|---------|
| With maximum difference feature selection | |
| Specificity | 97.50 |
| Sensitivity | 96.66 |
| Accuracy | 97 |
| With randomly feature selection | |
| Specificity | 92.30 |
| Sensitivity | 91.80 |
| Accuracy | 92 |

5. Discussion

In this paper, we have proposed a new CAD method to classify the tumoral mammogram. This method is fully automatic and does not need operator manipulation. At first, we select the area of mammograms with high cancer probability. Selected area contains the suspected region which is given for feature extraction process. The extracted features are classified into normal and abnormal, using ensemble supervised classification method. The performance of the proposed method is evaluated by the perfect test method, which gives the sensitivity and specificity of the result. The sensitivity and specificity of the proposed method are 96.66% and 97.50%, respectively. The proposed classification method gave the correct classification of 97% for the division into two categories according to BI-RADS standard on the DDSM. The obtained accuracy of the proposed method, 97%, is comparable with KNN (96%), SVM (87%) and Naive Bayes (89%). In this paper, by assembling three classifiers and applying single voting policy, we improve the classification results in comparison to the method proposed by Luo et al. (13) The obtained results show that our method has a slight improvement over the other proposed methods on the dataset, which is publicly available. Therefore, the proposed method is more reliable in order to assist the radiologist in the detection of abnormal data and to improve the diagnostic accuracy.

Acknowledgements

We want to thanks Massachusetts General Hospital (D. Kopans, R. Moore), the University of South Florida (K. Bowyer), and Sandia National Laboratories (P. Kegelmeyer) for providing the Digital Database for Screening Mammography.

Authors' Contributions

Study concept and design: Hossein Yousefi Banaem and Dr. Alireza Mehri Dehnavi. Analysis and interpretation of data: Hossein Yousefi Banaem. Drafting of the manuscript: Hossein Yousefi Banaem. Revision of the manu-

script for important intellectual content: Dr. Alireza Mehri Dehnavi. Clinical counselor: Dr. Makhtum Shahnazi.

Funding/Support

This study was supported by Isfahan University of Medical Sciences, Isfahan, Iran.

References

1. Verma B, McLeod P, Klevansky A. Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer. *Expert Syst Appl*. 2010;**37**(4):3344-51.
2. Shirazi Noodeh A, Ahmadi Noubari H, Mehri Dehnavi A, Rabani H. Application of wavelets and fractal-based methods for detection of microcalcification in mammograms: a comparative analysis using neural network.; *International Conference on Graphic and Image Processing (ICGIP 2011)*, 82857E.; September 30, 2011; 2011. p. 82857E.
3. Luukka P. Feature selection using fuzzy entropy measures with similarity classifier. *Expert Syst Appl*. 2011;**38**(4):4600-7.
4. Nithya R, Santhi B. Classification of normal and abnormal patterns in digital mammograms for diagnosis of breast cancer. *Int J Comput Appl T*. 2011;**28**(6):21-5.
5. Nithya R, Santhi B. Mammogram Classification Using Maximum Difference Feature Selection Method. *J Theor Appl Inform Technol*. 2011;**30**(2):21-25.
6. Yachoub MA, Mohamed AS, Kadah YM. A CAD system for the detection of malignant patterns in digitized mammogram films. *CARIO International Biomedical Engineering Conference*. 2006.
7. Verma B, Zhang P. A novel neural-genetic algorithm to find the most significant combination of features in digital mammograms. *Appl Soft Comput*. 2007;**7**(2):612-25.
8. Alolfe MA, Mohamed WA, Youssef A, Kadah YM, Mohamed AS, editors. Feature selection in computer aided diagnostic system for microcalcification detection in digital mammograms.; *Radio Science Conference, 2009. NRSC 2009. National.*; 2009; IEEE; pp. 1-9.
9. Chen HL, Yang B, Liu J, Liu DY. A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. *Expert System Appl*. 2011;**38**(7):9014-22.
10. Vasantha M, Subbiahbarathi V. Classification of mammogram images using hybrid features. *Eur J Sci Res*. 2011;**57**(1):87-96.
11. Huang CL, Liao HC, Chen MC. Prediction model building and feature selection with support vector machines in breast cancer diagnosis. *Expert Syst Appl*. 2008;**34**(1):578-87.
12. Prathibha BN, Sadasivam V. A kernel discriminant analysis in mammogram classification using with texture features in wavelet domain. *Int J comput intell*. 2010;**1**(1):146-156.
13. Luo ST, Cheng BW. Diagnosing breast masses in digital mammography using feature selection and ensemble methods. *J Med Syst*. 2012;**36**(2):569-77.
14. Wei L, Yang Y, Nishikawa RM. Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis. *Pattern Recognit*. 2009;**42**(6):1126-32.
15. Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer WP, editors. In Proceedings of the Fifth International Workshop on Digital Mammography; International Workshop on Digital Mammography; 2001; Medical Physics Publishing; pp. 212-8.
16. Niwas SI, Palanisamy P, Sujathan K, editors. Wavelet based feature extraction method for breast cancer cytology images.; *Industrial Electronics & Applications (ISIEA), 2010 IEEE Symposium on.*; 2010; pp. 686-90.
17. Bei H, Lin J, PengCheng X, editors. The research on diagnosis by mammography basing on Bayesian Classification.; *Computer Application and System Modeling (ICCASM), 2010 International Conference on.*; 2010; pp. V11-490-V11-493.
18. Braz Junior G, Cardoso de Paiva A, Correa Silva A, Cesar Muniz de Oliveira A. Classification of breast tissues using Moran's index and Geary's coefficient as texture signatures and SVM. *Comput Biol Med*. 2009;**39**(12):1063-72.