

Wavelet and Symmetric Stochastic Neighbor Embedding based Computer Aided Analysis for Breast Cancer

S. Mohan Kumar^{1*} and G. Balakrishnan²

¹Department of Computer Science and Engineering, New Horizon College of Engineering, Outer Ring Road, Near Marathalli, Bellandur Main Road, Bengaluru – 560103, Karnataka, India; drsmohankumar@gmail.com

²Indra Ganesan College of Engineering, Madurai Main Road (NH-45B), Manikandam, Thiruchirappalli – 620012, Tamil Nadu, India

Abstract

Mammography is the most perceptive method for the detection of early breast cancer. The abnormalities of breast are analyzed by digital mammogram images and the most important indicators of breast malignancy are microcalcifications and masses. An efficient Computer Aided Diagnosis (CAD) system for breast cancer classification is proposed in this study based on Discrete Wavelet Transform (DWT), Symmetric Stochastic Neighbor Embedding (SSNE) and Support Vector Machine (SVM) using digital mammogram images. Two technical approaches are employed for feature selection from the wavelet decomposed mammogram for classification. They are based on the application of SSNE over the decomposed image. At first, SSNE is applied to the whole wavelet decomposed image whereas in the second technique it is applied to individual sub band of the wavelet decomposed image. The whole mammogram classification system is implemented in two consecutive stages. The first stage of the proposed system classifies the mammogram image into normal or abnormal. The severity of the predicted abnormality is further classified either it is benign or malignant associated with mass or microcalcification images. The performance of the proposed mammogram classification system is evaluated using Mammographic Image Analysis Society (MIAS) database images.

Keywords: Digital Mammogram, Discrete Wavelet Transform, Mass, Microcalcification, Symmetric Stochastic Neighbor Embedding, Support Vector Machine

1. Introduction

A computer aided diagnostic system based on new texture shape feature coding is used to classify the masses is explained in¹. The mass classification on mammogram presents a great challenge for design of computer aided diagnostic systems due to the complexity of mass background and its mammographic characteristics. SVM is used for the classification. Manly

transformation and principal component analysis is used to reduce the high dimensional data. The classifiers like SVM and Bayesian Networks are trained and tested using a k-fold cross-validation method. Ranklet is used as a classification features for classifying the masses in³. The dimensionality of the classification problem is reduced by using a recursive feature elimination technique without affecting the classification performances.

*Author for correspondence

Artificial neural network is used to classify the masses, which performs benign-malignant classification on region of interest that contains mass is explained in⁴. Texture is the major characteristics of mass classification. The textural features used for characterizing the masses are mean, standard deviation, entropy, skewness, kurtosis and uniformity. This method is used to increase the effectiveness and the efficiency. Texture analysis based on curvelet transform for the classification of mammogram tissues is presented in⁵. The most discriminative texture features of regions of interest are extracted. Then, a nearest neighbor classifier based on Euclidian distance is constructed. The obtained results calculated using 5-fold cross validation. The approach consists of two steps, detecting the abnormalities and then classifies the abnormalities into benign and malignant tumors.

Computerized schemes to classify masses by optimizing certain criteria to classify cases into one of mutually exclusive classes are developed in⁶. It is based on Linear Discriminant Analysis (LDA). The features required to distinguish the benign from malignant masses are abnormality dependent. LDA in mixture with stepwise feature selection is trained and tested on morphological features extracted using the machine segmentation and radiologist segmentation⁷. A novel opposition based classifier is developed⁸ which classifies breast masses into benign and malignant categories. A Multi Layer Perceptron (MLP) network with a novel learning rule, called Opposite Weighted Back Propagation (OWBP), is utilized as classifier. The features include circularity, Zernike moments, contrast and average gray level.

The fractal modeling of the mammographic images and their background morphology is presented in⁹. For fractal modeling, the original image is first segmented into appropriate fractal boxes followed by identifying the fractal dimension of each windowed section. Then two-dimensional box counting algorithm is used which is based on the order of the computations; they are placed in an appropriate matrix to facilitate the required computations. Finally using eight features identified as characteristic features of tumors extracted from mammogram images.

A novel semi-supervised k-means clustering is proposed for outlier detection in mammogram classification in¹⁰. The shape features are extracted from the digital mammograms, and k-means clustering is

applied to cluster the features, the number of clusters is equal with the number of classes. A novel genetic association rule miner is applied with this reduced feature set to construct the association rules for classification. The performance is analyzed with rough set using receiver operating characteristic curve analysis.

A new classification approach using SVM for the detection of microcalcification clusters in digital mammograms is proposed in^{11,16,17,18}. Classifying data is a common task in machine learning. The microcalcification detection is formulated as a supervised learning problem and applies SVM as a classifier to determine at each pixel location in the mammogram if the microcalcification is present or not. In this study, the classification of breast malignancy, microcalcifications and masses is proposed based on DWT, SSNE and SVM.

2. Methodology

The proposed CAD system for mammogram classification is implemented based on DWT, SSNE and SVM.

3. Proposed Method

The proposed CAD system that uses digital mammogram images for the diagnosis of breast cancer composed of three modules; feature extraction, feature selection and classification. These modules are employed for both microcalcification and mass system individually. All the stages are explained in detail in the following sub sections

3.1 Feature Extraction

Feature extraction is an essential pre processing step for variety of machine learning and pattern recognition approaches. The first stage of the proposed mammogram classification system is feature extraction in which features are extracted by applying DWT on digital mammogram images. At the first, the Region Of Interest (ROI) mammogram image is decomposed by DWT at various levels of decomposition from 2 to 6. It decomposes the image into approximation and detailed sub bands, in which k level DWT produces $3k + 1$ sub bands. Figure 1. shows a sample mammogram image and its 2-level and 3-level wavelet decomposition.

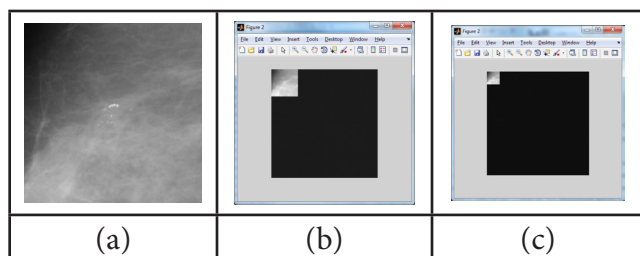


Figure 1. Digital mammograms decomposed by using DWT (a) original image (b) 2-level decomposition (c) 3-level decomposition.

The decomposition produces vast number of wavelet coefficients same as the size of the image. These high dimensional wavelet coefficients make the classification algorithm more complex. In order to ease the problem, the high dimensional feature space is reduced into low dimension space by dimension reduction techniques.

3.2 Feature Selection

Feature selection is a key process for simple and efficient breast cancer classification, where the computation complexity of the system is reduced. In order to improve the efficiency of feature selection, the proposed mammogram classification system employed SSNE to predict whether the features are relevant or not. After feature extraction the informative features are found out by SSNE from the extracted wavelet coefficients. The course of action is applied for all the training normal, abnormal and associated severity mammogram images. And the selected features are stored in database for further classification. The process of feature extraction and database creation is carried out separately for mass and microcalcifications images.

3.3 Classification Process

SVM is a powerful supervised learning classification approach, which is used as a classifier. Two-stage classification approach is implemented in the proposed mammogram classification system. In the first stage of classification, the sample mammogram image is classified into whether it is associated with normal or abnormal one by comparing the database 1 and the extracted unknown sample feature space. The features are extracted from the unknown sample image as same as what have done for training images. If the first stage classifier categorizes the unknown image as abnormal, it goes to the second stage of the classification. Database 2 is used to categorize the

abnormal image into associated severity such as whether it is benign or malignant. The classification process is done in separate manner for mass and microcalcification images with wavelet and wavelet sub band approaches.

4. Experimental Results

In order to evaluate the performance of the proposed system, experiments with MIAS database¹⁵ images are performed. The classification test for microcalcification and mass images are performed separately. The performance of the microcalcification system is carried on 99 normal images and 25 microcalcification images. Among the 25 abnormal images, there are 12 benign and 13 malignant images available. All the images are considered for the classification test. The number of samples used for training the SVM classifier and testing images for classification is given in Table 1. Figure 2. shows 2 sample training images of normal, benign and malignant categories of microcalcification and similarly Figure 3. shows 2 sample test images of each category.

Table 1. Number of training and testing samples for microcalcification system

Category	No. of Training Set	No. of testing Set
Normal	66	99
Abnormal	17	25
Microcalcification (Benign)	8	12
Microcalcification(Malignant)	9	13

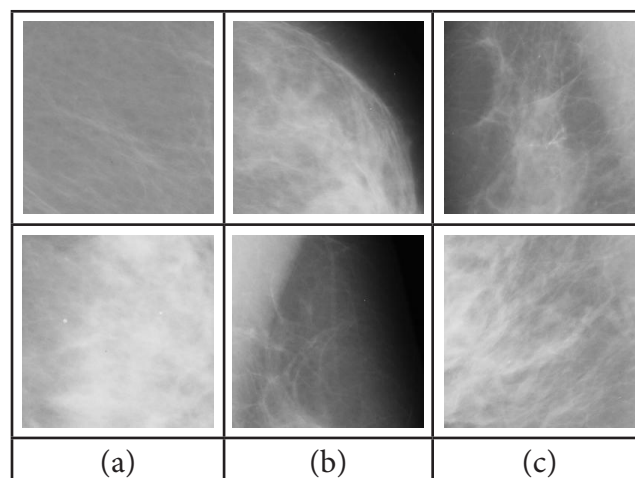


Figure 2. Sample training images used in microcalcification system (a) normal (b) benign (c) malignant.

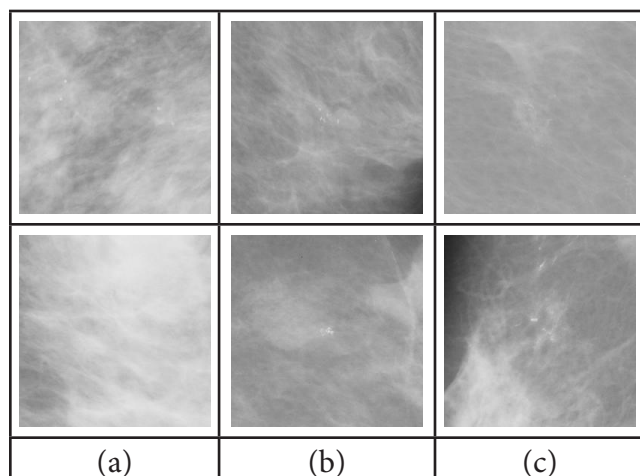


Figure 3. Sample testing images used in microcalcification system (a) normal (b) benign (c) malignant.

The performance of the mass system is carried on 99 normal images used in the microcalcification system and 56 mass images. Among the 56 abnormal images, there are 37 benign and 19 malignant images available. All the images are considered for the classification test. The number of samples used for training the SVM classifier and testing images for mass classification is given in Table 2. Figure 3. shows 2 sample training images of normal, benign and malignant images of mass categories and similarly Figure 4. shows 2 sample test images of each category.

Table 2. Number of training and testing samples for mass classification

Category	No. of Training Set	No. of testing Set
Normal	66	99
Abnormal	37	56
Masses(Benign)	25	37
Masses(Malignant)	13	19

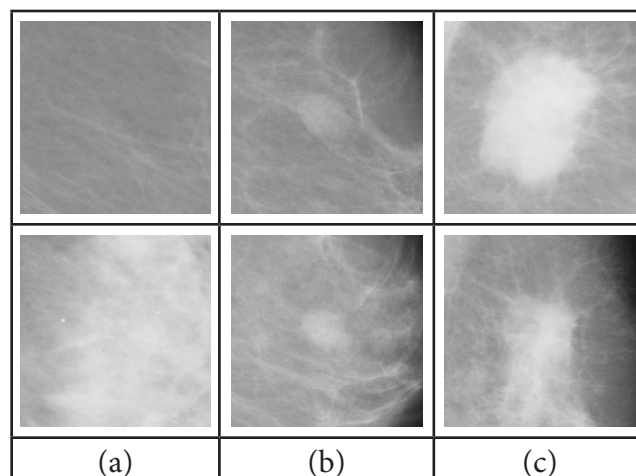


Figure 4. Sample training images used in mass system (a) normal (b) benign (c) malignant.

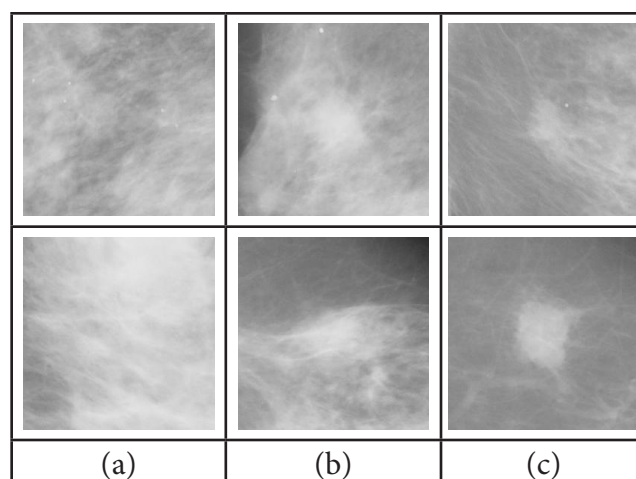


Figure 5. Sample testing images used in mass system (a) normal (b) benign (c) malignant.

Two technical approaches based CAD for mammogram classification system is presented in this study. At first, the feature selection technique SSNE is applied for

Table 3. Classification accuracy of first stage using Wavelet, SSNE and SVM

Level of decomposition	Wavelet + SSNE + SVM					
	Microcalcification			Mass		
	Normal (%)	Abnormal (%)	Average (%)	Normal (%)	Abnormal (%)	Average (%)
2	98.99	84	91.49	95.95	87.5	91.72
3	100	80	90	97.97	82.14	90.06
4	98.99	84	91.49	92.92	87.5	90.21
5	98.99	76	87.49	94.94	83.92	89.43
6	98.99	84	91.49	91.91	85.71	88.81

Table 4. Classification accuracy of final stage using Wavelet, SSNE and SVM

Level of Decomposition	Wavelet + SSNE + SVM					
	Microcalcification			Mass		
	Benign (%)	Malignant (%)	Average (%)	Benign (%)	Malignant (%)	Average (%)
2	91.67	84.62	88.14	100	84.21	92.10
3	91.67	92.31	91.99	97.29	89.47	93.38
4	91.67	84.62	88.14	97.29	89.47	93.38
5	100	84.62	92.31	97.29	89.47	93.38
6	91.67	84.62	88.14	97.29	84.21	90.75

Table 5. Classification accuracy of first stage using Wavelet sub-band, SSNE and SVM

Level of Decomposition	Wavelet sub-bands + SSNE + SVM					
	Microcalcification			Mass		
	Normal (%)	Abnormal (%)	Average (%)	Normal (%)	Abnormal (%)	Average (%)
2	100	68	84	97.97	78.57	88.27
3	100	72	86	97.97	80.35	89.16
4	100	72	86	94.94	82.14	88.54
5	100	72	86	95.95	80.35	88.15
6	100	68	84	98.98	78.57	88.78

Table 6. Classification accuracy of final stage using Wavelet sub-band, SSNE and SVM

Level of decomposition	Wavelet sub-bands + SSNE + SVM					
	Microcalcification			Mass		
	Benign (%)	Malignant (%)	Average (%)	Benign (%)	Malignant (%)	Average (%)
2	83.33	100	91.67	100	78.94	89.47
3	91.67	92.31	91.99	100	78.94	89.47
4	91.67	100	95.83	100	78.94	89.47
5	83.33	100	91.67	100	84.21	92.10
6	83.33	100	91.67	100	78.94	89.47

whole wavelet decomposed image and the corresponding features are stored for classification. In the second approach, SSNE is applied to individual sub band of the wavelet decomposed image and then the reduced features of each sub-bands are fused together and given to the classifier. The classification accuracy obtained by using DWT and SSNE for the classification of microcalcification and mass in digital mammograms using SVM classifier is shown in Table 3.

For the classification of microcalcification and mass, the maximum classification accuracy obtained is 91.49% and 91.72% respectively at 2nd level of decomposition. The improvement in the classification accuracy at higher level decomposition is insignificant. Table 4. shows the

classification accuracy obtained by SVM classifier at final stage of classification.

It is clearly observed from the Table 4, microcalcification benign patterns are classified very accurately with 100% classification and the mass benign patterns having 2.71% misclassification. In both cases, the maximum classification accuracy is obtained at 5th level of wavelet decomposition. The performance of this approach is better for mass classification as the classification accuracy obtained by the mass system is consistent with over 90%. The performance of SSNE is again evaluated based on the computation of SSNE on wavelet sub-bands individually. Table 5. shows the classification accuracy obtained by this approach at first stage of classification.

The application of SSNE on wavelet sub-bands produces better accuracy for normal cases of about 100% and 97.97%. However, this approach fails to classify the abnormal images of over 20%. In overall, the average accuracy obtained is below 89% in microcalcification and mass classification. Table 6 shows the classification accuracy obtained at final stage of classification.

It is clearly observed from the Table 6. mass benign patterns and microcalcification malignant patterns are correctly classified. The wavelet sub-bands produces better accuracy for microcalcification of over 91.67% than the mass system that achieves less than 92.10%. From the analysis, for mass classification system the wavelet based approach produces better result than wavelet sub-band based approach. At first stage classification of microcalcification system, the wavelet based approach gives higher accuracy and wavelet sub-band based approach is best for final stage classification.

5. Conclusion

Computer aided diagnosis system for microcalcification and mass classification using digital mammogram images is proposed in this study. Feature extraction and feature selection is carried out by DWT and SSNE techniques. The wavelet features are selected by SSNE in two ways and their performances are evaluated by using SVM classifier. The first stage of the system classifies the image into normal/abnormal, correspondingly second stage classifier categorizes the severity of the abnormality into benign/malignant. Experimental results show that the proposed microcalcification and mass system achieves over 90% classification accuracy.

6. References

1. Chen Y, Chang CI. New texture shape feature coding-based computer aided diagnostic methods for classification of masses on mammograms. Proceedings of the 26th Annual International Conference of the Institute of Electrical and Electronics Engineers (IEEE); 2004 Sep.
2. Samulski M, Karssemeijer N, Lucas P, Groot P. Classification of mammographic masses using support vector machines and Bayesian networks. Proceedings of SPIE on Medical Imaging. 2007; 6514:1–11.
3. Masotti M. Exploring rank lets performances in mammographic mass classification using recursive feature elimination. Proceedings of the 16th Institute of Electrical and Electronics Engineers (IEEE) Signal Processing Society Workshop on machine learning for signal processing; 2006. p. 265–9.
4. Islam MJ, Ahmad M, Maher A. Sid-Ahmed. An efficient automatic mass classification method in digitized mammograms using artificial neural network. International Journal of Artificial Intelligence and Applications (IJAIA). 2010 Jul; 1(3):1–13.
5. Eltoukhy MM, Faye I. Curvelet based feature extraction method for breast cancer diagnosis in digital mammogram. Institute of Electrical and Electronics Engineers (IEEE) International Conference on Intelligent and Advanced Systems; 2010 Jun. p. 1–5.
6. Cheng HD, Shi XJ, Min R, Hu LM, Cai XP, Du HN. Approaches for automated detection and classification of masses in mammograms. Pattern Recognition. 2006; 39:646–68.
7. Sahiner B, Petrick N, Chan HP. Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization. Institute of Electrical and Electronics Engineers (IEEE) Transactions on Medical Imaging. 2001; 20(12):1275–84.
8. Saki F, Tahmasbi A. A novel opposition-based classifier for mass diagnosis in mammography images. Institute of Electrical and Electronics Engineers (IEEE) Iranian Conference of Biomedical Engineering; 2010 Nov. p. 1–4.
9. Noodeh AS, Rabbani H. Detection of cancerous zones in mammograms using fractal modeling and classification by probabilistic neural network. Institute of Electrical and Electronics Engineers (IEEE) Iranian Conference of Biomedical Engineering; 2010 Nov. p. 1–4.
10. Thangavel K, Mohideen AK. Semi-supervised K-means clustering for outlier detection in mammogram classification. Institute of Electrical and Electronics Engineers (IEEE) Trendz in Information Sciences and Computing. 2010 Dec:68–72.
11. Dheeba J, Selvi ST. Classification of malignant and benign microcalcification using SVM classifier. Institute of Electrical and Electronics Engineers (IEEE) International Conference on Emerging Trends in Electrical and Computer Technology; 2011 Mar. p. 686–90.
12. Hinton G, Roweis S. Stochastic neighbor embedding. In Advances in Neural Information Processing Systems, MIT Press. 2003; 15:833–40.
13. Rejani YIA, Selvi ST. Early detection of breast cancer using SVM classifier technique. International Journal on Computer Science and Engineering. 2009; 1(3):127–30.
14. Smola AJ, Bartlett P, Scholkopf B, Schuurmans D. Advances in large margin classifiers. MIT press; 2000.
15. <http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html>.

16. Kumar SM, Balakrishnan. Classification of breast mass classification – CAD system and performance evaluation using SSNE. *International Journal of Innovative Science, Engineering and Technology (IJSET)*. 2015 Sep; 2(9): 417–25.
17. Kumar SM, Balakrishnan. Classification of breast mass classification – CAD system with performance evaluation. *International Journal of Engineering And Computer Science*. 2015 Sep; 4(9):14187–93.
18. Kumar SM, Balakrishnan. Classification of breast micro-calcification- CAD system and performance evaluation using SSNE. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2015 Sep; 5(9):824–30.