Methods towards the Classification of Clustered Microcalcification

S. Narasimhamurthy,
Research Scholars
Dept. of Electronics and Communication Engg.
P.E.S. C.E, Karnataka, India-571401

Arun kumar
Research Scholars
Dept. of Electronics and Communication Engg.
P.E.S. C.E, Karnataka, India-571401

H.S. Sheshadri
Professor
Dept. of Electronics and Communication Engg.
P.E.S. C.E, Karnataka, India-571401

ABSTRACT
Breast cancer is one of the leading causes of death among the women. Mammogram analysis is the most effective method that helps in the early detection of breast cancer. Microcalcification, masses, and architectural detection in the mammogram plays an important role in the later stages of diagnosis. In this paper we propose an effective method for the detection and classification of clustered microcalcification. We applied the proposed method in the MIAS datasets and found the effectiveness in the detection and classification of clustered microcalcification. We also brief out in this article the methods adopted to select the features for clustered microcalcification and technique to handle the class imbalance specific to microcalcification classification problem.

Keywords
Classification, Clustered Microcalcification, Imbalanced data sets, Mammography.

1.INTRODUCTION
Breast cancer is the second leading cause of cancer death among women and it is the leading cause of cancer deaths among women in the 40-55 ages. Mammography [1, 2, 3, 4] is a key screening tool for breast abnormalities detection, because it allows identification of tumor before being palpable. Interpretation of mammographic images are difficult for the radiologists because of poor quality and noises in the images.

With the advances of computer technology radiologists have an opportunity to improve their image interpretation because of its capabilities that can enhance the image quality of mammograms. Many computerized automatic schemes for detecting and diagnosing early stage micro calcification is developed. Several CAD systems that support MC detection have been deployed for clinical use. Literature reports show mixed results on the role of current CAD systems in practice, with some showing improvement and others showing no improvement. Some of these systems may tend to overemphasize the sensitivity in their detection ability at the expense of specificity. This may result in increased unnecessary biopsies when using such CAD systems. Most of the research work on classification of micro calcification cluster deal with feature extraction, and then classify using a suitable classifier. Until now no standard minimum feature set that accurately classify (> 90%) has been proposed. Only a particular type of features is not a good choice for the classification. It has found that a combination of different features from statistical and structural will leads to an accurate classification. Better understanding of MC characteristics as perceived by experts should also be considered for mammogram analysis.

Number of approaches [5, 6, 7, 8, 9, 10] are developed to classify the microcalcification as malignant and benign. However, the area under the ROC curve achieved by all the techniques is 0.85 maximum. The classification accuracy is reduced in most cases because the difficulty in dealing with imbalanced training set used for classification. Refer [11] to see methods summary on class imbalanced data problems. Since the classification of potential microcalcification has an issue of the class imbalance problem as we have highlighted in our work [11] we deal it with a new methods i.e. SMOTE + CMTNN. This method accurately classifies the potential microcalcification so that more accuracy is achieved in the next phase of clustered microcalcification classification.

The remaining sections are organized as follows. Previous works carried out for the microcalcification are summarized in section 2. Proposed methodology is explained in section 3. Section 4 details the experimental results. Conclusions are drawn in section 5.

2. LITERATURE SURVEY
Maurice Samulski [12] has developed a method to detect lesions in digital mammogram using temporal Bayesian classifiers. To develop the system he has included medical background knowledge of breast cancer domain. He argues that Bayesian classifiers often outperform more sophisticated network structure as well as other types of classifiers. The results obtained using this method is compared with existing techniques such as Support Vector Machines and Neural Networks by using FROC analysis method. A Oliver et al. [13] proposed an automatic classification method for identification of breast tissue in mammographic images. Classification is based on gross segmentation and the underlying tissue. The proposed method uses K-means algorithm for segmentation, k-NN and ID3 for classification. W Sheta et al. [14] proposed a method to classify the breast masses and normal tissues in screening mammograms using Linear Genetic Programming. Genetic Programming is capable of discovering complex solutions that cannot be easily captured by other classification techniques such as decision trees, statistical classifiers and neural networks. The classification was based on image features extracted from selected regions of interest using directional neighbourhood analysis. They have evaluated the LGP (Linear Genetic Programming) classifier using Receiver Operating Characteristics. E. Alberdi et al. [15] have developed a new

method for automatically detecting the abnormalities in a digitized mammogram. CADMIUM II, a system for the diagnosis of mammograms combines image processing with symbolic representations of clinical decisions. They have designed a scheme with 50 descriptors. This scheme was used by the radiologists to describe 40 sets of calcifications. Decision support presented in the system at a level of description is both relevant and meaningful to the user.

Monika shinde [16] proposed a new method for classification of mass and normal tissue. The Expectation Maximization (EM) method developed, automatically separate mass tissue from normal breast tissue in a digitized mammogram. Both the raw data and summary data generated by Law's texture analysis are investigated. The EM method achieved a 63% of sensitivity and 89% of specificity based on the 300 image data set consisting of 97 malignant and 203 benign cases. Renato Campanini et al. [17] proposed a novel approach to mass detection in digital mammograms. For the detection of region of interest they have not extracted any features, but exploited all the information available on the image. The detection task is considered as a two-class pattern recognition problem. They claimed that the sensitivity of the proposed system is nearly 80% with a false positive rate of 1.1 marks per image, estimated on images coming from the USF DDSM database. Discrete wavelet transform mod-max method is applied to the problem of mammographic mass classification by Bruce and Roza R Adhani [18]. Multi resolution features were extracted using this method. They developed three new features: - variation ratio mean, variation ration standard variation, and the Lipschitz sum to classify mammographic mass shapes. Strausz et al. [19] developed a CAD system for automatic diagnosis. This method uses hierarchical model of image features and neural networks to identify individual microcalcifications. They mainly used images from the MIAS digital mammography database for experimentation. They found that image features and network architecture have influence on the performance of the system. They have discussed the importance and problem of screening mammography.
H.S.Sheshadri et al. [20] have developed an image processing algorithm for detection of micro-calcifications and also a computer based decision system for early detection of breast cancer. They have suggested certain methods for texture feature extraction in digital mammograms. These methods are based on the digital filters together with a filter response energy measure as texture feature extractors. In their work they have made an attempt to improve the classification performance of the co-occurrence approach. The system is capable of detecting micro-calcifications by visual inspection of digital mammograms. H.D Cheng et al. [21] proposed a novel approach for the detection of microcalcification based on fuzzy logic. They also included the scale space techniques along with fuzzy logic. To fuzzify the image the fuzzy logic and fuzzy set theory is used. To detect the size and location of microcalcification Scale Space and Laplace Filter techniques are used. Using the developed technique microcalcification can be accurately detected. Leungng et al. [22] proposed a method for the detection and classification of two types of breast tumors. The method make an assumption that both type of breast tumors i.e., stellate lesions and circumscribed lesions appear approximately circular and bright with a fuzzy boundary. The method proposed by Andreadis et al. [23] use an approach based on the binary methodology support vector machines (SVM) for the classification and characterization of clustered micro-calcifications in digitized mammograms, using CAD system. They have tested the performance of various SVM schemes and compared them with the existing CAD system using a database of 155 (118 benign and 37 malignant) clinical mammograms provided from collaborating diagnostic centres focused on breast examination. Experimental results show that the method is accurate in classification.

3.0 PROPOSED METHODOLOGY

The presence of Micro calcification clusters are primary indicators of early stages of malignant types of breast cancer and its detection is important to prevent the disease. The goal of the proposed model is to identify the micro calcification clusters and also classify in order to determine which ones are malignant and which ones are benign.

For our work we select the mammograms provided by the Mammography Image Analysis Society (MIAS) [24].

The diagram of the proposed method is shown in figue1. The proposed system indicates suspicious regions with a strong likelihood of micro calcification cluster presence and classifies these clusters as benign or malignant.

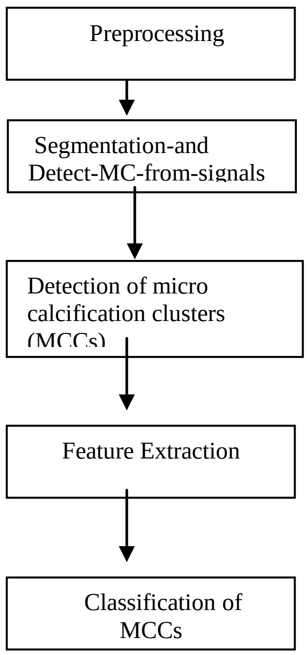

Figure 1. Proposed Method for micro calcification clusters

The preprocessing reduces the work area to a region that contains only the breast area so that background and isolated regions are deleted. The image after the segmentation step is a binarized image with potential micro calcification (signals). In the next stage these signals (potential microcalcification) are classified as microcalcification or not-microcalcification. In this stage the classification of potential microcalcification is done with SMOTE+CMTNN. Here we used one combination of SMOTE+CMTNN to handle the class imbalance in the classification of potential microcalcification. During the phase of *detection of micro calcification clusters* the micro calcification clusters are identified. The classification phase, classify the micro calcification clusters as benign or malignant. In this stage micro calcification clusters feature set is extracted and accurately classify them into benign or malignant using a classifier. The steps from preprocessing to the microcalcification detection and classification are discussed in our previous work [11].

3.1 Microcalcification detection and cluster Feature Selection

After the individual microcalcifications are detected an algorithm adds microcalcifications to their closest clusters. This is done at a reasonable distance until there are no more microcalcifications left. The entire detected cluster is then labelled using a region-labelled algorithm From the clusters 17 features are extracted. The features are Number of microcalcifications in the cluster, Average area of MC, Biggest area in the cluster, Smallest area in the cluster, Range of areas in the cluster, SD of the area, SD of the microcalcifications' distances, Variance of the average distance, Cluster Entropy, Average Entropy of microcalcifications, Standard Deviation of the Entropy, Average Cluster Intensity, SD of Average Intensity, Cluster Circularity, minimum diameter, minimum radius, and mean radius of the clusters. These features are passed through a feature selection process. We used WEKA tool for this feature selection. These features are selected using consistencySubsetEval + GreedyStepwise method of WEKA. We used only 4 features for the classification of clustered microcalcification.

3.2 Microcalcification cluster classification

In this phase microcalcification clusters are classified as benign or malignant. We used SVM as the classifier. The proposed model promises the improvement over the existing model by the following added features

1. Decreased unnecessary biopsies when using the CAD systems.
2. Accurate classification of the micro calcification clusters into benign or malignant using a less number of features from statistical and structural features
3. The set of features that optimally describe a mammogram Region Of Interest (ROI) is designed.
4. Better understanding of Micro Calcification characteristics as perceived by experts is also considered.

4. EXPERIMENTAL ANALYSIS

For our experimental work we have used MIAS database. The following sections details our experimental results

4.1 Preprocessing

We have applied some preprocessing steps to remove the artifacts, and noises in the mammograms. Also the pectoral muscle is segmented to reduce the processing area. .

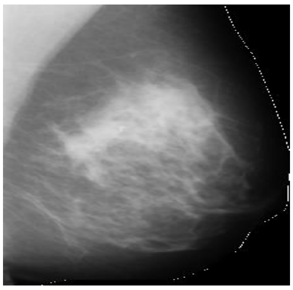

Fig 2. Breast contour superimposed on the original image

Wavelet decomposition of fourth level is used for the segmentation of pectoral muscle and edge detection is done with Sobel. Figure 3 shows the result of these steps in a MIAS image.

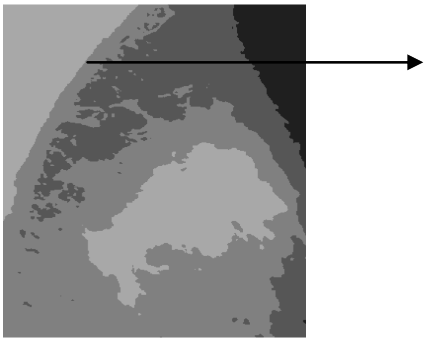

Fig 3. Arrow shows Pectoral muscle identified in the image

Identified pectoral muscle edge is superimposed on original image. For more details of our work refer [62]. Then binary images created from each segmented image helps in automatic cropping so the image will have the required ROI. Figure 4 shows some ROI extracted from the images.

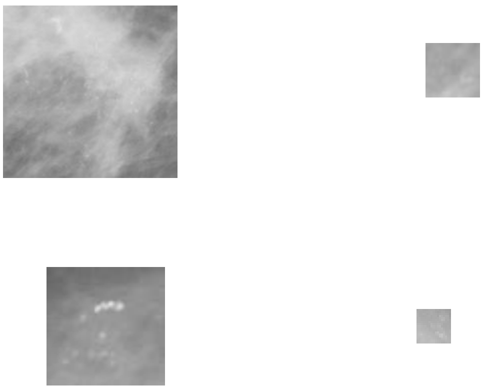

Figure 4. ROIs extracted from the images.

4.2 Detection of points and Signals (Potential Microcalcification)

Here we use only 22 MIAS images preprocessed. These images contain the microcalcification inside the boundaries of breast. To detect the points, Difference of Gaussian filters are applied sequentially to the images. These points are classified to signals (potential microcalcification) using some selection criteria based on the values calculated on pixels. Figure 5 shows the application of DoG filters in a ROI.

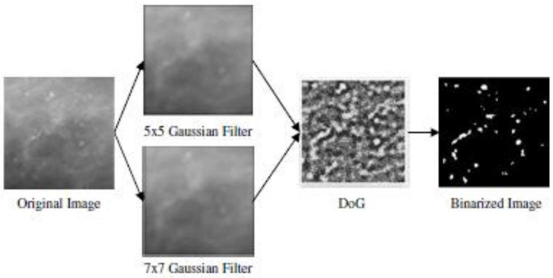

Figure 5. Application of DoG to ROI

4.3 Classification of Signals

The main objective of this phase is to classify the potential microcalcification to true microcalcification or not. With the available truth information and experts we have created a datasets which consists of only 4% of microcalcification and 96% of not-microcalcification. This dataset is an example for class imbalance. Since this biclass problem should deal with a data sampling approach we used SMOTE + Complementary Neural Network (CMTNN) and to assist the classification we have used SVM classifier. The accuracy achieved in this phase is more effective than any other CAD system.

4.4 Detection and Classification of clustered microcalcification cluster

After the signals are classified the individual microcalcifications are clustered. This clusters are classified as benign or malignant. Figure 6 shows some benign microcalcification clusters and figure 7 shows some malignant microcalcification clusters.

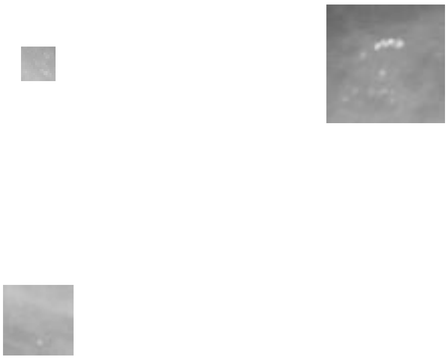

Figure 6. Benign clustered microcalcification.

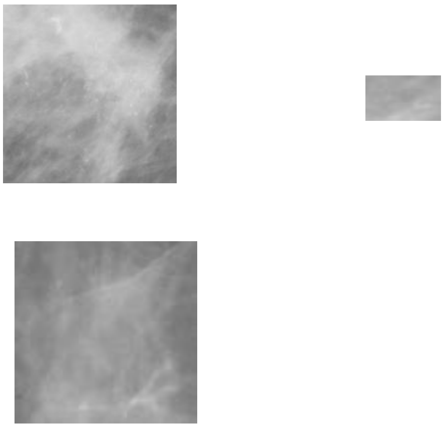

Figure 7. Malignant clustered microcalcification.

4. CONCLUSION

There are many methods developed for the microcalcification cluster detection and classification. The classification of clustered microcalcification will be effective if the detection of individual microcalcification is done effectively but many existing methods does not handle the class imbalance in the classification of individual microcalcification. To classify the individual microcalcification from the identified signals initially we used SMOTE+CMTNN sampling method because the classification of signals to microcalcification is a class imbalance problem and then SVM classifier. Since SMOTE+CMTNN effectively handled the data sampling for further classification of individual microcalcification using SVM, more efficiency is achieved in the classification of clustered microcalcification as benign or malignant.

5. REFERENCES

[1]. Wei D, Chan H.P, Helvie M.A, Shahiner B, Petrick N, and Alder D.D, "Classification of Mass and Normal Breast Tissue on Digital Mammograms", Med Physics, 22(9), 1501-13, Sept. 1995.

[2]. Undrill P.E, Gupta R, and Henry S, "The Use of Texture Analysis and Boundary Refinement to Delineate

Suspicious Masses in Mammography", SPIE, Medical Imaging: Image processing, SPIE 2710, 301-310, 1996.

[3]. American Cancer Society, "Cancer Facts and Figures", GA: American Cancer Society, 2003.

[4]. Mansees B S, "Evaluation of Breast Microcalcifications", The Radiologic Clinics of North America, Breast Imaging, Vol 33, 6, 1109-1121, Jan 1995 .

[5] I. El-Naqa, Y. Yang, N.P. Galatsanos, et al., A similarity learning approach to content-based image retrieval: application to digital mammography, IEEE Trans. Med. Imaging 23 (10) (2004) 1233–1244.

[6] I. El-Naqa, Y. Yang, M.N. Wernick, et al., A support vector machine approach for detection of microcalcifications, IEEE Trans. Med. Imaging 21 (12) (2002) 1552–1563.

[7] J. Ge, B. Sahiner, L.M. Hadjiiski, H.-P. Chan, J. Wei, M.A. Helvie, C. Zhou, Computer aided detection of clusters of microcalcifications on full field digital mammograms, Med. Phys. 33 (8) (2006) 2975–2988

[8] L. Wei, Y. Wei, Y. Yang, R.M. Nishikawab, Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis, Pattern Recogn. 42 (6) (2009) 1126–1132.

[9] L. Wei, Y. Yang, R.M. Nishikawa, et al., Relevance vector machine for automatic detection of clustered microcalcifications, IEEE Trans. Med. Imaging 24 (10) (2005) 1278–1285.

[10] K. Thangavel, M. Karnan, R. Sivakumar, A.K. Mohideen, Automatic detection of microcalcification in mammograms – a review, ICGST-GVIP J. 5 (5) (2005) 31– 61.

[11] Arun kumar M.N and H.S Sheshadri, " On the classification of imbalanced data sets", International Journal of Computer Applications (0975 – 8887) Volume 44– No.8, April 2012

[12]. Maurice Samulski " Computer Aided Detection of Lesions in Digital Mammograms using Temporal Bayesian Classifiers", University of Medical Center Nijmegen, Department of Radiology, 2005.

[13]. A.Oliver, J.Freixenet, A.Bosch, D.Raba, "Automatic Classification of Breast Tissue", Institute of Informatics and Applications, University of Gerona, Spain, 2005.

[14]. W.Sheta, N.Eltonsy, G.Tourassi, A.Elmagbraby, "Automated Detection of Breast Cancer from Screening Mammograms using Genetic Programming", IJICIS vol.5, No.1, July 2005.

[15].E. Alberdi, P.Taylor, R.Lee, "Elication and Representation of Expert Knowledge for Computer Aided Diagnosis in Mammography", Methods Inf Med 2004, pp. 239-46.

[16].Monika Shinde " Computer Aided Diagnosis in Digital Mammography:- Classification of Mass and Normal Tissue", Department of Computer Science and Engineering, University of South Florida, 2003.

[17]. Renato Campanini, Danilo Dongiovanni, Allessandro Riccardi, "A Novel Featureless Approach to Mass Detections in Digital Mammograms based on Support Vector Machines", Department of ZPhysics, University of Bologna, Italy, 2003.

[18]. Lari Inann Bruce and Roza R Adhani "Classifying Mammogram Mass Shapes using the Wavelet Transform Modulus-Maxima Method", IEEE Transactions on medical Imaging, pp 1170-1177, v.18, n.12, 1999.

[19].Strausz Gy, G. Horvath, B. Pataki, " Intelligent Solution for Mammography Image Diagnosis", Budapset University of Technology and Economics, 2002.

[20].H.S.Sheshadri, A.Kandaswamy, "Computer Aided Decision System for early Detection of Breast Cancer", Indian Jmed, res 124, Aug 2006, 149-154.

[21]. H.D Cheng, Jingly Wang, "Microcalcification Detection using Fuzzy Logic and Scale Space Approaches", Pattern Recognition, pp.363-375, v.37, 2004.

[22]. Shun Leungng and Walter F. Bischof, "Automated Detection and Classification of Breast Tumors", Computer and Biomedical Research, pp:218-337, v.25, 1992.

[23]. Andreadis, G. Spyrou, A. Antaraki, G.Zografos, D. Kouloheri, G. Giannakopoulou, K. S. Nikita & P. A. Ligomenides, "Combining SVM and Rule Based Classifiers for Optimal Classification in Breast Cancer Diagnosis", Informatics Laboratory, Academy of Athens, Greece, 2006

[24] Arun kumar M.N and H.S. Sheshadri, ―Breast contour extraction and pectoral muscle segmentation in digital mammograms‖, International Journal of Computer Science and Information Security, Vol 9, No.2, February 2011.