CLASSIFICATION OF MAMMOGRAMS ON ZONE BASED STATISTICAL FEATURES

[1] Asmat Ara Ainapur, [2]D.A.Kulkarni, [3]G.R.Udupi

[1,2,3]KLS Gogte Institute of Technology, Belgaum ,Karnataka, (India)

ABSTRACT

Mammogram breast cancer images have the ability to assist physicians in detecting breast cancer caused by cells abnormal growth. In this paper, we present a method for the classification of mammograms using zone based statistical features. In the first step mammograms are preprocessed to reveal the regions of interest, distinguish masses from background tissue through the segmentation. Zone wise features are extracted from segmented part of the image which is used to identify the commonly used mammography dataset as normal, benign, or malignant, using nearest neighbor classifier. The experimental results show that the method works efficiently with the accuracy of 88.2 %.

Keywords: Breast Cancer, Mammogram, Classification, Segmentation, Features

I INTRODUCTION

The design and development of mammogram classification methods has recently gained importance, which provides a support for physicians to decide what type of treatment they supposed to give to a patient. Breast cancer remains a leading cause of cancer deaths among women in many parts of the world. Early detection of breast cancer through periodic screening has noticeably improved the outcome of the disease. Cancer is an abnormal, continual multiplying of cells. Most types of cancer cells eventually form a lump or mass called a tumor, and are named after the part of the body where the tumor originates.

In this paper, we present a new method based on zone wise features for medical image classification. We extract features from images to train, which are then employed for classifier building and validation. Rigorous experimentation is performed, and we achieve classification accuracy on a previously studied mammogram dataset, demonstrating the efficacy of our technique.

The following sections of the paper are organized as follows. In Section II, we provide a brief introduction to related work. In Section III, we outline the methodology in detail; in Section IV, we present the results of our proposed methodology and final section conclusion.

II RELATED WORK

The classification of medical images is a difficult and often computationally overwhelming task. Digitized medical images contain labels, noise, and irregularities that must be minimized before computational methods can be used to analyze them.

There are large numbers of diagnostic methods currently available, among which mammography is the most reliable method, for detecting early breast cancer [1]. The analysis of mammograms by computer is roughly divided into three steps: Enhancement of pre-selected features, Localization of suspicious areas and Classification of these areas into benign or malignant tumor. A scheme [2] which preserves the breast area and eliminate the structural noises in the mammograms through preprocessing. Then, suspicious regions are selected from the breast area using the Sech template matching method and adaptive square regions of interest (ROIs) are segmented from the original mammogram corresponding to suspected regions.

A method of finding regions of interests that can be suspicious masses in mammograms by using the iterative thresholding algorithm has been proposed [3]. The contours of the regions of interests are extracted from the first mode obtained by applying the (BEMD) Bidimensional Empirical Mode Decomposition method. Finally, the masses are refined by the contours extracted.

A tumor detection scheme [4] which is an empirically optimized, rule-based algorithm driven by a series of observations made on a group of mammograms that contain malignant masses. Specifically, the detection scheme relies on a morphological model of breast cancer growth that is relevant to masses. Extensive evaluation of the detection scheme is performed on various groups of normal and abnormal mammograms. A new approach is proposed [5] for automated detection of tumors (of different kinds and scales) in mammograms based on some prognostic factors: mass size, mass shape, intensity variation around mass boundaries, and spread of primary shape. Classification is then performed using these two segments. The texture features from the original image to discriminate between micro calcification and the normal tissue in the breast is proposed [6]. An integrated method using marker controlled watershed segmentation algorithm and level set evolution without re-initialization for breast mass segmentation is proposed by Ming Zhang [7] to address the problem of segmentation. SAHEB BASHA [8] presented a novel approach to automatically detect the breast cancer mass in mammograms using morphological operators and fuzzy c – means clustering algorithm.

III PROPOSED METHODOLOGY

The proposed methodology consists of mammogram preprocessing, segmentation, features extraction, classifier training and classification. We have employed the breast mammogram database made available by the Mammography Image Analysis Society (MIAS) [9] and Digital Database for Screening Mammography (DDSM)[10] which are publically available.

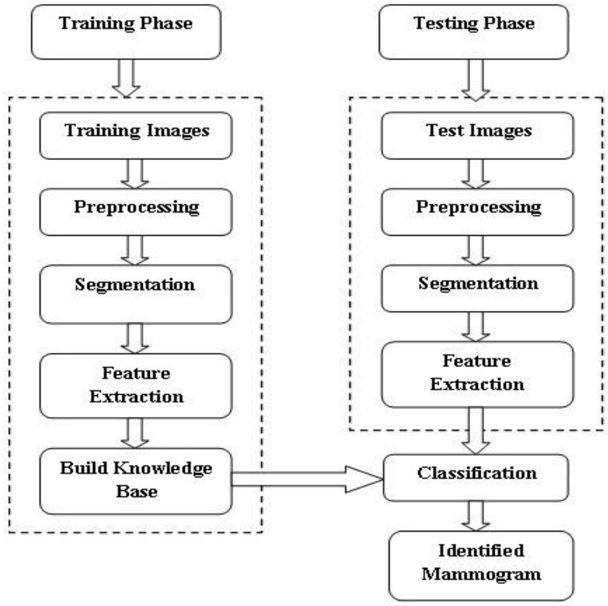


Fig 1. Proposed Methodology

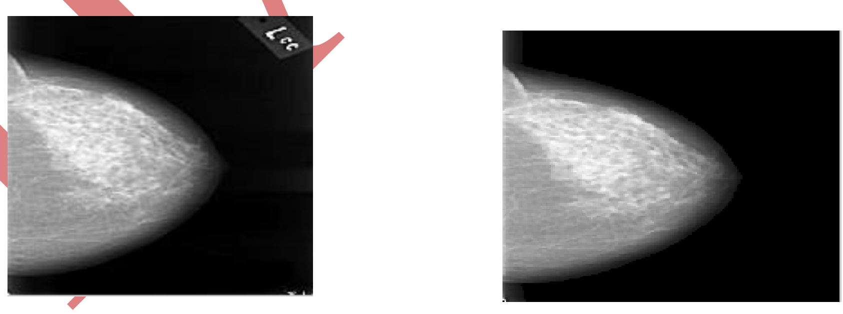
3.1 Preprocessing

During preprocessing stage, image is resized to 300X 300 pixels and median filtering is applied for enhancement of mammograms intensity and contrast manipulation, noise reduction, later a morphological operators are applied over binary image for making the foreground in the image which removes deliberately inserted identifiable labels. Dilation and erosion are defined by

Erosion: $E(I,SE) = I\Theta(-SE)$ ………(1)

Dilation: $D(I,SE) = I\Theta(SE)$ ………(2)

Figure 2 showed the preprocessing result.



(a) **(b)**

Fig2. (a) Original Image (b) Preprocessed Image

3.2 Segmentation

In the segmentation step, it is important to distinguish the suspicious region from its surroundings. The methods used to separate the region of interest from the background are usually referred as the segmentation process. In other words, the process of dividing an image into distinct, meaningful regions is called image segmentation. This is useful in the analysis of digital mammogram images because the image is divided into various components and areas of interest. The region is iteratively grown by comparing all unallocated neighboring pixels to the region. The difference between a pixel's intensity value and the region's mean, is used as a measure of similarity. The pixel with the smallest difference measured this way is allocated to the region. This process stops when the intensity difference between region mean and new pixel becomes larger than a certain threshold. In the proposed method threshold value is 15 and maximum distance we considered is 1500.

This function performs "region growing" in an image from a specified. Start with a single pixel (seed) and add new pixels slowly.

Step 1: Choose the seed pixel

Step 2: Check the neighboring pixels and add them to the region if they are similar to the seed

Step 3: Repeat step 2 for each of the newly added pixels; stop if no more pixels can be added.

(a) **(b)**

Fig3. (a) Original Image (b) Segmented Image

3.3 Feature Extraction

During the 3rd step that is feature extraction, where feature is used to denote a piece of information which is relevant for solving the computational task related to a certain application. A typical mammogram contains a great amount of heterogeneous information that depicts different tissues, vessels, ducts, chest skin, breast edge, the film, and the Xray machine characteristics. In order to build a robust CAD system that correctly classifies normal and abnormal regions of mammograms, we have to present all the available information that exists in mammograms to the diagnostic system so that it can easily discriminate between the normal and the abnormal tissue. However, the use of all the heterogeneous information, results to high dimensioned feature vectors that degrade the diagnostic accuracy of the utilized systems significantly as well as increase their computational complexity and calculation time. Therefore, reliable feature vectors should be considered that reduce the amount of irrelevant information thus producing robust Mammographic descriptors of compact size. In this phase, the preprocessed image of size 300x300 is divided into 36 zones each of size 50x50. Then the statistical features are computed from all zones and stored into a feature vector X. The sum of all pixels in every zone is used as a feature value. Totally 180 features are stored into feature vector X . viz, contrast, correlation, homogeneity, entropy and energy.

These extracted features are used to build the knowledge base.

During training, the features are extracted from all training samples and knowledge base is organized as a dataset of feature vectors. The stored information in the knowledge base sufficiently characterizes all variations in the input.

3.4 Classifier

In this phase, a test image is processed to obtain zone wise statistical features and stored into a feature vector. Then, the classifier determines nearness between the test sample and every record in the knowledge base using Euclidean distance measure as depicted below

$$\text{Distance(Test,KB)}=\sqrt{\sum (Test\ (f_i) - Train\ (X_j(f_i))^2}$$

Where,

$$1 \le i \le 36,\ 1 \le j \le N$$

The minimum distance between the test image and record in the knowledge base is used to recognize the character. The proposed methodology performs well for different image resolution. However, the method requires sufficient training of all variations in mammograms with different orientations.

IV RESULTS

The proposed methodology has produced good results for images of different size and orientation. The advantage lies in less computation involved in feature extraction and recognition phases of the method. During experiments it is noticed that, the zone wise features made samples separable in the feature space. Hence, the proposed work is robust and achieves an average recognition accuracy of 82%. The overall performance of the system after conducting the experimentation on the dataset is reported in Table I.

Test Case	Mammogram Type	Number of samples tested	Number of samples correctly classified	Number of samples incorrectly classified	Percentage of recognition accuracy	Overall Percentage
All samples are chosen from the trained set	Normal	54	54	0	100	98.7654
	Malignant	54	54	0	100	
	benign	54	52	2	96.2963	
Samples are chosen from trained set and untrained set	Normal	54	49	5	90.7407	92.5926
	Malignant	54	49	5	90.7407	
	benign	54	52	2	96.2963	
All samples are untrained	Normal	46	41	5	89.1304	88.9621
	Malignant	36	28	8	77.7778	
	benign	36	36	0	100.0000	

TABLE I. OVERALL SYSTEM PERFORMANCE

A closer examination of results revealed that misclassifications arise due to noise, more similarity between structural features and other degradations. It is also noticed that, zonal features takes care of variations in the orientation of the mammogram. It is also found that, if the knowledge base is trained for all variations and degradations, better performance can be obtained.

V. CONCLUSION

Mammogram classification is an overarching image classification problem. In this paper, we have presented. A method for the improvement of mammogram classification, which includes a preprocessing methodology for segmenting, Zone wise features for classification using nearest neighbor classifier . Our detailed experimental results demonstrate that our technique is valuable to existing techniques for mammogram classification.

REFERENCES

[1] Early Stage Detection Of Tumors In Mammogram S pawar B. V. And Patil Sushama Electronics Dept., W.C.E. Sangli _ 416415(Maharashtra), India.

[2] An Automatic Mass Detection System in Mammograms based on Complex Texture Features Shen-Chuan Tai, Zih-Siou Chen, and Wei-Ting Tsai 2013 IEEE.

[3] Mass Segmentation in Mammograms by Using Bidimensional Emperical Mode Decomposition BEMD Said Jai-Andaloussi, Abderrahim Sekkaki, Gw´enol´eQuellec, Mathieu Lamard, Guy Cazuguel, Christian Roux 35th Annual International Conference of the IEEE EMBS Osaka, Japan, 3 - 7 July, 2013

[5] Automated Detection of Tumors in Mammograms Using Two Segments for Classification. Mahmoud R. Hejazi and Yo-Sung HoY.-S. Ho and H.J. Kim (Eds.): PCM 2005, Part I, LNCS 3767, pp. 910 – 921, 2005. Springer-Verlag Berlin Heidelberg 2005

[6] Screening Mammogram Images for Abnormalities using Radial Basis Function Neural Network Dheeba.J Tamil Selvi.S 2010 IEEE

[7] An integrated method for breast mass segmentation in digitized mammograms Ming Zhang, Yingying Chai, Jianzhong Wang 2011 3rd International Conference on Advanced Computer Control (ICACC 2011)

[8] Automatic Detection Of Breast Cancer Mass In Mammograms Using Morphological Operators And Fuzzy C –Means Clustering .Saheb Basha, Dr.K.Satya Prasad Madina Engineering College, Kadapa, (A.P) - India, Jawaharlal Nehru Technological University, Kakinada, (A.P)- India.

[9]. MIAS Database. The PCCV Project: Benchmarking Vision Systems.http://peipa.essex.ac.uk/info/mias.htm.

[10]. Digital Database for Screening Mammography (DDSM) was developed through a grant from the DOD Breast Cancer Research Program, US Army Research and Material Command DAMD17-94-J-4015. http://marathon.csee.usf.edu/Mammography/Database.html