

BREAST CANCER CLASSIFICATION USING DATA MINING FEATURE EXTRACTION TECHNIQUE FOR MAMMOGRAM: A SURVEY

¹D.Janani, M.Phil scholar, Vels University, Chennai
²T.S.Poonguzhali, Asst.Professor, Vels University, Chennai
¹dhiyajanu@gmail.com, ²poonguzhali.scs.@velsuniv.ac.in

ABSTRACT

The visual representation is the most effective way to convey the information to each other. Digital Images are analyzed and executed by using automatic recognition systems to understand hidden information to be had in it. The mammogram scanning is the medical imaging technique for the breast examination. The mammogram examination is used to detect then analyzed early of breast cancer disease. The mammography segmentation is used in to detect the exact cancer on breast by applying proper preprocessing and segmentation technique. The mammogram imaging is the improved version of Computer Tomography (CT) that uses the features of X-ray system and utilizes the high resolution manner that can identify the cancer or tumors on the breast. The mammogram analysis can be studied by applying different algorithms to segment and classify the stages of the cancer image. The proper preprocessing technique has to be applied to restore and enhance the mammogram image matrix data to improve its quality. The preprocessing has two steps. In the first step, the matrix data has to be applied to de-noising technique to remove all the noise content occurred on the image matrix data due to the scanning device. In the second step, the image matrix data enhancement technique has to be applied to improve the image matrix data quality in terms of contrast and brightness to get high resolution output. Then image matrix data segmentation technique to be applied to detect the region. The ROI is the breast cancer

occurred on the mammogram image. The data are calculated from the ROI to estimate required data for data mining technique to classify the different stages of the cancer of the patient. Many research are proceeds on the mammogram image matrix data processing to provide the accurate and novel algorithms to improve the medical imaging field.

Keywords: mammogram, segmentation, image matrix data restoration, image matrix data enhancement, image matrix data segmentation, image matrix data classification, data mining, data extraction

1. INTRODUCTION

The imaging is processed to retrieve the facts without affecting the alternative capabilities which is also present within the imagery information. Processing of the imagery relies upon the utility domain and procedures to resolve it. The images are processed through primary steps both to suppress the unwanted information which include noise or improve the satisfactory of images for human visualization. A number of the application regions of image matrix data processing are scientific discipline, remote sensing, image matrix data recognition, video processing and microscopic image matrix data and so on.

Recently many preprocessing techniques for image matrix data filtering and enhancement are proposed by many scholars with appreciable novelty to give innovative solutions on mammography images.

The soft computing technique is used to segment the images that have cancer. The cancer region has to be properly segmented by applying the clustering technique to group the pixels exactly. Also some hybrid techniques and partial differential equations (PDE) are used to segment the image matrix data to detect the cancer. The segmentation is grouped into two parts. First part is clustering technique. Using the clustering technique the mammogram image matrix data has to be grouped into different pixel intensities. The threshold technique is the second part. The high intensity pixels are highlighted using threshold. There are many techniques are used for segmentation to extract the ROI from the mammogram images. Some of the methods Expectation Maximization (EM) algorithm, wavelet and fuzzy, wavelet and neural network, fuzzy and neural network, active contour model, genetic optimization technique, optimization techniques and some of the hybrid techniques are used to segment the cancer portion of the images.

E. Malar and A.Kandaswamy [1] are proposed following..

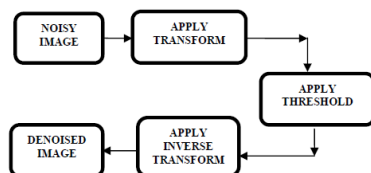


Figure A. Block diagram of de-noising

The above figure is the proposed system for mammogram image matrix data denoising. Denoising system accompanied.

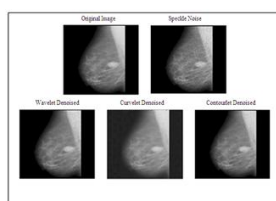


Figure B. Denoisy model

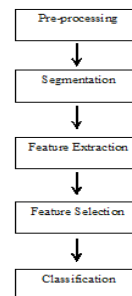


Figure 1: Image matrix data segmentation and classification

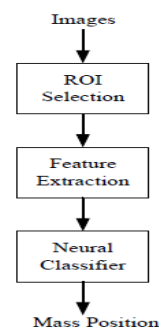


Figure 2: Mass extraction using neural network classifier

The figure 2 shows that the block diagram is proposed system by the authors.

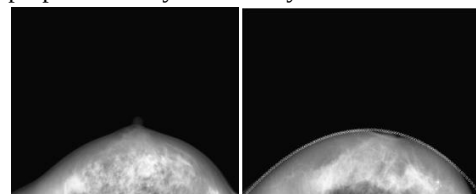


Figure 3: Input mammogram image, Figure 4: ROI extraction

2. ROI Mass segmentation on Matrix data of Mammogram

Zhen-zhong HAN and Hou-jin CHEN [4] are proposed a novel method of segmentation of

mass in mammogram images matrix data. The masses boundaries are transitional in nature of the human tissues. The mass frequently has an indistinct boundary. And there can be a few holes in the mass for the affection of noise. It is each hard to differentiate the mass via the conventional segmentation technique based on grey price. It is not possible to set the identical threshold for one of kind images. The area which belongs to the mass will not be divided into the tumor region, whilst its gray value isn't useful the threshold. And the form of the region could be very important for the following processing. Therefore, a completed mass segmentation is wanted. The PCNN-FCM (Pulse Coupled Neural Network-Fuzzy C Means) algorithm can be used to remedy the problem. Based on the time features generated with the aid of PCNN, the pixels are labeled by way of FCM clustering set of rules. A binary image matrix data is got after every PCNN iterative processing. Consequently, many binary images will be were given from the time perspective, but if examine each pixel alongside the time axis, $M * N$ ($M * N$ is the dimensions of the images) time collection could be located. It is easy to find that comparable pixel can have comparable series.

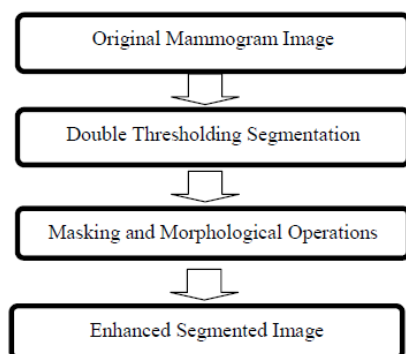


Figure 5: Flow chart for morphological operation based segmentation
The figure 5 shows that the flow chart for morphological operation based masses segmentation on mammogram imagery to detect

cancer. It uses the double threshold technique to convert the gray level image matrix data into binary data format as shown in figure below format.

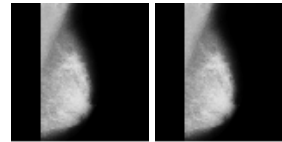


Figure 6.a. Input Image, Figure 6.b. Morphological Operation

3. Mass Lesions in Mammograms

Rabi Narayan Panda and BijayKetan [6] are proposed Feature Extraction for Classification of Micro-calcifications and Mass Lesions in Mammograms for breast cancer detection. Numerous traits that can characterize medical problems, which includes dissimilarities between imagery, and confluent densities related to fibrosis benign and calcifications etc are identified with the useful resource of Mammography.

4. Process of Calcification

Tiny put down of natural resources (calcium) that appear like contained medium depth spots within the mammogram are referred to as process of calcification. It is documented when micro-calcifications produce to be obvious from side to side mammography.

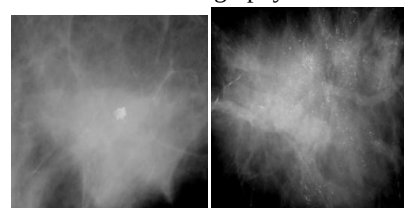


Fig.7. (a) coarse calcifications; (b) amicro calcification cluster.

Mammogram classification Data mining

Identity of the breast boundary may be very critical. By shape out the border of the breast, it's far likely to arrange of some object present out of doors the breast

vicinity that may have an effect on the performance of photograph evaluation and sample recognition techniques. Following notations are used to explain set of rules for skin line detection:

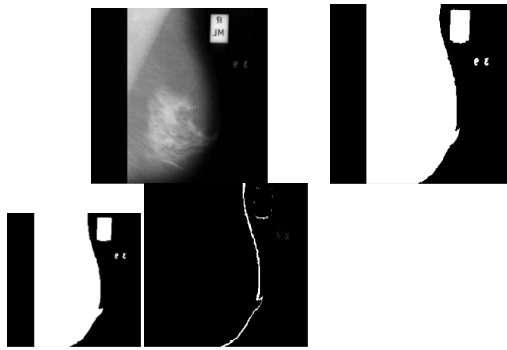


Fig. 8. Mammogram imagery outputs

Background removal manner can be defined for the absent lesion sight of mammogram procedure for the proper lesion view could be very similar so, it is not necessary to provide an explanation for both. To eliminate breast history, it's far necessary to create and observe a mask. Mask is created by means of making use of set of rules which is defined underneath to image with detected pores and skin line, proven in Fig. 2(d).

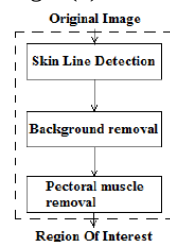


Figure 9. Flow chart of segmentation

1. Data extraction from mammogram matrix

Aswini Kumar Mohanti and Sukantha Kumar [8] are proposed Image Mining for Mammogram Classification by Association Rule Using Statistical and GLCM features for normal and abnormal breast cancer classification.

The mammogram image for this have a look at is taken from Mammography image analysis Society (MIAS) in England. Classification assessment determines. In this work the Contrast Limited Adaptive Histogram Equalization (CLAHE) changed into implemented to the matrix data Ranjit Biswas and Abhijit Nath [9] are proposed mammogram classification using gray-level co-occurrence matrix for diagnosis of breast cancer.

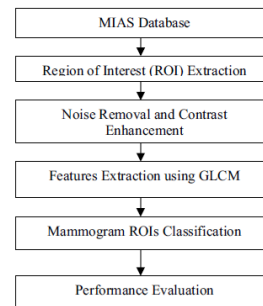


Figure 12. Proposed architecture

Mammogram image classification tool is proposed by the authors. The proposed system for class of mammogram image data is built primarily based on GLCM via making use of unique classifiers. The system is split into 5 tiers to categories mammogram image data. First step is the statistics set collection, 2nd is ROI extraction system, third is the pre-processing steps that's again divided into two steps filtering and enhancement, fourth is the feature extraction from GLCM and ultimate degree is class. Authentic mammogram image data have one-of-a-kind types of noises, artifacts of their heritage, pectoral muscle tissue and many others that are unwanted for feature extraction and classification. For this reason a cropping operation has been applied on mammogram image data to extract the ROIs which includes the abnormalities, other than the unwanted portions of the mammogram image data. MIAS

database offers all of the details about every mammogram imagery, viz., length in pixels, individual of history tissue, magnificence of abnormality, Xc and Yc coordinate fee of centre of abnormality, 'r' radius of circle enclosing the abnormality by the radiologists. ROIs extraction completed with the aid of guide cropping operation considering the centre of the unusual location as the centre of ROI and taking the approximate radius (in pixels) of a circle enclosing the unusual place as proven in Fig. 2. For the extraction of normal ROI, the identical cropping method has been completed on ordinary mammographic photographs with random choice of place. in this paintings, all the ROIs are resized in to 128 x 128 for uniformity.

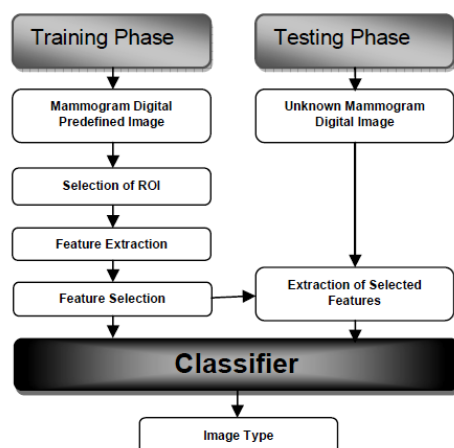


Figure 10: Classification of mammogram cancer type

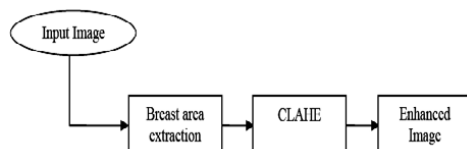


Figure 11: preprocessing block diagram

2. Histogram Equalization

In image processing, processing of massive records is time eating and much less classification. For reducing time, the enter records is transformed into decreased set of function vector. This variation method is called characteristic extraction method. This option vector contains relevant information and is used as input vector for classification. Capabilities may be labeled primarily based on color, texture and shape. In this proposed system, we are particularly involved about texture features and for extraction of functions; gray level Co-occurrence Matrix (GLCM) is applied because it has been demonstrated as an effective tool for data extraction. On this proposed system texture data features specifically contrast, correlation, Energy, homogeneity are estimated for data mining.

Some of the important data extraction for data mining is,

Correlation,

$$F2 = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{(ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

Energy,

$$F3 = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \{p(i,j)\}^2$$

Homogeneity,

$$F4 = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{p(i,j)}{1+(i-j)^2}$$

3. Neural Networks for data mining

Xiaofei Zhang and Yi Zhang [10] are proposed mammogram image classification with convolutional neural networks. Deep learning to know with Convolutional Neural Networks has emerged as one of the most powerful tool-gaining knowledge of medical image classification, the accuracy of virtually all different conventional classification strategies or even medical imaging. The convolutional system can simplify a mammogram image containing millions of pixels to a set of small

characteristic maps, thereby decreasing the measurement of input information at the same time as maintaining the maximum-critical differential features. The utility of CNNs to classify mammograms isn't entirely new. But, most of the work awareness at the classification of small patches, known as Region of interest (ROI). An ROI is the region that is possibly to comprise a tumor. That is commonly carved out of the complete image based on either clinical data or automatic segmentation. Their great effects have been 0.74 for complete image, 0.8 for whole image plus automatically detected small lesion patches, and 0.91 for complete image plus manually segmented small patches in the ROI.

High quality Mammogram facts from the university of Kentucky medical University had been acquired with institutional assessment board approval (IRB 17-0011-P3K). All mammography image data have been assessed via experienced radiologists. The dataset includes three,018 bad and 272 nice mammogram images data. Each of the high-quality mammogram data carries at the least one biopsy-verified malignant tumor. The bad image do no longer include malignant tumors showed with at least 2- 12 months poor mammogram observe-up assessed through radiologists, but may additionally have benign loads approved through biopsy or hooked up more than 2 -yr imaging balance. All tests within the dataset had been taken in either CC or MLO view or each. Tremendous samples originated from one hundred twenty five sufferers.

4. CONCLUSION

The mammogram images are used to detect the breast cancer early. The computer data mining technology is useful to provide the proper segmentation and classification. To provide the proper segmentation and classification the image has to be subjected to the preprocessing. The

preprocessing stage will give the filtered and enhanced mammogram matrix data that is useful for extracting the cancer region from the breast. The data has to be extracted from the ROI of the breast cancer image that is given to the classification using data mining technique of neural network.

5. REFERENCES

- [1]. A comparative study on mammographic imagedenoising technique using wavelet, curvelet and contourlet transforms, E. Malar, A.Kandaswamy, S.S.Kirthana, D.Nivedhitha, Department of Biomedical Engineering, PSG College of Technology, Coimbatore, India, 978-1-4673-2322-2/12/\$31.00 ©2012 IEEE
- [2]. A Novel Method of Mass Segmentation in Mammogram Zhen-zhong HAN, Hou-jin CHEN, Ju-peng LI, Chang YAOSchool of Electronics and Information Engineering Beijing Jiaotong University Beijing 100044, China, 978-1-4673-0199-2/12/\$31.00 ©2012 IEEE
- [3]. Detecting Masses in Digital Mammograms Based on Texture Analysis and Neural Classifier Guo-Shiang Lin¹, Yu-Cheng Chang¹, Wei-Cheng Yeh², Kai-Che Liu³, Chia-Hung Yeh⁴
¹Dept. of Computer Science and Information Engineering, Da-Yeh University, Taiwan 978-1-4673-2588-2/12/\$31.00 ©2012 IEEE
- [4]. Mammogram enhancement and segmentation methods: classification, analysis, and evaluation mariam biltawi¹, nijad al-najdawi², sara tedmori³, Department of Computer Science, The King Hussein School for Information Technology, Princess Sumaya University for Technology.
- [5]. Breast Cancer Detection with Mammogram Segmentation: A Qualitative Study Samir M. Badawy¹ Industrial Electronics and Control Eng. Dept, Faculty of Electronic Engineering, Menoufia University, Egypt, (IJACSA)

International Journal of Advanced Computer Science and Applications.sss

[6]Dharmarajan, K., and M. A. Dorairangaswamy. "Discovering Student E-Learning Preferred Navigation Paths Using Selection Page and Time Preference Algorithm." International Journal of Emerging Technologies in Learning (iJET) 12.10 (2017): 202-211.

