Computer-Aided Diagnostic System for Digital Mammography

Yang Xingyao[*], Muhammad Fayaz[$], Tanmay Ghosh[#]

[*]School of Information Science and Engineering, Xinjiang University Urumqi, Xinjiang,
China

[$]University of Engineering and Technology, Peshawar, Pakistan

[#]Department of Microbiology, Dinabandhu Andrews college, Baishnabghata, South 24 Parganas
West Bengal

Corresponding author
Khan Alamgir
School of Information Science and Engineering, Xinjiang University Urumqi, Xinjiang, China
EMAIL: JOHNALAM4@GMAIL.COM

1 - Abstract

In this work, Computer-Aided Detection (CADe) and Computer-Aided Diagnosis (CADx) systems are developed and tested using the public and freely available mammographic databases named MIAS and DDSM databases, respectively. CADe system is used to differentiate between normal and abnormal tissues, and it assists radiologists to avoid missing a breast abnormality. At the same time, CADx is developed to distinguish between normal, benign and malignant breast tissues, and it helps radiologists to decide whether a biopsy is needed when reading a diagnostic mammogram or not. Any CAD system is constituted of typical stages including preprocessing and segmentation of mammogram images, extraction of regions of interest (ROI), features removal, features selection and classification. In both proposed CAD systems, ROIs are selected using a window size of 32×32 pixels, then a total of 543 features from four different feature categories are extracted from each ROI and then normalized. After that, the selection of the most relevant features is performed using four different selection methods from MATLAB Pattern Recognition Toolbox v.5 (PRtool5) named Sequential Backward Selection (SBS), Sequential Forward Selection (SFS), Sequential Floating Forward Selection (SFFS) and Branch and Bound Selection (BBS) methods. We also utilized Principal Component Analysis (PCA) as the fifth method to reduce the dimensions of the features set. After that, we used different classifiers such as Support Vector Machines (SVM), K-voting Nearest Neighbor (K-NN), Quadratic Discriminant Analysis (QDA) and Artificial Neural Networks (ANN) for the classification. Both CAD systems have the same implementation stages but different output. CADe systems are designed to detect breast abnormalities while CADx system indicates the likelihood of malignancy of lesions. Finally, we independently compared the performance of all classifiers with each selection method in both modes. The evaluation of the proposed CAD systems is done using performance indices such as sensitivity, specificity, the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curves, the overall accuracy and Cohen-k factor. Both CAD systems provided encouraging results. These

results were different corresponding to the selection method and classifier.

2 - Key words

Digital Mammography; Breast Cancer; Computer-Aided Diagnosis; Machine Learning; Medical Image Processing.

3 - Introduction

Breast cancer is the abnormal growth of breast cells, which usually starts in the lobules of the inner lining of the milk ducts. Different types of breast cancer exist with different stages, truculence and genetic makeup. 10-years disease-free survival rates vary from 98% to 10% with the best treatment plan. Breast cancer treatment includes various strategies such as surgery, chemotherapy, hormone therapy and radiation. Breast cancer is still being a significant public healthcare problem among women, and it is the most common cancer around the world. The cause of this disease remains undetermined, and this makes primary prevention to be impossible. It is believed that early detection of breast cancer is the most promising way to lower the number of women suffering from it and this improves the chances to provide proper treatment options so that treatment will work with better results [2]. Among women in U.S., breast cancer is the most commonly diagnosed cancer besides skin cancer, and its death rates are the highest among other cancers besides lung cancer. In 2016, it is expected to diagnose approximately 246,660 new cases of invasive breast cancer in women, and the expected diagnosed cases of non-invasive breast cancer are 61,000 women

in the U.S. Also About 2,600 new patients of invasive breast cancer in men are expected to be discovered [3]. Mammography is a particular type of radiography, using low radiation levels to acquire images for a breast to diagnose a consequent exist of abnormal structures that imply a disease like cancer. The early detection of strange mammary pathologies like nonpalpable breast masses and calcifications is extremely important for successful cure of breast cancer patients. Mammography is the standard screening tool that is used to perform the task of breast cancer detection, and it results to reduce at least 30% in breast cancer deaths in the world [4], [5]. The benefit of mammography screening has undergone some recent contention since definite evidence relating mammography with mortality may not be proven. In contrast, an Institute of Medicine Report on Mammography suggests that using mammography screening for earlier detection of breast cancer may be an important factor in decreasing mortality from breast cancer in recent years [6]. The computer-based systems may provide a second supportive alternative in detecting breast abnormalities by completing the expert knowledge of radiologists, and two may help to make concise diagnostic decisions. Computer-aided diagnosis (CAD) is a diagnostic procedure performed by the help of computer algorithms. Computerized systems for such a purpose are called CAD systems. It has been proved that the radiologist's performance in diagnosing mammography images can be increasingly improved by assisting them with the results obtained by the CAD system. Therefore, motivations become keen to develop robust CAD systems to support

radiologists' ability in reading different mammograms accurately [7].

4 - Important contents

Recently, many CADe systems have been developed by research groups for detection and classification of breast abnormalities. In many systems, some typical stages should be performed to find suspicious breast lesions. The main steps of such a method are shown in figure 3.1. The implementation of the proposed CADe system is accomplished using digital mammograms from MIAS database. The preprocessing is the first stage of the CADe system by which the breast region is enhanced and segmented by applying image processing techniques to improve the quality and reduce the noise ratio of the mammogram. The next stage is the selection of a region of interest (ROI), where a group of suspicious ROIs are selected to further delineate or classify them as normal or abnormal lesions. The following stage is the feature extraction that aims to characterize the lesions to differentiate actual lesions from falsely detected candidates, where the selected ROI is used to calculate a set of features. After feature extraction comes to the feature selection step which is considered an essential part of any classification scheme where the success of classification task largely depends on the selected features and the extent of their role in the model. The final stage is the classification where the selected features are then inputted into a classifier. The classifier is trained to distinguish normal from abnormal lesions. The design and experimental verification of the proposed study are achieved through two main phases. The first phase is the training phase, while the other is the testing phase, with 50% of the database. In the training phase, the system is trained to differentiate between healthy and cancerous cases by inputting the selected features of predefined normal and malignant images to the classifier. Then, in the testing phase, we test the performance of the system by entering elements of a test image to evaluate the correctness degree of the system decision.
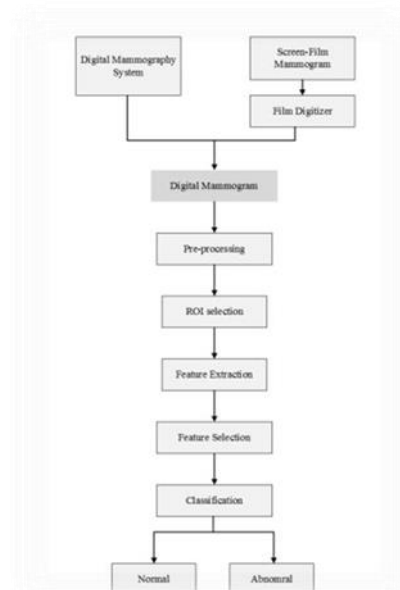

Figure 3.1: Block diagram of a generic CADe system

In future work, It is imperative to continue the development of CAD systems that can contribute in the study of the breast cancer, innovating techniques capable of assisting the radiology experts and helping to reduce their subjectivity and examination time. The proposed systems are suitable to continue working on, by using datasets with a large number of samples representing different cases, including various classes of breast tissue abnormalities especially micro calcifications, extracting new powerful textural features and using more robust

classification algorithms utilizing hybrid classifiers.

We can additionally use Enhancement of Mammography Images Using Peripheral Region Equalization

5 - Conclusions

Early detection of breast cancer is of great importance to increase the survival rate and improve the chances to provide proper treatment options so that treatment will work with better results. Mammography is the gold standard tool for the early detection of breast cancer. Still, the sensitivity of mammography is usually affected by the image quality and the radiologist's level of expertise. CAD systems have been developed to support radiologists' decision and thus decreasing the false positive rate. In this study, we proposed computer-aided detection (CADe) and computer-aided diagnosis (CADx) systems. Both systems have the same implementation techniques, but each system is designed to perform a different task. CADe system is used to detect the abnormalities in breast lesions while CADx system is intended to diagnose and determine the malignancy of the suspicious breast tissues. MIAS database was used to develop the CADe system, whereas the CADx system was designed using the DDSM database. The same combination of different features extracted from each ROI was used in both CAD systems. Also, both CADe and CADx systems have the same feature selection methods and classifiers.

In almost all mammography systems, during the acquisition of a mammogram, the breast is compressed between the compression paddle and the support table then taking an image of the compressed breast tissue. Due to the forces that are applied on the upper surface of the breast by the upper plate, the deformation of the breast will happen. When compression is used, the top plate is tilted, which results in variation in breast thickness up to 2 cm from the chest wall to the breast margin. Variation in breast thickness affects image analysis by its impact on the grey level values of the image at the peripheral area of the breast, which causes changes in contrast at the breast periphery [20]. The outer space of the chest always tends to have lower intensity than the central area of the breast. So, a radiologist typically must adjust a window level setting when reading different areas of the image to assess a copy of the entire breast. This process may make only portions of the image comparable at any one window level setting and increases the time it takes to read the image especial with a massive number of patients. Peripheral equalization (P.E.) method is a dedicated image processing technique improved for mammogram enhancement. It is used to enhance the visibility of the outer area of the breast to make image features visible in both central and peripheral regions of the chest with one window level setting. The technique is also referred to as peripheral enhancement or thickness correction. Fatty tissues in the interior and outer regions of the breast have similar grey level values after equalization [19]. Figure A.1 shows an example of the process of peripheral equalization.
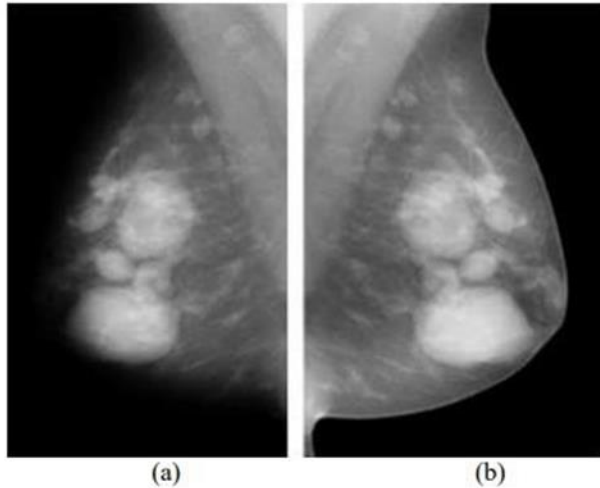
Figure A.1: Example of a corrected mammogram . (a): an MLO image. (b): The thickness corrected vision.

The peripheral enhancement technique proposed by Tao Wu et al. [8] is used as a preprocessing stage in our CAD systems. The general concept described here is to estimate the normalized thickness profile (NTP) of a breast from a mammogram image and enhance the peripheral area. The algorithm is described in detail as follows: The first step is the segmentation of a projection mammogram where segment the breast muscle or breast region from the background using adaptive threshold value computed using the Otsu thresholding. Separating a mammogram into breast and background regions can be done by Otsu' threshold value. A segmentation image (S.I.) was generated in which pixels were assigned a first value (e.g. amount of one) in a breast region and a second value (e.g. the value of zero) in the background region. Figure A.2 shows the segmentation step by Otsu thresholding for MIAS data set.
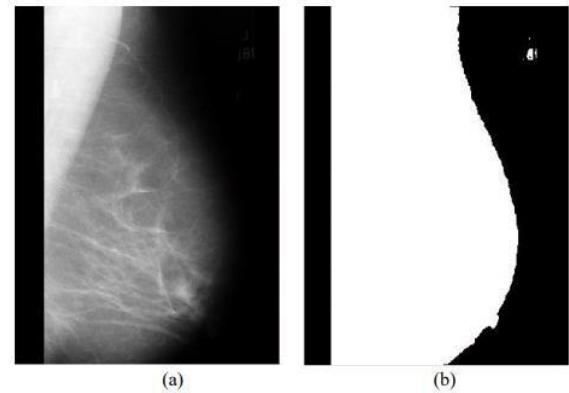

Figure A.2: Segmentation step of a mammogram (a): The original mammogram. (b): The mammogram image segmented by Otsu thresholding.

The second step is a blurred image (B.I.) generation. In this step, a two dimensional (2D) lowpass filter (LPF) was applied to the original image in the frequency domain to obtain a blurred vision, which primarily reflected variations in breast thickness. The Gaussian low-pass filter GLPF with cutoff frequency equal to 20 was used in this study. The filter function of GLPF in the frequency domain is:

(A.1)

$$H(u, v) = e^{-\frac{D(u,v)}{2D_0^2}},$$

1.   Multiply the input image f (x,y) by (−1) x+y to centre the transform.

2.   Compute F(u, v), the DFT of the image from step (1).

3.   Multiply F(u, v) by a filter function H (u, v).

4.   Compute the inverse DFT of the result in (3).

5.   Obtain the real part of the result in (4).

6.   Multiply the result in (5) by$(−1)$ x+y. A perspective plot, image display of a Gaussian LPF function is shown in figure A.3.
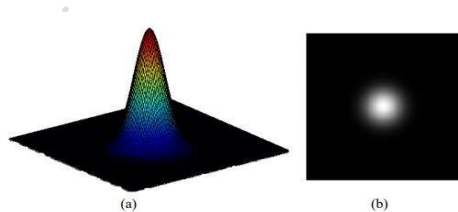

Figure A.3: Gaussian low-pass filter GLPF. (a): Perspective plot of a GLPF transfer function (3D). (b): Filter displayed as an image (2D).

6 - References

1. G. A. Sulik, Pink Ribbon Blues: How Breast Cancer Culture Undermines Women's Health. Oxford University Press, 2011.

2. H. D. Cheng and M. Cui, "Mass lesion detection with a fuzzy neural network," in 2003
   IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03), vol. 2, no. 6, p. II-849-52

3. American Cancer Society. Cancer Facts & Figures 2016. Atlanta: American Cancer Society; 2016.

4. M. P. Sampat, M. K. Markey, and A. C. Bovik, "Computer-Aided Detection and Diagnosis in Mammography," in Handbook of Image and Video Processing, Second Edi., vol. 10, no. 4, Elsevier, 2005, pp. 1195–1217.

5. M. Akay, Ed., Wiley Encyclopedia of Biomedical Engineering. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006.

6. R. A. Castellino, "Computer aided detection (CAD): an overview," Cancer Imaging, vol. 5, no. 1, pp. 17–19, Aug. 2005.

7. S. Yu and L. Guan, "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films.," IEEE Trans. Med. Imaging, vol. 19, no. 2, pp. 115–26, Feb. 2000.

8. T. Wu, R. H. Moore, and D. B. Kopans, "Multi-threshold peripheral equalization method and apparatus for digital mammography and breast tomosynthesis." Google Patents, 27Jul-2010.

9. R. Sivaramakrishna and R. Gordon, "Detection of breast cancer at a smaller size can reduce the likelihood of metastatic spread: A quantitative analysis," Acad. Radiol., vol. 4, no. 1, pp. 8–12, Jan. 1997.

10. S. V. Rajkumar and L. C. Hartmann, "Screening Mammography in Women Aged 40- 49 Years," Medicine (Baltimore)., vol. 78, no. 6, pp. 410–416, Nov. 1999

11. . S. A. Feig and R. E. Hendrick, "Radiation risk from screening mammography of women aged 40-49 years.," J. Natl. Cancer Inst. Monogr., no. 22, pp. 119–24, 1997.

12. . R. a. Smith, A. C. von Eschenbach, R. Wender, B. Levin, T. Byers, D. Rothenberger, D. Brooks, W. Creasman, C. Cohen, C. Runowicz, D. Saslow, V. Cokkinides, and H. Eyre, "American Cancer Society Guidelines for the Early Detection of Cancer: Update of Early Detection Guidelines for Prostate, Colorectal, and Endometrial Cancers: ALSO: Update 2001-- Testing for Early Lung Cancer Detection," CA. Cancer J. Clin., vol. 51, no. 1, pp. 38–75, Jan. 2001.

13. M. Elmanna and Y. Kadah, Computer Aided Diagnosis System for Digital Mammography. Germany: Scholars' Press, 2015.

14. American College of Obstetricians & Gynecologists, "Breast Cancer Screening. Practice Bulletin No. 122," Obs. Gynecol, vol. 118, pp. 372–82, 2011.

15. A. Qaseem, V. Snow, K. Sherif, M. Aronson, K. B. Weiss, and D. K. Owens, "Screening mammography for women 40 to 49 years of age: a clinical practice guideline from the American College of Physicians.," Ann. Intern. Med., vol. 146, no. 7, pp. 511–5, Apr. 2007.

16. "Recommendations on screening for breast cancer in average-risk women aged 40- 74 years," Can. Med. Assoc. J., vol. 183, no. 17, pp. 1991–2001, Nov. 2011.

17. "Mammomat Inspiration with PRIME Technology The reference in low-dose mammography." Siemens Healthcare GmbH, Germany, 2016.

18. R. Perez, Design of Medical Electronic Devices. Academic Press, 2002.

19. P. R. Snoeren and N. Karssemeijer, "Thickness Correction of Mammographic Images by Means of a Global Parameter Model of the Compressed Breast," IEEE Trans. Med. Imaging, vol. 23, no. 7, pp. 799–806, Jul. 2004.

20. X. H. Wang, W. F. Good, B. E. Chapman, Y.-H. Chang, W. R. Poller, T. S. Chang, and L. A. Hardesty, "Automated Assessment of the Composition of Breast Tissue Revealed on Tissue-Thickness-Corrected Mammography," Am. J. Roentgenol., vol. 180, no. 1, pp. 257–262, Jan. 2003.

21. D. Lavanya et al. / Indian Journal of Computer Science and Engineering (IJCSE) Analysis of feature selection with classification: Breast Cancer Datasets.