STUDY ON MAMMOGRAPHIC USAGE OF IMAGE PROCESSING TECHNIQUES

Wiraj Udara Wickramaarachchi[1] ,R.K.A.R.Kariapper[2] , S.L. Abdul Haleem[3] and S.L. Roshan Rukshana[4]

1

[1] *School of Computer Science and Technology, Wuhan University of Technology, P.R. China*
[2] *Department* of *Computing and Information Systems, Sabaragamuwa University of Sri Lanka*, *Sri Lanka*
[3] *Department* of *Information & Communication Technology, South* Eastern *University of Sri Lanka, Sri Lanka*
[4]*Faculty of Arts & Culture, South Eastern University of Sri Lanka*, *Sri Lanka*

wirajudara@gmail.com[1] , *rk@appsc.sab.ac.lk ahaleemsl@seu.ac.lk[3]* ,
roshanslr93@gmail.com[4]

ABSTRACT: *Cancer is a division of abnormal cells without control and it is aggressing to other tissues as well as spreading to other parts of the body through the blood and lymph systems. It could not be consider as just one disease but many diseases. There are numerous different types of cancer. Breast Cancers is one of the most heard diseases in present society. It has become a major public health issue which has spread out through the whole world. There are several methods to diagnose breast cancer using breast tissues. One of the most recognized method is mammography. The study reveals an insight of selected available breast cancer detection techniques based on mammographic images, and their highlighted characteristics. Further, the details and pros and cons of the techniques are also discussed.*

Keywords: Mammography, Image Processing, Breast Cancer, Preprocessing, Tumor Identification

1. INTRODUCTION

According to the global census reports which were published in recent times, around 1.15 million of new cases for breast cancer have been recorded from all over the world and it has caused more than 500,000 numbers of deaths (Parkin, Bray, Ferlay, & Pisani, 2002) (Hortobagyi, et al., 2005). Almost half of these have been recorded from industrialized nations (Parkin, Bray, Ferlay, & Pisani, 2002) (Parkin & Fernández, 2006) and more developed regions of the world. They are most likely attributed to the availability of screening programs used to detect breast cancer, which may otherwise have never been diagnosed (Vainio, Bianchini, & Heseltine, 2002).

The percentage of life up to five years of an infected patient could be increased by the detection of breast cancer at the earliest stage.. This cancer has four stages. In the first stage the size of mass varies between 0-2cm and diagnosis at this stage improves the life quality and 5 year survival rate of 96% (Radiological Society of North America (RSNA), 2016). At the last or fourth stage, there may be a need of partial or total removal of breast. At this stage 5 year survival rate drops to 20%.The total overall cost for the treatment of patients with breast cancer increases with higher stages of diseases (Groot, Baltussen, Groot, Anderson, &

Hortobágyi, 2006).Therefore, screening and diagnosis of breast cancer at earlier stages is beneficial to the patient and minimizes the financial burden (Groot, Baltussen, Groot, Anderson, & Hortobágyi, 2006).

Medical Image Processing is the recent improvement in the medical field and more specifically with the involvement of information technology in the medical field. Medical Image Processing is not limited to diagnose cancer disease, but also for the diagnosis of different kinds of diseases and it is clear through statistics.

It is easier to detect tumor from an infected breast and diagnose breast cancer with the help of image processing techniques. The risk of the most unnecessary outcome of this disease (death) can be minimized with the help of a proper diagnosis and treatment (Yasmin, Sharif, & Mohsin, 2013). This review is an attempt to provide an insight of the available different breast cancer detection techniques and their numerous impacting factors which uses Mammographic images. Also this provides the performance and accuracy of the discussed techniques. Each discussed technique is unique in its nature and targets a special kind of scenario. The details and pros and cons of methodologies are discussed.

2. MAMMOGRAPHY

First paragraph should start here. (Sentence Case, size 12, Justified, 1.15 Spacing)

- ***Computer Aided Detection (CAD) systems***

This method uses a digitized mammographic image that can be obtained from either a conventional film mammogram or a digitally acquired mammogram. The computer software then searches for abnormal areas of density, mass, or calcification that may indicate the presence of cancer. The CAD system highlights these areas on the images, alerting the radiologist to the need for further analysis.

- ***Breast Tomosynthesis:***

Also called three-dimensional (3-D) breast imaging and is a mammography system where the x-ray tube moves in an arc over the breast during the exposure. It creates a series of thin slices through the breast that allow for improved detection of cancer and fewer patients recalled for additional imaging (Radiological Society of North America (RSNA), 2016).

The discussion of the reviewed techniques can be stated as below. These techniques basically uses above concepts.

2.1   Identify Abnormal Mass Using Digital Mammography

Mammography cannot detect every kind of breast cancer but still it is world widely used process for breast cancer detection due to its inexpensiveness and low complexity. Mammography detects around 80% to 90% of breast cancers (Highnam & Brady, 1999). The breast area is extracted as an image and processed before printing on the film for better visualization of size, location and angle of the mass. These optimized images are then observed by radiologist for detection of possible abnormalities. These observations can be specific and may vary from one patient to another.

The survey identified two major stages in the process of most the Digital Mammographic Algorithms. They are:

- Preprocessing
- Post processing

The mammogram images including all type of breast tissues used by this algorithm (Kekre, Sarode, & Gharge, 2009) are taken from mini MIAS database (Suckling, et al., 1994). These images are in the map format of 8-bit having 256 grey levels. The algorithm will identify abnormal masses to ease further investigation. Also it focused on calcification, circumscribed, speculated and other ill-defined masses.

Preprocessing: Mammogram images are difficult to use as it is. Therefore they are processed before used for segmentation. This process includes removal of unwanted or irrelevant areas by setting a certain threshold value and to make highlight the area of interest by increasing the contrast.

Post processing: In this stage the preprocessed mammogram image is divided into 2x2 small blocks of pixels. Then all pixels values of the block are scanned and the value having maximum occurrence within the block is assigned to all pixels of that block. Recursively this is done by combining these 2x2 homogenous blocks in to 4x4 pixel blocks and then constructed 4x4 block pixels into 8x8. This process highlights the abnormal mass areas (Ball & Bruce, 2007). After highlighting the area, segmentation of breast is performed through color quantization technique (Kekre, Sarode, & Gharge, 2009). Through this technique the 8x8 homogenous mammogram image can be segmented to different color regions with each region representing specific part and properties.

The enhanced mammogram image are analyzed by constructing the histograms in MATLAB where a comparison is made between a normal and a diseased breast and then calculating ratio for finding

area of the region or the set of pixels were the abnormality lies (Khuzi, Besar, Zaki, & Ahmad, 2009).

There a lot of researches (Alhadidi, Zubi, & Suleiman, 2007) (Mohanalin, Kalra, & Kumar, 2008) (Wang, Qin, Fisher, & Zhao, 2006) have been done for the early diagnosis of breast micro calcifications using the digitalized mammograms with the application of digital image processing techniques and different preprocessing algorithms. All these papers follow almost the same steps with some difference in techniques. The common steps are preprocessing, segmentation of breast image, feature extraction, tumor segmentation and extracted feature classification.

2.2 Segmentation of Breast Cancer Mass using Screening Mammography

Screening mammography is the imaging technique used here (Yao, 2004) (Zadeh, Janianpour, & Haddadnia, 2013). A computer Aided Detection (CAD) system and screening mammography are combined in order to obtain accurate and efficient segmentation results. Image enhancement is done by applying Gaussian and Convolution operations (Zadeh, Janianpour, & Haddadnia, 2013) as spatial and frequency domain filters and low pass filter in order to smooth the edges of the image. Laplacian and gradient filters used as high pass filters to make the details sharp and enhanced (Gonzalez & Wintz, 2010).

Fourier transformation is done for frequency domain filtering as the first step. Then transformed image gets multiplied with filter mask. The resulted image shows the foremost abnormal area. Fractal analysis is performed after high pass filtering to remove the normal tissues in order to make segmentation step efficient.

The morphological algorithm used to extracts the boundaries in order to do segmentation. This algorithm also consists of two major steps same as digital mammography. They are:

- Preprocessing
- Segmentation

A method of thresholding is used to convert an image in to binary image. The tumor tissues have greater intensity may be represented by 1. After this process two morphological operations of dilation and erosion are applied to highlight tumor pixels or white pixels. Then tumor area is extracted from binary image by performing morphology boundary extraction. Through this way tumor is extracted from the mammogram images.

2.3   Image processing techniques with MRF and HPACO algorithms

A Survey (Yasmin, Sharif, & Mohsin , 2013) on this topic shows that these techniques also follow above procedure. The considered algorithm is used to remove the noise from the digital mammographic image (Sultana, Ciuc, Strungaru, & Florea, 2010). Then a median filter is added to the resulted image inorder to remove the noise and high frequency components without disturbing the edges. The infected region or region with micro calcifications is segmented using threshold value which derived from combination of Markov Random Field (MRF) and HPACO algorithm (Thangavel, Karnan, Siva Kumar, & and KajaMohideen, 2005).

Dependency Matrix (SRDM)and Grey Level Difference Matrix (GLDM) are used to classify extracted features of segments. Receiver operating characteristics (ROC) is evaluated for classification performances (Irwin, Downey, Gardi, & Fenster, 2008). The results of the research have shown that the detection rate ofthis algorithm is 94.8%.

2.4   Tumor Identification through Image Segmentation and 3d Structure Analysis

A survey paper (Yasmin, Sharif, & Mohsin , 2013) shows that we can capture the image for cancer diagnosis from many sources like mammography and electromagnetic imaging. For breast examination, the technique that uses high contrast X-ray system is known as a high resolution mammography. This imaging technique gives a bit inaccurate result in detection and determining the actual stage of tumor, in resulting a need to add accuracy if this type of imaging technique is used.

Electromagnetic imaging uses microwave technology and the concept of back scattering from the water content present in the tissues. This technique is further divided into three categories. They are:

- Microwave hybrid approach
- Microwave tomography
- Ultra wide band radar technique

But the paper only highlights the imaging techniques being used for early stage breast tumors detection as well as lungs tumors. The previous CAD system that identifies the early stage breast and lungs tumor with the help of image processing and artificial intelligence gives 85%-93.1% accurate results. However 100% accuracy for both breast and lungs cancer could identify the early stage tumor with the enhanced CAD system. This proposed system provides the exact size of tumor with the help of image processing and 3D structure analysis.

The detection technique needs to be dynamic because of different nature of image problems at the time of detection (Arabi, Muttan, & Suji, 2010). For this all types of edge operators are applied and the one which gives the ideal result for the particular nature of image is adopted. Next step of feature extraction is applied to get the area of tumor and then to calculate this area to reconstruct the 3D image. This is done by taking multiple images of the same infected patient from different angles so that a complete tumor image can be reconstructed with the help of MATLAB 3D graph utility. Exact size of the tumor is calculated by the summation of 2D segmented pixels and the pixels of tumor thickness. Area calculation of tumor then proceeds towards the stage classification by the proposed CAD system (El-Shenawee, 2011).

2.5 Use of K-Nearest Neighbor and Fuzzy C-Means

The intensity discriminations are optimized as a tool to detect micro calcifications that further develop into a breast tumor (Moore, Nall, & Tim, 2016). Addition of standard deviation and window means to the visual figure has improved the segmentation results produced by the knn-means (Bezdek, Hall, Clark, Goldgof, & Clarke, 1997). The technique mentioned in this paper used two methods and they are supervised (k-Mean) and unsupervised (fuzzy c-Means) (Singh & Mohapatra, 2011). Both these methods use the physical labeling and clustering of input image by the operator or human.

FCM will tend to run slower than K-means, since it is actually doing more work (Singh & Mohapatra, 2011). Each point is evaluated with each cluster, and more operations are involved in each evaluation. K-means just needs to do a distance calculation, whereas FCM needs to do a full inverse-distance weighting.

This proposed work makes these differences or weakness of strong point for full detection of breast cancer. From this algorithm was able to find out the masses as well as the cancerous area. And the result shows whether the pixel is a tumor tissue or not a tumor tissue.

2.6 Vector Quantization

The method of clustering and texture analysis is also known as vector quantization. This is the method used for segmentation of mammographic images that implemented the generation of codebook or training vector by dividing the whole image into small boxes of equal size. There is use of LindeBuzo and Grey Algorithm (LBG) for vector quantization classifying the complete image into training vectors (Kekre, Sarode, & Gharge, 2009).

After that a centroid is computed and the Euclidean distance is computed between training vectors on the basis of which 8 clusters and a codebook of size 128 are formed. But image constructed using first code vector displayed the result correctly.

2.7 Detect Breast Cancer using blood sample

An Improvement of traditional Watershed algorithm was used in (Shete, Kharate, & Rege, 2012). The Estrogen Receptor (ER), an element of the cell, is used to identify cancer cells in the tissue sample. The methods involved in image processing are HSV color conversion from RGB image, Hue, saturation and value based object-background separation, morphological operations such as dilation and closing, and area based filtering for preliminary preparation of image for detailed analysis. Then a modified watershed algorithm designed for eliminating errors arising due to over-segmentation in traditional watershed algorithm is proposed to provide comparatively more accurate results.

Further, intensity based thresholding is performed for identifying and categorizing the cancerous cells into levels of severity of damage done to cells due to cancer. The proposed modified watershed algorithm is compared with the original watershed algorithm and an accuracy of almost 96% was observed and verified (Shete, Kharate, & Rege, 2012).

3. DISCUSSION

All techniques discussed above have different characteristics. This survey tried to give a brief review of existing techniques which uses different methods to diagnose mammographic images. Table 1 gives a comparison with advantages, characteristics and limitations of those different techniques for cancer diagnosis on mammographic images.

Table 1. Comparison between Different Diagnosis techniques on mammographic images

Technique/Algorithm	Advantages/Characteristics	Limitations
Digital Mammography	Calcification, circumscribed, speculated and other ill-defined masses can be diagnosed	Frequent intensity changes may not give 100% diagnostic results
Screening Mammography	Diagnosis is based on the roughness between normal and tumor tissues	-
MRF and HPACO algorithms	Success rate is 94.8%	-

Image Segmentation and 3d Structure Analysis	Exact size and thickness of tumor can be calculated	Need to combine CAD system with segmentation algorithm
K-Nearest Neighbor and Fuzzy C-Mean algorithm	Change of intensity is used as a discriminating feature	-
Vector Quantization	Absence of over and under segmentation	Lot of segmentation areas and information
Enhanced Watershed algorithm	Accuracy is almost 96%	Threshold level for minimum distance used in the algorithm is manually chosen

Other than mammographic images, other types of images such as Thermal Infrared Images, Microscopic Slide Images, Ultrasound Images, MRI and CT scan Images are being used in cancer identification.

4. CONCLUSION

This study briefs some selected techniques used in breast cancer detection based on mammographic images. New enhancements in image processing techniques increase the performances and accuracy of detection of breast cancers. The introduction of new techniques in image processing area open up numerous future enhancement topics in breast cancer identification due to rapid development and involvement of various computing technologies such as Artificial Intelligence, Big data Analysis and etc.

5. REFERENCES

Alhadidi, B., Zubi, M., & Suleiman, H. (2007). Mammogram Breast Cancer Image Ditection Using Image Processing Functions. *Information Technology Journal, 6*(2), 217-221.

Arabi, P. M., Muttan, S., & Suji, R. (2010). Image enhancement for detection of early breast carcinomaby external irradiation. *IEEE International conference on Computing, Communication and Networking Technologies*, (pp. 1-9).

Ball, J. E., & Bruce, L. M. (2007). Digital Mammogram Spiculated Mass Detection and Spicule Segmentation using Level Sets. *International Conference of the IEEE Engineering in Medicine & Biology Society* (pp. 4979-4984). Lyon, France: IEEE.

Bezdek, J. C., Hall, L. O., Clark, M. C., Goldgof, D. B., & Clarke, L. P. (1997). Medical image analysis with fuzzy models. *Statistical Methods in Medical Research, 6*(3), 191-214.

Breastcancer.org. (2015). *What Is Breast Cancer?* Retrieved 02 12, 2016, from Breastcancer.org:
http://www.breastcancer.org/symptoms/understand_bc/what_is_bc

El-Shenawee, M. (2011). Electromagnetic Imaging for Breast Cancer Research. *IEEE Biomedical Wireless Technologies, Networks, and Sensing Systems*, (pp. 55-58).

Gonzalez, R. C., & Wintz, P. (2010). *Digital image processing.* Addison-Wesley Publishing Co.

Groot, M. T., Baltussen, R., Groot, C. A.-d., Anderson, B. O., & Hortobágyi, G. N. (2006, January). Costs and health effects of breast cancer interventions in epidemiologically different regions of Africa, North America, and Asia. *The Breast Journal, 12*(1), 81-90.

Highnam, R., & Brady, M. (1999). Mammographic Image Analysis. In R. Highnam, M. B., & M. A. Viergever (Ed.), *Computational Imaging and Vision* (Vol. 14). Kluwer Academic Publishers.

Hortobagyi, G. N., Salazar, J. G., Pritchard, K., Amadori, D., Haidinger, R., Hudis, C. A., et al. (2005, December). The Global Breast Cancer Burden: Variations in Epidemiology and Survival. *Clinical Breast Cancer, 6*(5), 391-401.

Irwin, M., Downey, D., Gardi, L., & Fenster, A. (2008). Registered 3-D Ultrasound and Digital Stereotactic Mammography for Breast Biopsy Guidance. *IEEE Transactions on Medical Imaging, 27*(3), 391-401.

Kekre, H. B., Sarode, T. K., & Gharge, S. M. (2009). Tumor Detection in Mammography Images using Vector Quantization Technique. *International Journal of Intelligent Information Technology Application, 2*(5), 237-242.

Khuzi, A. M., Besar, R., Zaki, W. W., & Ahmad, N. N. (2009). Identification of masses in digital mammogram using gray level co-occurrence matrices. *Biomedical Imaging and Intervention Journal, 5*(3).

Mohanalin, J., Kalra, P., & Kumar, N. (2008). Fuzzy based micro calcification segmentation. *5th International Conference on Electrical and Computer Engineering*, (pp. 49-52). Dhaka.

Moore, K., Nall, R., & Tim, J. (2016, July 25). *Calcification*. (J. Marcin, Editor) Retrieved August 15, 2016, from healthline.com: https://www.healthline.com/health/calcification#overview1

National Cancer Institute (NCI), the U.S. . (2016). *Diagnosis*. Retrieved 02 15, 2016, from Cancer.gov: https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis

Parkin, D. M., & Fernández, L. M. (2006, January). Use of Statistics to Assess the Global Burden of Breast Cancer. *The Breast Journal, 12*(1), 70-80.

Parkin, D. M., Bray, F., Ferlay, J., & Pisani, P. (2002). Global Cancer Statistics. *A Cancer Journal for Clinicians, 55*(2).

Radiological Society of North America (RSNA). (2016). *Breast Cancer Screening*. Retrieved 02 15, 2016, from RadiologyInfo.org: https://www.radiologyinfo.org/en/info.cfm?pg=screening-breast

Shete, P. G., Kharate, G. K., & Rege, S. C. (2012). Breast Cancer Cell Detection Using Image Processing. *International Journal of Engineering Research & Technology (IJERT), 1*(9).

Singh, N., & Mohapatra, A. (2011). Breast Cancer Detection in Mammograms Using K-means and C-means Clustering. *International Journal of Computer Applications, 22*(2), 15-21.

Suckling, J., Parker, J., Dance, D. R., Astley, S., Hutt, I., Boggis, C. R., et al. (1994). The Mammographic Image Analysis Society Digital Mammogram Database. *Digital Mamography*, 375-378.

Sultana, A., Ciuc, M., Strungaru, R., & Florea, L. (2010). A new approach in breast image registration. *IEEE International Conference on Intelligent Computer Communication and Processing* (pp. 149-154). IEEE Computer Society.

Thangavel, K., Karnan, M., Siva Kumar, R., & and KajaMohideen, A. (2005). Segmentation and Classification of Micro calcification in Mammograms Using the Ant Colony System. *International Journal on Artificial Intelligence and Machine Learning, 5*, 29-40.

Vainio, H., Bianchini, F., & Heseltine, E. (2002). Breast cancer screening. In *IARC handbook of cancer prevention.* Lyon, France: IARC Press.

Wang, K., Qin, H., Fisher, P., & Zhao, W. (2006). Automatic registration of mammograms using texture-based anisotropic features. *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro*, (pp. 864-867). Arlington, VA, USA.

Yao, Y. (2004). Segmentation of Breast Cancer Mass in Mammograms and Detection using magnetic resonance imaging. *IEEE Image Processing*.

Yasmin, M., Sharif, M., & Mohsin , S. (2013, October). Survey Paper on Diagnosis of Breast Cancer Using Image Processing Techniques. *Research Journal of Recent Sciences, 2*(10), 88-98.

Zadeh, H., Janianpour, S., & Haddadnia, J. (2013). Racognition and Classification of the Cancer Cells by Using Image Processing and LabVIEW. *International Journal of Computer and Engeneering, 5*(1), 104-107.