

Signal Processing Techniques and Computer-Aided Detection Systems for Diagnosis of Breast Cancer – A Review Paper

Susama Bagchi and Audrey Huong*

Universiti Tun Hussein Onn Malaysia, Malaysia;
bagchi.susama@gmail.com, audrey@uthm.edu.my

Abstract

This work aimed to review different Computer-Aided Detection (CAD) systems that were proposed as alternative means to replace the tedious and erroneous double reading procedure via radiologists. These CAD systems include the use of different signal processing techniques such as Wavelet Transform and Curvelet Transform, image processing technique, pattern recognition, artificial intelligence technologies namely the Artificial Neural Network and Fuzzy Logic system, and different algorithms of computer sciences. Among the developed algorithms proposed for this purpose include k-nn algorithm, fuzzy C-means (FCM), swarm algorithm, genetic algorithm and multi-resolution techniques. It was found from this study that multiscale curvelet transform has the highest classification accuracy with the reported value up to 98.59 %, followed by the Swarm Optimization which produced a percentage error of 1.7 %. Meanwhile it was observed that multi-resolution technique along with genetic algorithm produced the highest error of 20.8 ± 8 % in its diagnosis. This work concluded that curvelet transform and swarm algorithm is, thus far, the most suitable CAD techniques to be used before clinical investigation of malignant breast tissues. In the future, these techniques may be improved further to detect different stages of breast cancer.

Keywords: Breast Cancer Detection, Computer-Aided Detection System

1. Introduction

Cancer is the major health concern over the last few decades. Approximately 1.7 million new breast cancers were detected in 2012 which is nearly 11.9 % of total new cancer cases and ranks fifth as cause of death¹. In Malaysia, approximately 5,000 women with breast cancer are diagnosed every year. They are mostly aged between 30 to 60 years and almost half of those are under 50-years. Among all imaging techniques, mammogram (Special X-ray for breast tissues) is the preliminary method to diagnose breast cancer; an individual with suspected cancerous cell would be referred for biopsy to confirm diagnosis.

Mammogram diagnosis is the widely accepted technique to date. Mammogram is of two types, namely film mammography and digital mammography. This review considers only digital mammogram largely due

to the ease of image manipulation and the fast screening time. Two methods can be followed to find out the presence of cancerous breast tissues depending on mammography, which is either screening it by more than one senior radiologist or by using CAD system².

The chances of life risk of a breast cancer patient can be minimized by early detection³. However, the cost associated with breast cancer detection, starting from mammography screening to clinical examination, is always a concern throughout the world and it varies in different countries depending on several factors⁴⁻⁶. Implementation of regular breast cancer screening program can help in this regard which in turn will increase the overall cost⁴. From the report of⁵, in the USA, the Medicare fee-for-service program pays out more than 45 % of their total spent per year on breast cancer screening. According to⁷, yearly biopsies for breast cancer cases are executed beyond 1 million and among them benign cases

* Author for correspondence

are about 80 %. Biopsy is an aggressive invasive method that involves several risks and patient discomfort³. It is also a tough job to separate a benign tumor from the malignant one since it is dependent on the experience of the radiologist. Statistics reveals that cancerous cases are only 20–30 % of all breast biopsies and in 10 % cases mammography is unable to detect breast cancer⁸. Furthermore, the digital mammograms are considered as one of the highly complicated medical images amongst all since there are several kinds of breast tissues to be studied which differs a lot in all aspects. Initial signs of masses and microcalcification clusters are the important visual clues for breast cancer detection. Unfortunately all these indications are very illusive and appeared differently at primitive stage and hence, even for the specialists also the diagnosis becomes harder⁹. A better accuracy can be achieved in detecting the cancerous tumor if more than one radiologist is recruited to perform mammogram screening, but this can be time consuming and costly. Hence, as an effort to minimize the number of avoidable breast biopsies and overall cost of mammography screening⁷, CAD systems are evolved to facilitate the radiologists with single reading. Although in¹⁰, it has been stated that the CAD system enhances performance of a single reading and yields an increased sensitivity, but statistically its recall rate is not significantly different than the existing double reading. Therefore, researches are going on to improve the outcome of the CAD systems to find out the malignant breast tissues so that refined criteria can be provided before going for biopsy.

This paper provides a brief overview on different CAD systems that are recently used for breast cancer detection and they are structured as follows: In Section 2, masses and microcalcifications features are explained. In Section 3, different signal processing techniques and other state of the art of CAD systems to diagnose breast cancer are discussed. In Section 4, a comparative study has been done for the performance of different approaches. The conclusion is included in Section 5.

2. Types of Tissues/Tumors in Breast Cancer

There are two types of tissues in mammography on which radiologists rely during breast cancer diagnosis, namely microcalcifications and masses. The diagnosed results are of three types: normal in which no cancerous cells are found in mammogram, benign in which tumors without

any malignant cell are found in mammogram and malign in which tumors with cancerous cell are found in mammogram⁸. The comprehensive mammography images of circumscribed masses, spiculated masses and microcalcifications can be found in the work⁸.

A mass is a space-occupying lesion generally identified through its features like density, margin and location⁸. Benign masses are round shaped; have low-density and smooth and properly defined margins. Stellate shaped, spiculated high-density masses with inadequate margins are generally recognized as malignant. Architectural distortion and bilateral asymmetry are other characteristics of masses.

Microcalcifications emerge as minute bright marks in the mammogram which are actually very small calcium depositions. Their average size is about 0.3 mm¹¹ and they may be with or without mass⁸. Benign microcalcification is usually coarse, large (diameter around 1–4 mm), identical in size and shape, round or oval, dispersed or diffused. Malignant tissue is microscopic, innumerable, stellate-shaped, of different size and shape and may be found in group or cluster or as linear branching. If more than five microcalcifications are found in a cluster, then usually they are malignant⁸.

3. Different Signal Processing Techniques and Other CAD Techniques

Figure 1 shows the working principle of a CAD system. The pre-processing stage involves cropping of an image. Image enhancement, segmentation, feature extraction and cluster detection are performed in detection stage. The final stage of the process is classification which involves feature selection and it is an important step in diagnosis.

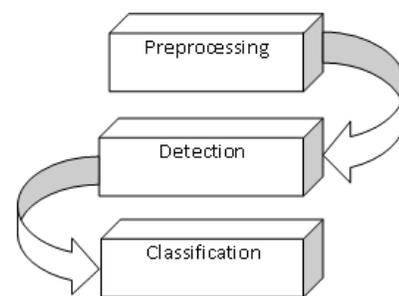


Figure 1. The Basic Work Principle of a Computer-Aided Detection (CAD) System.

Several signal processing techniques, like microwave imaging, ultrasound imaging, wavelet transform and curvelet transform are used for breast cancer detection as a part of CAD system. Here this study highlights the distinctive features of wavelet and curvelet as they seem to be promising in providing better performance. In addition, MATLAB as a signal processing tool, artificial neural network (ANN), fuzzy logic and neuro-fuzzy logic will also be highlighted briefly in the following sections.

3.1 Wavelet Transform

Wavelet transformation is the time-frequency illustration of a signal using small waveform, known as wavelet. The latter is time limited and has an average value of zero. Here the time-domain signal travels through various high-pass and low-pass filters and breaks down into a shifted and scaled version of the initial wavelet by filtering out either the high frequency or low frequency segments of the signal. Whenever for every new cycle this method is repeated with a slightly reduced (or increased) window, it will remove some part of the signal related to some frequencies of that signal. The signal-cutting problem can be solved by using a completely scalable modulated window and every time by shifting the window, the spectrum will be measured¹². If it is considered that $f(x)$ is the one dimensional signal, then after first decomposition two outputs will be obtained, where the coefficients for low frequency and high frequency parts are mentioned as $\varphi(x)$ and $\psi(x)$ respectively. Since the high frequency components generally contain more noise, it may be discarded and low frequency components can be further decomposed to obtain a better de-noising effect as shown in Figure 2. After the high-pass and low-pass filtering, the decomposed signal is down sampled to keep the number of samples of the output signal same as the input signal.

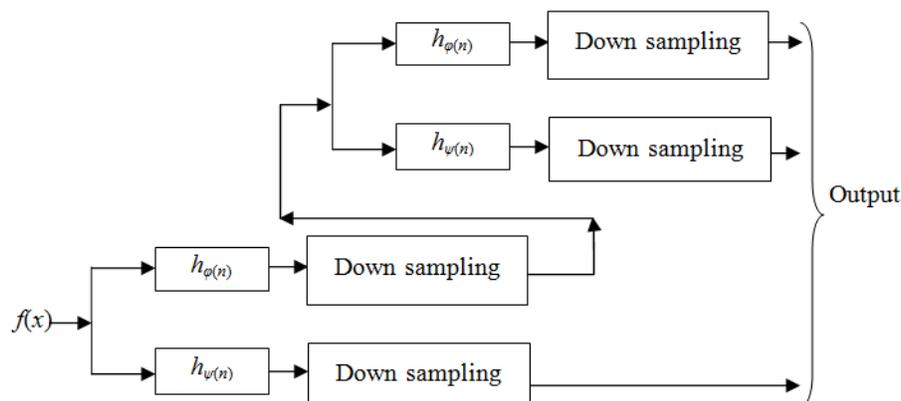


Figure 2. Two Stage Decomposition of Wavelet Transform.

The $h_\varphi(n)$ is known as the scaling function and $h_\psi(n)$ is the wavelet function.

In Continuous Wavelet Transform (CWT), a continuous time-frequency function is divided into wavelets for time and frequency localization. Here the frequency is dependent on the scaling factor. For lower scaling factor, signal will be compressed while higher factor will stretch the same. The disadvantage of CWT is that the compressed signal will not last for its entire duration while the stretched signal will fail to capture the details.

When a wavelet is discretely sampled for a particular scaling value, it is known as Discrete Wavelet Transform (DWT). This transform acquires both frequency and location information, and it has better efficiency in removing noise while leaving the image intact. Hence, by using DWT it is easy to identify subtle abnormalities in images and enhance it further. However, DWT is translation invariant¹³ and to overcome this, undecimated wavelet transform can be used¹³. The latter omits both down-sampling in the forward and up-sampling in the inverse transform. It must also be mentioned that thresholding of wavelet coefficient has major effect on image de-noising.

The Dyadic Wavelet Transform can be deduced from CWT¹⁴. This transform has very good applications where wavelets are required for sharp prescribed localization and fast decay in spatial and as well as in frequency domain.

3.2 Curvelet Transform

Curvelet transform can be derived from wavelet transform. It can produce images at various scales and its degree of localization differs with scale. This transform can identify skinny ridges with a precisely determined orientation and verify the multi-dimensional features at

wedges. Therefore, it is mostly used for feature extraction to distinguish the smooth tissues with disruptions near curves for cases such as sparser illustration of edgy matters, reformation of an image for severely damaged problems in its best way and optimal sparse representation of wave propagators³.

3.3 Other CAD Systems

Feature selection is a very important process for classification to differentiate between normal and abnormal categories of breast tumors. This includes the use of artificial intelligence, artificial neural network (ANN), fuzzy logic and neuro-fuzzy system.

The ANN describes a computer model by assuming how the human brain works. A set of neurons or nodes are the building blocks of ANN. The nodes are the processing units and must be internally connected. Based on the received input signal a node generates an output signal to solve a particular problem. It has adaptable gains or weights that are gradually accustomed through the repetition of the input-output instruction provided to the ANN. The working process is divided into two phases, learning and recall. The machine learning procedure may be supervised or unsupervised. In the supervised method human intervention is required in different phases to train the system¹⁵. The unsupervised method is dependent on machine learning to describe hidden structure of unlabeled data. In learning phase, ANN is trained to its task through the adaptation of its weight and the problem is solved through the recall phase. This learning process is viable with feedforward or feedback method.

The fuzzy logic reflects some attributes of human reasoning processes like logic reasoning, hypothesis, inference and can deal a certain level of uncertainties, where the true value may range between completely true and completely false. This technique is easy to understand and implement, and is able to provide a user friendly approach of presentation efficiently. It is also advantageous when neither assumption can be made for both the linearity and time invariance of the controlled process, nor the process has a properly framed mathematical representation. When the work procedure of the model totally deviates from human perception, then also fuzzy logic can provide better result¹⁶. However, it is really difficult to create a model using fuzzy system since it requires precise tuning and simulation before operation and it is tough to identify its proper membership values. Fuzzy controller basic work principle has 4 segments: (i) Fuzzification interface

changes and transforms inputs to proper linguistic values. (ii) A set of instructions are maintained in Rule Base as its know-how to regulate the system in best way and compare the outcome of the fuzzification interface. (iii) Inference mechanism decides the control rules which are required for given problem and then opts for the correct input. (iv) The result of the inference mechanism is then modified by the Defuzzification Interface into crisp decisions or control signals¹⁶.

The neuro-fuzzy logic is a hybrid intelligent system to resolve any problem in human-like reasoning style by using fuzzy sets consisting linguistic model of *if-then* rules and exploiting the features of artificial neural networks, like robustness, massive parallelism and learning in data-rich environment. The ANFIS (Adaptive Neuro-Fuzzy Inference System) is a type of heterogeneous neuro-fuzzy system which utilizes *if-then* rules of fuzzy system and the learning procedure can be done through feedforward or back-propagation method. It can approximate the non-linear function through a set of directly connected nodes. Here each node acts as a processing unit to perform the static node functions on its incoming signals and produces a single node output. Usually every node has different functions. The node to node signal flow is directed by each link which does not have any weight or parameter in general. The node functions can be changed by changing its parameters which in turn will modify the overall behavior of the adaptive network.

3.4 MATLAB as Signal Processing Tool

MATLAB is a performance oriented multi-paradigm numerical computing environment that incorporates visualization, computation and programming in a user-friendly way where known mathematical representations are used to specify the problems and their solutions¹⁷. It has toolboxes for image processing, signal processing, neural network and fitting. Image processing toolbox supports image enhancement and its analysis, Region of Interest (ROI) activity, linear filtering and designing of filter. Engineers and scientists use this software to achieve the required result by selecting proper tools and the features.

4. Comparative Study of Computational Approaches

Masses are far complicated than microcalcifications

during detection since the traits of the former are hard to perceive and sometimes same as normal breast tissues². Since microcalcifications have higher contrast than rest of the region of ROI and corresponds to high frequency components, they may be easy to detect by image enhancing and de-noising as it was done in¹¹ by using dyadic wavelet processing. But masses have low contrast, varying densities, spiculated structures and correspond to low frequency components. Multiscale curvelet transform can provide a better result for mass detection due to its advantages over wavelet and the outcome can be seen in³ which suggested the classification accuracy rate of 98.59 %. Whereas according to⁸, for microcalcification cases the achieved classification rate was best with the use of wavelet analysis and ANFIS by extracting features at 2–3 levels and wavelet decomposition represents them at the utmost level. Alternatively for masses, the obtained rate of classification was best when from 3–4 levels the features were extracted since they are visible, of bigger size, and characterized by low frequency data and wavelet decomposition represents them at the lowest levels⁸.

In¹⁸, the researchers made a computational technique to detect and divide regions in mammographic images using genetic algorithm and multi-resolution techniques, and the corresponding work presented a relatively high accuracy result. The researchers concluded that this transform algorithm presents some advantages in terms of shift insensibility, high directionality and its ability to provide phase information. It was also mentioned that curvelet and shearlet transforms are another examples of possible future investigation as a way to improve the efficiency of the proposed algorithm and to obtain directional information¹⁸.

The supervised method of segmentation in digital mammograms by using crisp k-nn (k-nearest neighbor) algorithm and unsupervised method of segmentation by using fuzzy c-means (FCM) is attempted in¹⁵ and the results revealed better misclassification rates with intensity as the differentiating feature. Using window means and standard deviations as extra features, the amount of unspecified pixels in few regions within the image can be reduced, but the result is not much of satisfaction¹⁵. Meanwhile, according to¹⁹, the accuracy for nodules and calcifications were obtained 76.67 % and 83.34 % respectively by using Fuzzy Omega Algorithm. It also suggested that as fuzzy sets the BI-RADS (Breast Imaging-Reporting and Data System) features can be

used for which it is feasible to establish the relevance of the result of every BI-RADS type and gives the user an opportunity to execute the quantitative analysis. In other studies different other algorithms are used, such as in²⁰, association rule (AR) was implemented in reducing the element from the database of breast cancer and at classification part Apriori algorithm in Neural Network (NN) was tried. This work achieved the accuracy of 95.6 %. Likewise in²¹, FCM was used for segmentation and the classification of the features were done by applying CAT (algorithm) Swarm Optimization (CSO) with Optimal Brain Damage (OBD) pruning neural network for which obtained accuracy was 98.3 %.

The brief summary of the reported classification rate from the previous studies is incorporated in Table 1 along with the used CAD techniques.

Table 1. Brief Summary of Signal Processing Techniques and Other CAD Systems used in Breast Cancer Determination and their Reported Accuracy

| No. | Proposed CAD Techniques | Reported Classification Accuracy |
|-----|---|--|
| 1. | Genetic Algorithm and Multi-resolution Techniques ^[18] | 79.2 ± 8 % |
| 2. | BI-RADS standard and Fuzzy Omega Algorithm ^[19] | 76.67 % for nodules and 83.34 % for calcifications |
| 3. | Wavelet Analysis for feature extraction and ANFIS for classification ^[8] | 87.5 % for microcalcification at level 2-3, 93.7 % for masses at level 3-4 |
| 4. | AR and NN (Apriori Algorithm) ^[20] | 95.6 % |
| 5. | FCM and CSO with OBD pruning ^[21] | 98.3 % |
| 6. | Multiscale Curvelet Transform ^[3] | 98.59 % |

5. Conclusion

This paper discusses the recent signal processing techniques and other CAD systems that are utilized for the breast cancer detection. It is clear from all the works done up to this date regarding digital mammogram that none of the breast tissue classification system using CAD is able to provide 100 % accuracy. Different wavelet analysis can be tried for de-noising so that microcalcifications are prominent in mammogram which in return may facilitate the classification. On the other hand, curvelet analysis may provide better result for masses for its special characteristics. In classification, different feature selection

and threshold may help to improve the classification accuracy.

Meanwhile for the CAD system, no noteworthy differences were observed between the existing double reading and single reading system. Nonetheless, the former has a better sensitivity. It was also found that the existing CAD systems are overstressing on the sensitivity for the diagnosis of microcalcifications at the expense of specificity for which the number of unnecessary biopsies are increasing. Hence, there are rooms of improvement on the performance of CAD systems for early detection so that actual cases of breast cancer can be referred for biopsies and the number of mortality can be reduced at the same time. If the recall rates can be minimized, then the overall cost for screening may also be reduced.

6. Acknowledgments

We are grateful to Ministry of Education Malaysia under FRGS grant number 1581 for financially supporting this work.

7. References

- Jacques F, Isabelle S, Rajesh D, Sultan E, Colin M, Marise R, Donald MP, David F, Freddie B. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer*. 2015 March; 136(5): E359–E86.
- Jinshan T, Rangaraj MR, Jun X, Issam EN, Yongyi Y. Computer-Aided Detection and Diagnosis of Breast Cancer with Mammography: Recent Advances. *IEEE Transactions on Information Technology in Biomedicine*. 2009 March; 13(2): 236–51.
- Mohamed ME, Ibrahima F, Brahim BS. Breast Cancer Diagnosis in Digital Mammogram Using Multiscale Curvelet Transform. *Computerized Medical Imaging and Graphics*. 2010 June; 34(4): 269–76.
- Robert AS. Cost-effectiveness of Population-Based Breast Cancer Screening. In: *IARC Handbooks of Cancer Prevention, Volume 7: Breast Cancer Screening*. International Agency for Research on Cancer Publ.: USA. 2002; 157–70.
- Cary PG, Jessica BL, Joseph SR, Maysa MAK, Rong W, Brigid KK, Heather TG, Anees BC, Xiaomei M. The Cost of Breast Cancer Screening in the Medicare Population. *JAMA Internal Medicine*. 2013 February; 173(3): 220–6.
- Quirine LO, Gerrit D, Arno DK, Martin LB, Harry JDK. Breast Cancer Screening Policies in Developing Countries: A Cost-Effectiveness Analysis for India. *Journal of the National Cancer Institute*. 2008 September; 100(18): 1290–1300.
- David G, Luisa PW, Amy HK, Lara AH, Gordon SA, Ratan S, Jules HS. Trends in Recall, Biopsy, and Positive Biopsy Rates for Screening Mammography in an Academic Practice 1. *Radiology*. 2005 May; 235(2): 396–401.
- Rafayah M, Qutaishat M, Abdallah M. Breast Cancer Diagnosis System Based on Wavelet Analysis and Fuzzy-Neural. *Expert Systems with Applications*. 2005 May; 28(4):713–23.
- Thangavel K, Roselin R. Fuzzy-Rough Feature Selection with Π -Membership Function for Mammogram Classification. *IJCSI International Journal of Computer Science Issues*. 2012 May; 9(4).
- Matthew G. Comparison of Computer-Aided Detection to Double Reading of Screening Mammograms: Review of 231, 221 Mammograms. *American Journal of Roentgenology*. 2008 April; 190(4): 854–9.
- Arianna M, Marcello S, Roberto L, Manuela F, Federica C. Mammographic Images Enhancement and Denoising for Breast Cancer Detection Using Dyadic Wavelet Processing. *IEEE Transactions on Instrumentation and Measurement*. 2008 July; 57(7): 1422–30.
- Supriya G, Pradeep T, Gurjit K. A Fuzzy Based Approach for Denoising of ECG Signal Using Wavelet Transform. *International Journal of Bio-Science and Bio-Technology*. 2016; 8(2): 143–156.
- Aglia G, Chandrika K, Imola KF. Undecimated Wavelet Transforms for Image De-Noising. Report Lawrence Livermore National Laboratory. 2002 November; 12: 1–12.
- Pierre V, Jean-François G. Directional Dyadic Wavelet Transforms: Design and Algorithms. *IEEE Transactions on Image Processing*. 2002 April; 11(4): 363–72.
- Tobias CC, Melanie AS, James CB. Breast Cancer Detection Using Image Processing Techniques. In *Fuzzy Systems, 2000. FUZZ IEEE 2000. The Ninth IEEE International Conference on (Vol. 2, pp. 973–976)*. IEEE. 2000;2:p.973-76.
- Sunny S. A Review on Efficacy of Artificial Neural Networks in Medical and Business Areas. *International Journal of Recent Trends in Engineering and Research*. 2016 April; 2(04).
- Basim A, Mohammad HZ, Hussam NS. Mammogram Breast Cancer Image Detection Using Image Processing Functions. *Information Technology Journal*. 2007; 6(2): 217–21.
- Danilo CP, Rodrigo PR, Marcelo ZDN. Segmentation and Detection of Breast Cancer in Mammograms Combining Wavelet Analysis and Genetic Algorithm. *Computer Methods and Programs in Biomedicine*. 2014 April; 114(1): 88–101.
- Gisele HBM, Joaquim CF. Computer-Aided Diagnosis System Based on Fuzzy Logic for Breast Cancer Categorization. *Computers in Biology and Medicine*. 2015 September; 64: 334–46.
- Murat K, M. Cevdet I. An Expert System for Detection of Breast Cancer Based on Association Rules and Neural Network. *Expert Systems with Applications*. 2009 March; 36(2): 3465–69.
- D. Thuthi S, Arthi K. Efficient Breast Cancer Classification Using Improved Fuzzy Cognitive Maps with Csonn. *International Journal of Applied Engineering Research*. 2016; 11(4): 2478–85.