Computer-Aided Detection and Diagnosis for Microcalcifications in Mammogram: A Review

Norlia Md Yusof[1], Nor Ashidi Mat Isa[2] and Harsa Amylia Mat Sakim[3]
School of Electrical & Electronic Engineering,
Universiti Sains Malaysia, Engineering Campus,
14300 Nibong Tebal, Penang, MALAYSIA

Summary
Breast cancer continues to be a significant public health problem among women around the world. It has become the number one cause of cancer deaths amongst Malaysian women. The key to improve the breast cancer prognosis is by early detection. The important sign for the breast cancer detection is the presence of lesion such as microcalcification clusters (MCCs). In this review paper, the mammogram-based approach will be focused, as it is particularly suitable for detecting this type of lesion. To date, mammography remains the most effective diagnostic techniques for early breast cancer detection. However, due of some limitations, not all breast cancer can be detected by mammograms. The main objective of this paper is to discuss the computer-aided detection and diagnosis systems that have been proposed, designed and developed by previous researchers in order to overcome the drawbacks of mammograms by assisting the radiologists in detecting the specific abnormalities and improving the diagnostic accuracy in making the diagnostic decisions.

Key words:
Breast Cancer, Microcalcifications, Mammography, Computer-aided detection, Computer-aided diagnosis.

1. Introduction

Breast cancer is the most common cancer and continues to be a significant public health problem among women around the world. It has become the number one cause of cancer deaths amongst Malaysian women [10]. Primary prevention seems impossible since the cause of this disease still remains unknown [5]. It is believed that the most promising way to decrease the number of patient suffering from the disease is by early detection. The earlier breast cancer is detected, the better the chances that treatment will work and the better a proper treatment options can be provided.

To date, mammography remains the most effective diagnostic technique for early breast cancer detection. A mammogram is an x-ray system to examine breasts. Radiologists interpret the mammogram images for detect the abnormalities of cancerous cells such as clustered microcalcifications (MCCs), masses, architectural distortion, asymmetry between breasts, breast edema and lymphadenopathy. Then, they will diagnose the abnormalities to determine the status of breast cancer whether it is benign or malignant. In the study by Lo *et al.*, a focus is given to MCCs detection and diagnosis since its presence is one of the most important and sometimes the only sign of cancer on a mammogram [7].

While mammography is the best detection of breast cancer available today, however, not all breast cancer can be detected by mammograms [13]. For MCCs, the interpretations of their presence are very difficult because of its morphological features. For example, the sizes of MCCs are very tiny, typically in the range of 0.1mm-1.0mm and the average is about 0.3mm, implying it can easily be overlooked by a radiologist. While in some dense tissues, and/ or skin thickening, MCCs areas are almost invisible to be seen by examining radiologist. The dense tissues especially in younger women may easily be misinterpreted as MCCs due to film emulsion error, digitization artifacts or anatomical structures such as fibrous strands, breast borders or hypertrophied lobules that almost similar to MCCs. Other factors that contribute to the difficulty of MCCs detection are due to their fuzzy nature, low contrast and low distinguishability from their surroundings [5].

With the advances of computer technology, radiologists have an opportunity to improve their image interpretation because of its capabilities that can enhance the image quality of mammograms. Over the past two decades, many attempts have been made by computer scientists to assist a radiologist in MCCs detection and diagnosis by developed a computer-aided mammography (CAM). Image processing and intelligent systems are two mainstream of

Manuscript received June 5, 2007
Manuscript revised June 20, 2007

computer technologies that constantly being exploited in the development of CAM. Generally, a CAM system consists of two categories, i.e. computer-aided detection and computer-aided diagnosis, abbreviated as CAD and CADx respectively. A study by [5] reveals that the readers' sensitivity can be increased by 10 percent with the support of CAD systems in diagnosing breast cancer. [8] also reports that radiologists' performance increase when they incorporate computerized image analysis in their decision-making process for both the detection and diagnosis of cancer. Thus, a development of CAD/ CADx is highly desirable in order to assist radiologist's interpretation of specific abnormalities and to improve the diagnostic accuracy in making diagnostic decisions.

The main objective of this study is to discuss the computer-aided detection and diagnosis systems that have been proposed, designed and developed by researchers in order to overcome the drawbacks of mammograms in detecting and diagnosing the MCCs.

2. Computer-Aided Mammography Systems

As aforementioned in previous section, a CAM system is divided into CAD and CADx. CAD is a system that capable to detect a suspicious lesion from digitized or digital mammogram. Once the lesion has been detected manually by radiologists or automatically by CAD, CADx systems then assist the radiologist to classify that lesion in making a decision whether the examined lesions consist of malignant or benign tissue. The main goal of CAD is to improve the sensitivity by assisting radiologists to detect the suspicious lesion which might otherwise have been missed, while CADx is basically to improve the specificity, such as by avoiding unnecessary benign biopsies [7].

In this study, a review is given to CAD systems and CAD and CADx that incorporated into one ensemble systems, which refer as computer aided detection and diagnosis (CAD/CADx). There is no review on CADx alone like CAD since the study focused on mammogram-based approach which requires an image interpretation. Normally, researchers whom focus to the development solely on CADx, use the data that acquired from database such as Wisconsin breast cancer datasets. Some of them may use the data that can directly collect from other breast cancer detection modalities like fine needle aspirates (FNA).

A typical CAM system can be described by the block diagram in Figure 1. In the preprocessing module, mammograms will be digitized in order to be processed by computer. Since more than one-third of a mammogram is dark breast background that comprised with noise [12] and only provides very little information [14], it is better to eliminate this unwanted information. The fast retrieval and storage factor are another reason arises why the breast region extraction is needed in preprocessing stage. Next, the region of interests (ROIs) that contains possible MCCs are selected. However, some of detected pixels in ROIs may contain noise or breast tissue, so in order to extract the genuine MCCs, contrast enhancement and segmentation process are really important. The purpose of contrast enhancement is to improve the low contrast of calcified pixels while segmentation will segment the detected MCCs from the breast region. Lastly, the features of MCCs will be extracted to determine that ROIs are true cluster of MCCs. The feature extraction process in CADx extracts the features of MCCs to be an input for the classification technique in order to classify the MCCs into benign, malignant, suspicious and normal. Some researchers classify the detected clustered MCCs into five categories according to BI-RADS (Breast Imaging Reporting and Data System) i.e. negative, benign finding, probably benign finding, suspicious abnormality and highly suggestive of malignancy.

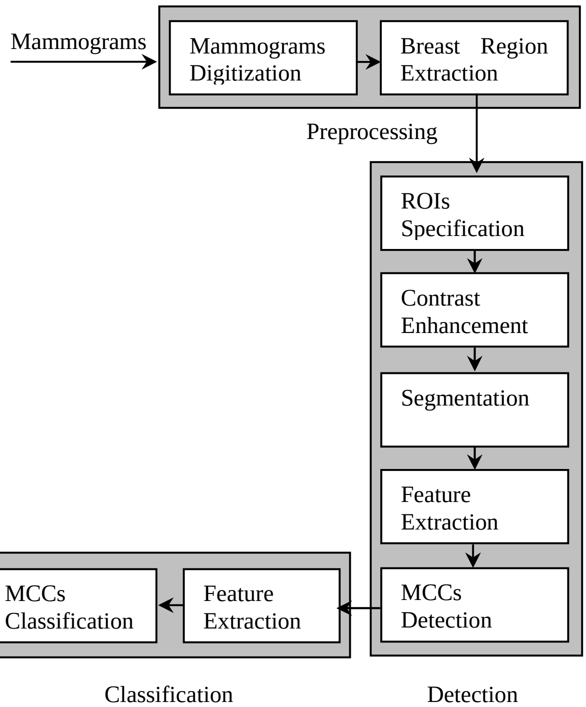

Fig. 1. A block diagram of CAM systems.

3. Computer-Aided Detection Systems

[16] have proposed a novel system for detecting clustered MCCs in Small Field Digital Mammography. The selection of ROIs uses a filter of finite impulse response where the filter accentuates the standard appearance of

MCCs. Then, the MCCs will be grouped into clusters by a series of mathematical morphology operation. Finally, a two-dimension discrete wavelet transform and median filtering are used to enhance the MCCs contrast and remove the noise from mammograms. The evaluation of the CAD systems is done by radiologists. The encouraging results show that the CAD significantly improves the detection of MCCs in Small Field Digital Mammography.

A hybrid intelligent systems (HIS) was presented by [1] for the identification of MCCs clusters. In preprocessing stage, a skin-line segmentation procedure is applied for breast region extraction process. For ROIs specification, neighbouring pixels with connectivity of eight are grouped together to create a possible MCCs. With the objective to categorize the specified ROIs as true MCCs cluster or not, HIS that consist of rule based and neural network sub system is employed. The results are evaluated as the receiver operating systems (ROC) performance and quantified by the area under the ROC curve (A_z). The Nijmegen and Mammographic Image Analysis Society (MIAS) dataset is tested and a satisfactory result is achieved. The detection specificity of the two datasets is 1.80 and 1.15 false positive clusters per image, at the sensitivity level higher than 0.90 respectively.

[15] developed a CAD systems that based on feature extraction and neural network classifier. Firstly, the potential MCCs pixels are segmented out using wavelet features and gray level statistical features. Then, a multilayer feedforward neural network classifier was implemented to generate a likelihood map of potential MCCs. Secondly, an individual MCC is detected by 31 features that extracted from the potential individual MCCs objects. General regression neural network was implemented to analyze and select the most discriminatory features. A free-response operating characteristics (FROC) curve was used to evaluate the performance of the system. Results show that the proposed system gives quite satisfactory detection performance.

[7] investigated the performance of four different approaches of feature extraction, i.e. image processing technique, radiologist-extracted, demographic data and combination between the first three approaches. The first three feature extraction approaches are known as local model while the combination of three approaches is known as ensemble model. After the breast region extraction, unsharp masking is applied to enhance the high frequency content of breast region. Then, local histogram analysis is employed to segment the individual MCCs on small and overlapping ROIs. To determine whether each ROIs contain a MCCs, histogram features are extracted and merged using backpropragation neural network classifier. Linear discriminant analysis (LDA) models are designed to evaluate the performance of local and ensemble models. From the ROC curves, each local model performed poorly; the best is one based upon image processing features which yielded A_z of 0.59 ± 0.03 and partial A_z of 0.08 ± 0.03, while for ensemble models, it improves the performance significantly with A_z of 0.69 ± 0.03 and partial A_z of 0.21 ± 0.04. This demonstrates the value of the radiologist-extracted features as a source of information for MCCs cluster detection.

Modified seed based region growing (MSBRG) was successfully detect the MCCs cluster in mammograms automatically. This new algorithm of CAD systems that modified from conventional seed based region growing technique has been proposed by [9]. The first stage in this detection algorithm is finding the threshold value by classifying two regions of clusters, i.e. the object of interest (the MCCs) and the background. This task is accomplished by moving K-means clustering algorithm, a modified version of *k*-means clustering. Once the threshold value is found, the MSBRG algorithm would then be applied. The results show that the MSBRG is capable to detect MCCs from mammography image by distinguish them from unwanted noise and background.

The Computer Assisted Library for Mammography (CALMA) that employed by [17] is a research of CAD that combined the software with grid technology. They utilize the technology due to its ability in allowing remote image analysis and interactive online diagnosis by supporting an effective tele and co-working between radiologists, cancer specialists and epidemiology experts. The MCCs detection in this system involves two different neural networks where fee-forward neural network classifies the segmented image at the initial stage. Then, if the image is classified as positive MCCs, it will become an input for the second neural network, that using the principal component analysis. The PCA classifies the ROIs as MCCs clusters by pointed the MCCs area using marker if the output value exceeds a threshold. The research obtains a sensitivity and specificity of 92% and proves that the CALMA system can be used as second reader and can be industrially developed.

4. Computer-Aided Detection and Diagnosis Systems (CAD/CADx)

[21] carried out a new algorithm using optimal thresholding and zonal Hough transform to suppress the pectoral muscle as it does not provide any meaningful information. Furthermore, by removing a pectoral muscle, the detection region could be reduced. All the MCCs are detected using discrete wavelet transform (DWT) and

filling dilation technique was used to segment the MCCs from the background. In this experiment, 10 features are selected to represent the MCCs, including area, mean intensity, contrast, coherence, compactness, ratio of pits, number of hollows, elongatedness, fractal dimension, and clustering number. Multi-layer perceptron (MLP), a conventional technique of NN, is used for the classification purposes with the architecture of 10 input nodes, 20 hidden nodes and one output node. This experiment achieves a promising result with the true positive rates of 96.9 % and only 0.2 false positive per image.

The proposed CAD/CADx systems by [12] emphasized on wavelet analysis for feature extraction and adaptive neuro-fuzzy inference systems (ANFIS) for classification purpose. In the preprocessing stage, image pruning that involve the process of background removal is applied using a cropping operation. While global gray level thresholding and histogram equalization deal with enhancing the contrast of MCCs, features are then extracted from the enhanced images based on the wavelet decomposition process. The research has shown that this method is very effective for the automatic detection and classification of MCCs in digital mammograms.

The CAD/CADx that constructed by [2] is a continuity from the earlier work [1] of CAD that has been described in previous section. Hence, in this section only CADx part will be explained. The characterization of each MCCs cluster as benign or malignant is employed using two different classification schemes that based on support vector machine (SVM) and neural network. ROC analysis is applied to measure the performance of the classification method in two well established mammographic datasets, the Nijmegen and MIAS. From the experiment, SVM with Gaussian kernel function provides the best performance of classification compared to neural network.

To support an effective resource sharing and co working among radiologist from geographical distant location, [11] exploited the grid technology that embedded into CAD/CADx systems. A grid approach allows remote image analysis and interactive online diagnosis among clinicians in the interpretation of mammographic data. They developed a MammoGrid that based on wavelet analysis for CAD and neural network for CADx. The wavelet based filter processes the breast region after the identification of breast skin line. Decomposition of breast region in NxN pixel wide partially overlapping sub image is the main step of automatic feature extraction. Feedforward neural network has been chosen to perform the classification which based on the backpropagation algorithm. The performance of MammoGrid achieves a sensitivity of 82.2% at a rate of 4.15 false positive per image.

Due to its ability in enhancing the local contrast of MCCs, [4] focused on the wavelet multiresolution analysis for the enhancement of MCCs. The seed growing technique is applied for the segmentation purpose. Two techniques are being used to classify the MCCs i.e. geometrical (shape features) classification and cluster classification. The systems gives good performance qualities by achieved 82.3% for geometrical classification and 73.7% for cluster classification.

[6] examined the efficiency of three set of feature extraction descriptors i.e. gray level histogram moments (GLHM), spatial gray level dependence matrix (SGLD) and independent component analysis (ICA) in order to classify the MCCs for the breast cancer diagnosis. GLHM and SGLD are based on textural analysis while ICA is a signal processing methods. To limit the dimensionality of features set, principal component analysis is implemented as a preprocessing step. Radial Basis Function classifier is employed to recognize any type of cancer in mammograms. Classification result shown that ICA has the highest recognition rates, with 79.31% compared to 72.41% for SGLD and 70.68% for GLHM respectively. The proposed new approach of ICA gives better performance in accuracy of diagnosis than textural analysis that is widely approach used in CAD/CADx system by various researchers.

Breast Cancer Detection System (BCDS) was developed by [3] with the aim to investigate the effectiveness of a combination features to be extracted. 14 features are selected to be most significant combination features based on neural network classification rate. Initially, each feature is fed as single input into neural network. If its classification rate is increased or unchanged, this feature will be included to the input vector. Only five most significant features are selected i.e. skew, entropy, number of pixels, histogram and standard deviation. BCDS has achieved a promising result, with 88.9% classification rate.

A research conducted by [14] presented other automatic detection and classification of MCCs. A block region growing and K-means clustering-based thresholding is employed to extract the breast region. Then, a blanket method finds and locates the suspicious areas of possible MCCs clusters. The MCCs detection module is developed to automatically extract the MCCs from the ROIs. Among the image processing that are involved in this module are gradient enhancement, contrast enhancement and Gaussian filters. The segmentation of MCCs from the background is done using entropy-based thresholding. Shape cognitron which is based on a neural network-like shape recognition systems is introduced as a classification technique of MCCs. The systems achieved as high as 95% classification rate with 93% detection rate. Table 1 illustrates the comparison among the techniques that have been employed in each stage of CAD/CADx.

Table 1: An analysis features of CAD/ CADx

Features	References							
	[21]	[12]	[2]	[11]	[4]	[6]	[3]	[14]
Breast Region Extraction	Iterative thresholdi-ng ; Local mean square deviation; zonal Hough transform	Crop operation in image processing	Breast skin-line localization	Breast skin-line localization	Not applicable	Not applicable	Not applicable	Block region growing; K-means clustering based thresholding method
Denoise Techiques	Not applicable	Not applicable	Not applicable	Not applicable	The seed growing algorithm	Not applicable	Not applicable	Gaussian filter
ROIs Detection	Discrete wavelet transform	Not applicable	Morphologic al Descriptors	Annonated by radiologists	Thresholding technique	Annonated by radiologists	Fuzzy detection algorithm	Blanket method
Contrast Enhancement	Intensity - remapping method	Global gray level thresholding; Histogram Equalization	Contrast enhancement filter	Wavelet based filter	Wavelet multiresoluti on analysis	Not applicable	Not applicable	Gradient enhancement
Segmentation	Filling dilation	Not applicable	Not applicable	Not applicable	The seed growing technique	Not applicable	Not applicable	Entropy-based thresholding
Feature Extraction	Morpholog-ical features	Wavelet decomposit-ion process	Discriminat-ive morphologi-cal and textural features	Auto associative Neural Network	Morphologi-cal features	Independent component analysis	Individual MCCs features; Shape features	Not applicable
Classification	MLP	ANFIS	Rule based system; NN; SVM	Feed forward Neural Network	Geometrical and Cluster classification	Radial basis function	BPNN	Shape Cognitron
Class	Benign; Malignant	Benign; Malignant	Benign; Malignant	Benign; Malignant	Benign; Malignant	Benign; malignant	Benign; Malignant	BIRADS
Evaluation	Classificati -on rate result	Classificati-on rate result	ROC analysis	Classificati-on rate result	Classificati-on rate result	Classificati-on rate result	Classificati-on rate result	ROC analysis
Result	96.9 % of classificati-on rate; 0.2 false positive per image	87.5 % of classification rate	Nijmegen SVM – A_z = 0.79 (original feature set); 0.77 (enhanced feature set) NN – A_z = 0.70(original feature set); 0.76(enhanc ed feature set)	82.2 % of classification rate; 4.15 false positive per image	82.3% for geometrical classification and 73.7% for cluster classification .	ICA has the highest recognition rates, with 79.31% compared to 72.41% for SGLD and 70.68% for GLHM respectively.	88.9% of classification rate	95% classification rate; 93% detection rate, 0% false alarm rate

5. Conclusions

This paper discusses the CAM systems that have been proposed and developed for MCCs detection and diagnosis for early breast cancer prognosis. Basically, CAM consists of two important sub systems, i.e. CAD and CADx. Since the scope of this review paper is based on mammography-based approach, only CAD and CAD/CADx are discussed.

There are three main stages in CAM, namely preprocessing, MCCs detection and MCCs classification (diagnosis). In this review paper, every technique that has been employed in each stage of CAD and CAD/CADx is explained and including the techniques that has been used to measure the performance of proposed systems. For CAD, image processing is a technique mainly implemented to interpret the mammographic images, while neural network is popular technique for classification of MCCs cluster among researchers. ROC and FROC analysis are standard methodologies for measurement of performance of detection and diagnosis algorithms in CAM. Currently, many researchers evaluate their system's performance using these evaluation methodologies.

The vast amount or research related to analysis of mammography, as well as widespread interest from the medical community stimulates the development of commercial CAD systems [7]. To date, there are three commercial systems of CAD that are successfully developed and get the approval from the US Food and Drug Administration (FDA), i.e. ImageChecker [18], MammoReader[19], and Second Look[20]. Although there are many outstanding performances have been achieved by CAM systems, the challenges and future directions of research are still remaining. [5] have listed some suggestions on how to improve the performance of CAM in future.

References

[1] A. Papadopoulos, D. I. Fotiadis, and A. Likas, "An Automatic Microcacalcification Detection System Based On a Hybrid Neural Network Classifier," *Artificial Intelligence in Medicine*, vol. 25, pp. 149-167, 2002.

[2] A. Papadopoulos, D. I. Fotiadis, and A. Likas, "Characterization of Clustered Microcalcifications in Digitized Mammograms Using Neural Networks and Support Vector Machine," *Artificial Intelligence in Medicine,* vol. 34, pp. 141-150, 2005.

[3] B. Verma, and J. Zakos, "Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy Neural and Feature Extraction Techniques," *IEEE Transactions on Information Technology in Biomedicine*, vol. 5(1), pp. 46-54, 2001.

[4] F. Ballesteros, A. Oropesa, L. Martin, and D. Andida, "Mammography Classification Using Wavelets," *Proceedings of 5th Biannual World Automation Congress,* vol. 13, pp. 293-300, 2002.

[5] H. D. Cheng, and M. Cui, "Mass Lesion Detection with a Fuzzy Neural Network," *Pattern Recognition*, vol. 37, pp. 1189-1200, 2004.

[6] I. Christoyianni, A. Koutras, E. Dermatas, and G. Kokkonakis, "Computer Aided Diagnosis of Breast Cancer in Digitized Mammograms," *Computerized Medical Imaging and Graphics*, vol. 26, pp 309-329, 2002.

[7] J. Y. Lo, M. Gavrielides, M. K. Markey, and J. L. Jesneck, "Computer-Aided Classification of Breast Microcalcification Clusters: Merging of Features From Image Processing and Radiologists," *Medical Imaging 2003: Image Processing,* vol. 5032, pp. 882-889, 2003.

[8] M. L. Giger, "Computer-aided Diagnosis in Medical Imaging – A New Era in Image Interpretation," *Medical Imaging Ultrasound*, pp. 75-79, 2000.

[9] N. A. Mat Isa, S. Sabaruddin, U. K. Ngah, K.Z. Zamli, and M. Mohd Noor, "The Potential Use of Modified Seed-based Region Growing Technique for Automatic Detection of Breast Microcalcifications and Tumour Areas," *Jurnal Teknologi Siri D*, vol 44, Jun 2006.

[10] National Cancer Registry, "Second Report Of The National Cancer Registry - Cancer Incidence In Malaysia 2003", 2006.

[11] P. Delogu, M.E. Fantacci, A.P. Martinez, A. Retico, A. Stefanini, and A. Tata, "A Scalable System for Microcalcification Cluster Automated Detection in a Distributed Mammographic Database," *IEEE Nuclear Science Symposium Conference Record*, pp. 1530-1534, 2005.

[12] R. Mousa, Q. Munib, and A. Moussa, "Breast Cancer Diagnosis System based on Wavelet Analysis and Fuzzy-Neural," Expert *Systems with Applications*, vol. 28, pp. 713-723, 2005.

[13] RadiologyInfo. "Mammography," 2006.

[14] S. K. Lee, C-S. Lo, C-M. Wang, and P-C. Chung, "A Computer-Aided Design Mammography Screening System for

Detection and Classification of Microcalcifications.” *International Journal of Medical Informatics*, vol. 60, pp. 29-57, 2000.

[15] S. Yu, and L. Guan, “A CAD System for the Automatic Detection of Clustered Microcalcifications in Digitized Mammogram Films,” *IEEE Transactions on Medical Imaging*, vol. 19, pp. 115-126, February 2000.

[16] T. Arodz, M. Kurdziel, T. J. Popiela, , E. O. D. Sevre, and D. A. Yuen, “Detection of Clustered Microcalcifications in Small Field Digital Mammography,” *Computer Methods and Programs in Biomedicine*, vol. 81, pp. 56-65, 2006.

[17] U. Bottigli, P. Cerello, P. Delogu, M.E. Fantacci, F. Fauci, and G. Forni, “A Computer Aided Detection Systems for Mammographic Images Implemented on a GRID Infrastructure,” 2003.

[18] U. S. Food and Drug Administration, “Summary and Safety and Effectiveness Data: R2 Technologies”, *P970058*, 1998.

[19] U. S. Food and Drug Administration, “Summary and Safety and Effectiveness Data: ISSI”, P010038, 2002..

[20] U. S. Food and Drug Administration, “Summary and Safety and Effectiveness Data: CADx Medical Systems”, *P010034*, 2002.

[21] W. Xu, S. Xia, and H. Duan, “A Novel Computer Aided Diagnosis System of the Mammogram,” *Springer-Verlag*, pp. 639-644, 2006.

Biography

[1]**Norlia Md Yusof** received BSc. in computer science from Universiti Teknologi Malaysia and MSc. in Intelligent Systems from Universiti Utara Malaysia. Currently, she is doing her PhD in medical imaging and artificial intelligence which focusing in breast cancer.

[2]**Nor Ashidi Mat Isa** obtained his BEng (Hons) in Electrical Engineering from Universiti Sains Malaysia in 2000 and PhD in Image Processing and Neural Networks from the same university in 2003. He is currently lecturing at the School of Electrical and Electronics Engineering, USM Engineering Campus in Transkrian. He specialises in the area of image processing, neural networks for medical applications and software engineering.

[3]**Harsa Amylia Mat Sakim** was born in Kuala Lumpur, Malaysia, in November 1971. She received her B.Eng degree from Liverpool University, United Kingdom in July, 1994. She subsequently undertook research at Newcastle University, United Kingdom whom honoured her with an M.Sc degree in 1996. She later pursed her Ph.D degree in breast cancer diagnosis and prognosis based on neural network techniques until February 1998. Subsequently, she pursued her research on similar topics at Universiti Sains Malaysia, whom then honoured her with Ph.D degree. She has published a number of journal and conference paper and her main interests are in biomedical engineering and cancer research specifically in disease management. Currently, she is a lecturer at School of Electrical & Electronic Engineering, Universiti Sains Malaysia.