

A BI-RADS Based Expert Systems for the Diagnoses of Breast Diseases

¹Umi Kalthum Ngah, ¹Shalihatun Azlin Aziz, ²Mohd. Ezane Aziz, ²Mazeda Murad,
²Nik Munirah Nik Mahdi, ²Ali Yeon Mat Shakaff, ¹Nor Ashidi Mat Isa,
³Mohd. Yusoff Mashor, ¹Mohd. Rizal Arshad

¹School of Electrical & Electronic Engineering, Universiti Sains Malaysia(USM),
Engineering Campus, 14300 Nibong Tebal, Penang, Malaysia

²Radiology Department, Hospital Universiti Sains Malaysia(HUSM)
Kubang Kerian, Kelantan, Malaysia

³School of Computer & Communication Engineering, Universiti Malaysia Perlis,
Jalan Kangar-Arau, Jejawi, Perlis, Malaysia

Abstract: We proposed an expert system based on the interpretation of mammographic and ultrasound images that may be used by expert and non-expert doctors in the interpretation and classifying of patient cases. The expert system software consists of a mammographic(MAMMEX) and breast ultrasound(SOUNDEX) medical expert systems which may be used to deduce cases according to the Breast Imaging Recording and Data System (BI-RADS) based upon patients' history, physical and clinical assessment as well as mammograms and breast ultrasound images. The systems were tested on a total of 179 retrospective cases from the Radiology Department, Hospital Universiti Sains Malaysia (HUSM), Kubang Kerian, Kelantan. The accuracy, sensitivity and specificity of MAMMEX were 97%, 96% and 92% and values of 99%, 98% and 100% were found for SOUNDEX respectively.

Key Words: Expert Systems, Knowledge-Based Systems, Mammograph, Ultrasound, Breast Cancer

INTRODUCTION

Breast cancer is the most common form of malignant disease amongst women. However, mortality rates fell noticeably especially in the United Kingdom partly because of the widespread practice of breast screening^[1]. For this, the usage of ultrasound in conjunction with physical and mammography examination has been propagated in order to obtain a thorough assessment in breast screening.

Breast screening is not without its problems. The implementation of mass screening would result in increased caseloads for radiologists which would in turn give rise to chances of improper diagnosis. Diagnosticians with the training and experience to interpret mammographic images and breast ultrasounds are scarce. This is further aggravated by the requirement of having two radiologists reading a case in certain practices. Mammography reading is a very hard skill to teach, requires years of experience and frequent scrutinizing^[2]. Radiologists training for mammography

traditionally involve viewing large numbers of films and they need to maintain a high throughput (approximately 7000 cases per year) in order to perform well in reading and interpreting the mammograms^[2]. Sensitivity and specificity are very crucial in clinical practice as only 15-30% of patient referred for biopsy are found to have a malignancy. Unnecessary biopsies increase health care costs and may cause patient anxiety and morbidity. It is therefore important to improve the accuracy of interpreting mammographic lesions^[3], thereby improving the positive predictive values of detection modalities.

Routine and repetitive use of computer-based systems developed for experiments would bring several benefits. Radiologists could be trained to evaluate the perceptual features appropriately^[4]. The existence of an expert system would make diagnostic expertise more widely and readily available in the clinical community. The availability of this system would facilitate computer aided study and learning. This system would also prove to be useful in the training of radiologists in

Corresponding author: Umi Kalthum Ngah, School of Electrical & Electronic Engineering, Universiti Sains Malaysia (USM), Engineering Campus, 14300 Nibong Tebal, Penang, Malaysia. Tel: +60-4-5995999 extn. 6022, fax: +60-4-5941023

the early part of their career. The archiving of knowledge gathered in this area with patient cases would also promote the interpretation of images in a more consistent and standard manner and may be referred to from time to time.

The lack of standardized description and categorization of breast assessment of patients^[5,6,7] led to the realization of an urgent need for the development of a certain standard form of guideline or system. In view of this, the American College of Radiology called upon a task force on breast cancer in the late 1980s and appointed a committee to develop guidelines for standardized reporting which was initially used for mammographic findings. This was published as the “Breast Imaging Reporting and Data System”, referred to as BI-RADS^[8,9] which was intended to standardize the terminology in reporting starting with the mammogram report. BI-RADS was not in routine use until the year 1997^[10]. The standardized assessment categories used to describe findings on mammography initially and then extended into the other modalities. On the basis of the level of suspicion, detected lesions or abnormalities can be placed into one of the BI-RADS assessment categories.

Even though BI-RADS was introduced to help standardize feature analysis and final management of breast modality findings, there still exists variations in their interpretations. Continued efforts to educate radiologists to promote maximum consistency still need to be done^[11].

The earliest study encountered was by Cook & Fox^[12], where mammographic image analysis was investigated using a decision table to represent all the parameters and possibilities in 41 rules that were created, all centred upon masses and lesions. The other related works were mostly based on Artificial Neural Networks(ANN) for decision making in the diagnoses of breast cancer and some of the work are related to breast biopsy decisions^[13,14,15,16,17]. A study was also carried out by Floyd *et al.*^[18] using case-based reasoning but none quite fits exactly in what this study is intended to achieve.

The Proposed Method: It usually takes five to ten person years to build even a moderate expert system and the most crucial stage in the technology of expert systems is the process of knowledge acquisition^[19] as it involves efforts dedicated to the identification of the facts that comprise the knowledge base. It is often very difficult for clinicians and health care providers to sketch systematically on paper, their knowledge base and/or algorithms that they use in diagnosis and/or treatments.

Certain guidelines may be adopted that proved to ease the whole process. The three steps in the Knowledge Acquisition Process are shown in Figure 1, processes involving acquiring explanations from the experts, actually capturing the knowledge and organizing the knowledge. The capturing stage is the process of documenting the objects, relations and actions that make up the knowledge. On the other hand, the organization stage is the process of ordering the knowledge in such a form that it would be ready to be mapped into the knowledge base being developed.

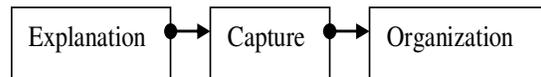


Fig 1: The three steps in the Knowledge Acquisition Process

Knowledge Explanation includes interviews, the interview environment, the do’s and the dont’s that need to be abided by the interviewer as well as obtaining knowledge existing in written forms and capture.

In addition to this, knowledge acquisition through a human expert is a delicate task that needs to be well thought out carefully and deliberately conducted. Also, it most typically lacks an organizational format to guide the activity and it has to undergo the processes as in the following Figure 2 in order to arrive at the domain definition.

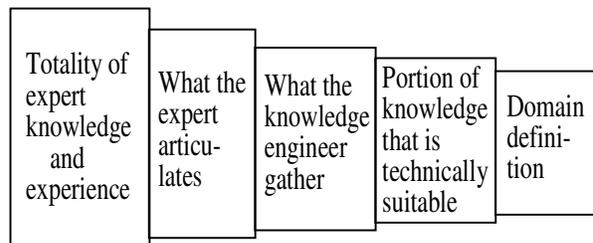


Fig. 2: The processes in acquiring knowledge from an expert.

As the development of the knowledge base is the most important task that the knowledge engineer performs, a stringent and diligent process needed to be employed to produce a systematic, thoughtful procedure in the knowledge-base construction. In the development of the expert system in this study, the process of knowledge acquisition and knowledge representation proceeded virtually hand in hand as this was absolutely vital for the integrity of the end result.

The complete knowledge base or expert system was then gradually developed in an incremental manner. The system development was based on production rules and therefore, decision tables were considered. The formation of the rules were based on the different modalities with their associated features.

The knowledge base or expert system developed for this work is divided into two parts, namely MAMMEX for the Mammogram Expert System and SOUNDEX for the Ultrasound Expert System. It is envisaged that these two experts systems will produce results which are at par and consistent with those generated by the primary domain expert. In other words, each time MAMMEX and SOUNDEX were consulted, they were expected to provide the same advice as an expert, which in this case, is the category of BI-RADS that the radiologist usually classifies at the end of the assessment for each case. This established a benchmark against which MAMMEX and SOUNDEX will be tested.

Amongst the several criteria that were considered in the initial stage of development of MAMMEX and SOUNDEX included choosing the most suitable language, the working environment i.e. the software should be able to be stored and run on a portable computer in a Windows environment, be simple enough so much so that an end-user would be able to learn to run them, should have the ability to exhibit high performance in terms of speed and reliability in order to be a useful tool, able to propose correct and consistent solutions in a reasonable amount of time and the screen definition language is powerful enough to customize the way the questions are asked and have the ability to allow interface calls to other external routines or programs. After several deliberations and careful considerations including the usage of established knowledge based shells, it was decided that the Builder C++ language environment be chosen as the medium for the implementation.

Developing a Prototype: The process of interpreting mammograms is in fact a multifaceted medical decision-making task^[20] as there is a constellation of characteristics, a plethora of features to be considered before a certain conclusion or decision can be made. As such, the diagnoses of breast diseases requires a more thorough investigation of all possibilities and procedures. Moreover, there is a poorly structured collection of many isolated facts and it is unclear what kinds of distinctions between the facts are the important ones. It was necessary to solve the possibilities by including heuristic or appropriate methods which did not require perfect data and the solutions derived by the

system may be proposed with varying degrees of certainties. Also, it was important to obtain explanations that infer how the expert system arrived at the answer and justifications for the knowledge itself. Therefore the use of rules or assertions was preferred to represent the knowledge.

The creation of the rule base proceeded from discussions with practicing clinicians and radiologists and through the extraction of rules from journals and texts on established practices for patients who present themselves in for assessment and complaints.

When the system was run, questions pertaining to the patient history, clinical and physical assessment, mammographic features and eventually, ultrasound features will be displayed on the screen in a windows environment. This was how data was obtained or needed as input to arrive at a decision.

The type of question asked was multiple choice. A question will display a statement ending in a verb, followed by a numbered list of possible choices to complete the sentence. The user will be requested to enter the number of the correct choice for the situation by a click on the mouse. Questions will continue to be displayed one after another depending on the path designated by the choice of answers. To illustrate this point, for example, if the user finds that there is no presence of mass on mammogram assessment, then the subsequent questions pertaining to mammogram mass will be skipped and questions pertaining to the next matter which involves another feature, calcifications (for example) will then have to be dealt with.

The user will be required to answer all the questions that were displayed by the system. At the end, the system will provide the user with a conclusion listing the categories of the BI-RADS for the particular modality. The highest numeric value associated with the particular category will be taken as the answer returned by the system.

The consultation is essentially a search through a tree of goals. The top goal at the root of the tree is the action part of the goal rule, i.e. the suspicious level returned or the diagnosis of the disease. Subgoals further down the tree include determining the other features involved and seeing if these are significant. Many of these subgoals have sub-subgoals of their own, such as mammographic features for example, the presence of mass and its details.

The special kind of structure called the tree is very useful for representing the way in which goals can be expanded into subgoals by a program. The basic idea is that the root node of the tree represents the main goals, the terminal nodes represent primitive actions that can

be carried out and the non-terminal nodes represent subgoals that are susceptible to further analysis.

The search strategy implied and the manner in which the rules were executed may be described in more technical terms as a Forward Chaining Search and Inference Technique with pruning, a natural way to design expert systems for analysis and interpretation. That is, the process begins with a certain data concerning the category that is most likely; for example, its mass features (if any), calcification features (if any) and so on. These data, along with the constraints, serve to highlight the potential alternatives and to decimate the unlikely ones. This is consistent with the way a domain expert reacts when confronted with patients' cases when they arrive at the hospital for check-up. The expert first needs to gather some information and then tries to infer from it whatever can be inferred_[19]. Thus the search ultimately arrive at a listing from which a final selection is made based on the highest score. The pruning process results in a reduction in search requirements. The actual premises for each of the different modalities are listed in Tables 1, 2, 3 and 4.

Table 1 lists the premises used in the classifier system for patient history, Table 2 shows the premises used in the physical and clinical assessment, Table 3 is concerned with the premises used in the mammographic assessment and lastly Table 4, the premises for the ultrasound assessment. The premises for each of the sections underwent numerous changes. Detailed discussions were held from time to time which entails numerous and endless trips to HUSM which were deemed necessary to ensure that the work progresses in an acceptable frame of reference. Finally, the above premises were obtained with the consensus of the radiologists.

An attempt had also been made to incorporate patients' images to be included in the expert systems developed to allow for image manipulations and processing.

The Quest For Information - Gathering Of Facts, Figures And Building The Decision Table: Facts and information had to be gathered in order to facilitate enough knowledge to be incorporated in the expert system. Based upon the various premises listed previously, work then began in developing the framework of the expert system.

Table 1: The premises used in the classifier system for a patient's history

The patient's age is - (20-29, 30-39, 40-49, 50-59, greater than 60, not available)
Patient's age of menarche - (9-10, 11-12, 13-14, 15-16, greater than 17, not available)
Patient has menopausal - (yes, no, not available)
If patient is menopausal than age of onset - (30-39, 40-49, 50-59, not available)
Patient's marital status - (married, single, divorced)
Number of children that patient has is (none, less than 5, more than 5, not available)
Patient has history of breast feeding - (yes, no, uncertain)
Patient lactating or nursing recently - (yes, no, not known)
Patient has other family members with history of cancer i.e. ovarian, prostate cancer - (yes, no, uncertain)
Patient has previous abnormal breast biopsy i.e. atypical ductal hyperplasia - (yes, no, uncertain)
Patient has previous history of breast trauma - (Yes, no, uncertain)
Patient has previous breast surgery/implant-(yes, no, uncertain)
If patient has had history of breast surgery, then surgery is - (mastectomy, lumpectomy, others)
Patient is on hormone replacement therapy('HRT')(yes, no, not available))
Patient had an oophorectomy or TAHBSO done - (yes, no, not available))
Patient is on contraceptive pills - (yes, no, not available)
Patient has history of extramammary malignancy (yes, no, not available)

For each of the modality and its various features, information digging and fact-finding had to be endeavoured. To determine reasonable numerical values associated with each and every factor making up the sections in MAMMEX and SOUNDEX, numerous papers were mined. In other words, each and every piece of information relating to the main modalities had to be investigated and gathered. From here, a more reliable numerical value would be found and thus be used in the eventual knowledge based system to be developed.

Table 2: The premises used in the classifier system for the physical and clinical assessment of a patient

The lump is palpable - (yes, vague lump, no, uncertain)
The nature of the lump is - (firm, hard, soft, uncertain)
The lump is mobile - (yes, no, mobility is uncertain)
The location of the lump is - (upper outer quadrant, upper inner quadrant, lower outer quadrant, lower inner quadrant, retroareolar, inner middle, outer middle, upper middle, lower middle, uncertain)
There is nipple retraction - (yes, no uncertain)
Presence of nipple discharge - (yes, no, uncertain)
If there is nipple discharge then discharge is - (from one breast and one orifice, not from one breast one orifice, uncertain from one breast, one orifice)
If there is nipple discharge, then discharge is - (white, yellowish, greenish-grey, clear serous, bloody, uncertain)
Discharge is with pus - (yes, no, not available)
Mass is present in axilla - (yes, no, uncertain)
There is breast tenderness/pain - (yes, no not available)
If there is breast tenderness/pain then it is - (localized, diffused, uncertain)
There is/are skin changes - (yes, no, not available)
If there is skin change then skin change is (retraction/puckering, skin oedema, ulcer, others)

Table 3: The premises used in the classifier system pertaining to the mammographic features

Mass/Masses present is - (positive, negative, uncertain)
The location of the mass is - (upper outer quadrant, upper inner quadrant, lower outer quadrant, lower inner quadrant, retroareolar, inner middle, outer middle, upper middle, lower middle, uncertain location)
The margin of the mass is (well defined, sharp halo, microlobulated, macrolobulated, ill-defined, irregular, obscured, uncertain)
The shape of the mass is (round, oval, irregular, stellate, uncertain)
The size is (less than 1.0cm, equal o 1.0 cm, greater than 1.0 cm, uncertain)
The density of the mass - (fat density, low density, isodense, high, has central lucency)
The mass are (multicentered, multifocused, multicentred/multifocused, uncertain)
If the mass is multicentred or multifocused, then they are also bilateral(yes, no)
There is no architectural distortion(yes, no, uncertain)
There is skin thickening(yes, no, uncertain)
There is nipple retraction/abnormality(yes, no, uncertain)
Calcifications are present(yes, no, uncertain)
The calcification is (micro, macro, mixed(macro,macro), uncertain)
The morphology of calcifications are (lucent-centered, parallel tracks/linear tubular, coarse/popcom like, large rod-like, round, eggshell/rim, milk of calcium, suture calcifications, dystrophic, punctate, amorphous/indistinct, granular sand-like, pleomorphic/heterogeneous/granular, fine linear/ fine linear branching/casting)
The calcification distribution is (grouped/cluster, linear, segmental, regional, diffused/scattered)
The number of calcifications per cubic cm is (1, less than 5, greater than 5, uncertain)
There is presence of node in axilla(yes, no, uncertain)
There are multiple nodes(yes, no, uncertain)
The shape of node in axilla is (round, ovoid/ellipsoid, bean-shaped, slightly lobulated, spiculated, uncertain)
The margins of node is well-circumscribed(yes, no, uncertain)
Nodes are bilateral(yes, no, uncertain)
Size of node is(less than 2.0 cm, more than 2.0 cm, is uncertain)
Node has central lucency(yes, no, uncertain)

Table 4: The premises used in the classifier system pertaining to the ultrasound features

The mass is detected on the ultrasound image (yes, no, uncertain)
Location of breast mass is (on the upper outer quadrant, upper inner quadrant, lower outer quadrant, lower inner quadrant, retroareolar, inner middle, outer middle, upper middle, lower middle, uncertain)
The shape of the mass is (round, ovoid/ellipsoid, irregular, lobular, spiculated, uncertain)
The orientation of axis of the mass is (taller than wide, wider than tall, is almost equal, uncertain)
Overall mass margin is (smooth/well-circumscribed, gentle lobulations, radial/ductal extension, branch pattern, angular margin, uncertain)
The number of lobulations are (less than 3, greater than 3, uncertain)
Echo pattern of mass is (anechoic, hypoechoic, hyperechoic, isoechoic, mixed, uncertain)
Posterior to the mass, (there is acoustic enhancement, normal/no enhancement/shadowing, complete shadowing, uncertain)
The lesion has calcifications(yes, no, uncertain)

Collating Past Works And The Use Of Decision Tables: To illustrate the whole process, it was certainly helpful that a table be developed prior to actually

embarking on the fact finding. This formed the decision tables to accommodate the various previous works that were based on each and every characteristic of the overall processes involved in breast assessment. For example, consider the case for the mammographic feature assessment. The presence of mass entails the search for facts related to the mass margin for example. Then, these premises will be listed in the left hand side of the table, making up the rows. The top headers of the table will be in the form of columns whereby resources and previous studies found to support evidences on the characteristics mentioned, will be recorded.

For each of the study that was gathered, The Positive Predictive Values (PPVs) for the associated benign and malignant features mentioned in the resources for the associated premises were entered accordingly in the appropriate cells of the decision table. Some papers focused on benign features only; some deal with mainly malignant features while some studies which scrutinize on a much broader basis and encompassed the benign as well as the malignant features.

After 'exhaustive' searching from previous studies, the mean of the positive predictive values were then calculated for the benign and malignant values. As such, values pertaining to the benign and malignant cases emerge. These were taken as to represent the possible range of certainty values for the benign and malignant values each differing for the different features for the different modalities. An example of this is depicted transparently in Table 5 which shows a sub-section of the eventual whole decision table.

The entire decision table represents a method for visualizing the large number of possible situations in a single table. Rules can then be created directly from the decision table. From the decision tables whereby the knowledge base or expert system may begin to be constructed and developed, certainty values may be formulated and built based upon the fact findings from the collection of referred papers on the various parameters. Work then proceeded in the direction of determining the framework of the knowledge-based system i.e. constructing the backbone of the entire system in the set of decision tables.

RESULTS AND DISCUSSIONS

Data of patients were obtained from the Radiology Department of HUSM, Kubang Kerian, Kelantan. A total of 179 cases (mammographic cases with corresponding ultrasounds) were gathered i.e. cases spanning the years 2002 until the year 2005, each with its respective patient history, clinical and physical assessment particulars, mammographic images and

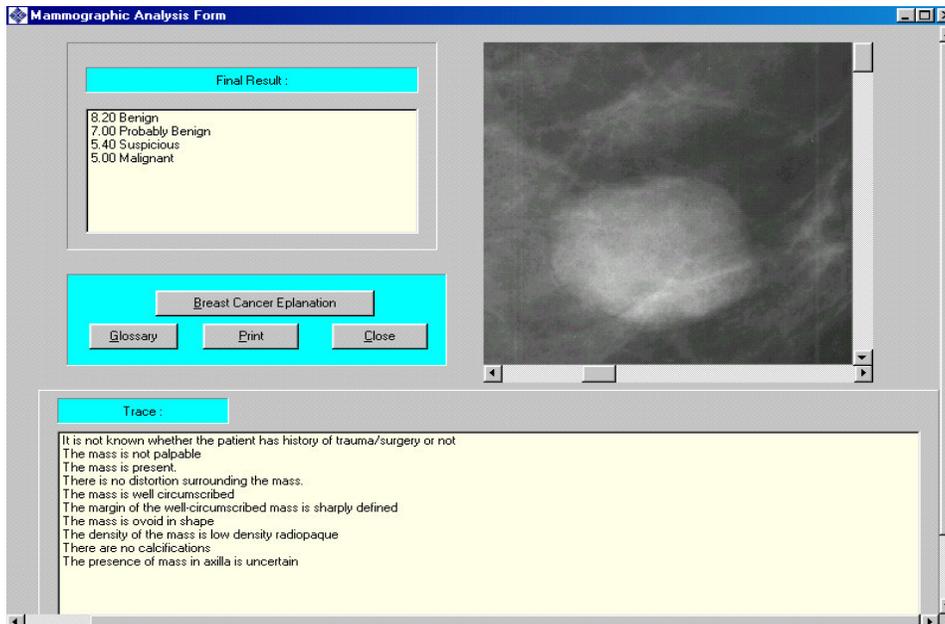


Fig. 3: The screen display of mammographic image based expert system.

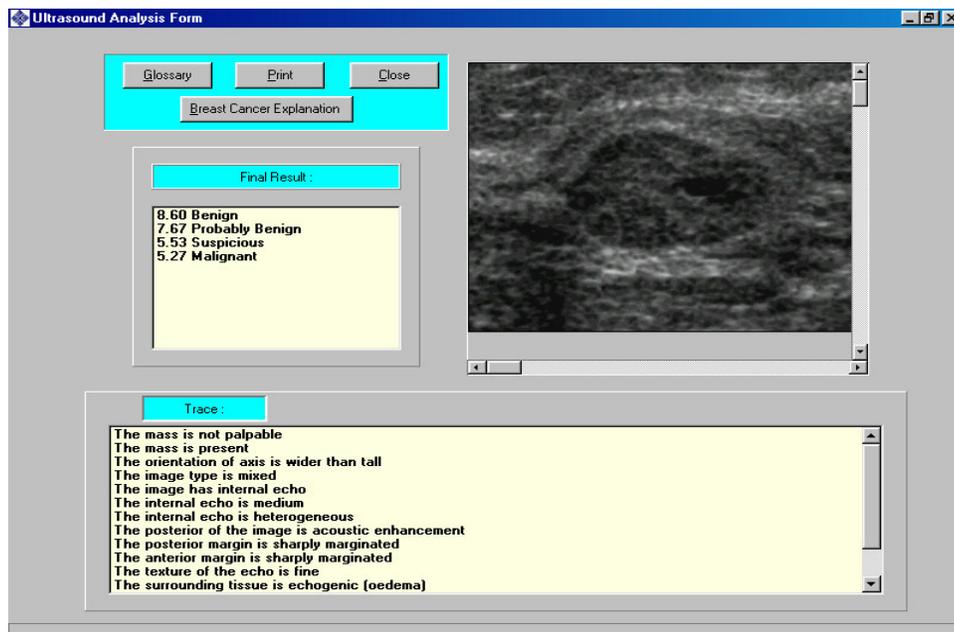


Fig. 4: The screen display of the breast ultrasound image based expert system.

Table 5: A subset of the decision table

(Benign/Malignant) Mmgr. Features	Dronkers <i>et al.</i> , (2002)	Hong <i>et al.</i> , (2005)	Average Benign (1-10)	Average Malignant (1-10)	Liberman <i>et al.</i> , (1998)	Hong <i>et al.</i> , (2005)
Mass margin:		
Well defined	9		10	8.3	0.39	0.9		0
Sharp halo	9			9.32	0.14			
Microlobulated		...	5	2.67	5.65	1.7	...	5
Macrolobulated				6.6	1.5			
Ill-defined	1		4	2.83	6.77	4.4		6
Irregular	1	...		1.00	6.5		...	
Spiculated			0.3	0.6	8.82	8.1		9.7
Obscured		...	8.4	8.4	4.6	3.3	...	1.6

their corresponding ultrasound images. Prior to these years, the database of patient images and cases in DICOM format were non-existent.

Table 6: The Two-by-Two Table

	Disease present D+	Disease absent D-
Test positive(T+)	True positives (TP)	False positives (FP)
Test negative(T-)	False negative (FN)	True negatives (TN)

Table 7: Evaluation of the accuracy of test T

Test T	Malign	Benign	Totals
Positive	65	43	108
Negative	9	126	135
Total	74	169	243

Table 8: The results obtained after executing MAMMEX and SOUNDEX on the set of cases

	MAMMEX	SOUNDEX
True Positive(TN)	127(from total of 130)	113(from total of 115)
True Negative(TP)	48(from total of 52)	64(from total of 64)
False Positive(FP)	2.7%	1.2%
False Negative(FN)	1.6%	0
Accuracy	97%	99%
Sensitivity	96%	98%
Specificity	92%	100%

Out of the 179 cases, radiologists have classified 52 cases as normal cases (BI-RADS 1), 73 as benign (BI-RADS 2), 22 as probably benign (BI-RADS 3), 24 as suspicious (BI-RADS 4) and 8 cases as malignant (BI-RADS 5) categories according to the mammographic features assessment. Radiologists also have classified 63 cases as normal (BI-RADS 1), 90 as benign (BI-RADS 2), 12 as probably benign (BI-RADS 3), 8 as

suspicious (BI-RADS 4) and 6 as malignant (BI-RADS 5) categories accordingly for the ultrasound features assessment. The screen displays of MAMMEX and SOUNDEX are shown as in figures 3 and 4.

The simple two-by-two table is one of the most intuitive methods for the analysis of diagnostic examinations^[21]. Despite this, the method is capable of displaying strength and power in illuminating understanding the performance and analysis of diagnostic examinations. The basic idea of a diagnostic test interpretation is to calculate the probability a patient has a disease under consideration given a certain test result. For this, a Two-by-Two table is used as a mnemonic device^[22]. The table is labelled with the test results on the left side and the disease status on top as shown in Figure 6.

Table 7 shows a fictitious data from an experiment to evaluate the accuracy of a certain test T for a certain set of patients with clinical suspicions. The data are numbers of women with malignant or benign breast tumours. Referring to the table, the sensitivity and specificity of the two-by-two table may be illustrated. By using the numbers in the 'Malignant' column, the sensitivity of a fictitious test T in the sample of women is approximately 88% (65 out of 74) and specificity of test T is calculated as 75% (126 negative results out of 169 women with benign lesions).

The Modified Two-By-Two Table: The categorization for cases in certain studies is not as straightforward as in Table 7. This is because measures of accuracy: sensitivity, specificity require a positivity threshold for classifying the test results as either positive or negative^[23,24]. In mammography and breast ultrasound, the BI-RADS scoring system was used to classify the modalities as normal, benign, probably benign, suspicious and malignant findings (i.e. BI-RADS B1, B2, B3, B4 and B5 respectively). Therefore, it was suggested that certain modifications be made based on the categories^[25], called the Modified Two-by-Two Table Analysis.

Assuming there are N cases altogether and cases have been classified by independent evaluators A as the

number of normal or BI-RADS B1 cases, B is the number of cases in the BI-RADS B2 cases, C is the number of cases in the BI-RADS B3 cases, D is the number of cases in the BI-RADS B4 cases and E is the number of cases in the BI-RADS B5 cases.

We may also blanket the positivity threshold of B2 and B3 cases together (noted as BI-RADS B₂₃ cases) and B4 with B5 (noted as BI-RADS B₄₅ cases). So, we may clump together the total number of cases for BI-RADS B₂₃ cases and BI-RADS B₄₅ cases i.e. (B + C) and (D + E) respectively. Out of these BI-RADS B₂₃ and BI-RADS B₄₅ cases, we can denote F_{TP} as True Positive (TP) in the BI-RADS B₂₃ category while G_{TP} denotes the TP cases in the BI-RADS B₄₅ category, making the number of collective TP cases evaluated by the test T in question to be the sum of F_{TP} and G_{TP} altogether.

Assuming that out of A, A_{TN} are classified by the test T in question as True Negative. Since we have made the modifications to the actual classifying, cases that are found to be normal for the category B₂₃ need to be summed up with cases found to be normal (if any) that fall under the B₄₅ category and those that are found to be B2 or B3 under that of the B₄₅ category.

This may be illustrated as in the equation below:

$$FN = \frac{B_{23}(bTf - B1) + (B_{45}(bTf - B1) + B_{45}(bTf - B_{23}))}{B + C + D + E} \quad (1)$$

where B₂₃(bTf - B1) means the category "Doctors findings are B2 and B3 but test T found these cases to be B1" and likewise. B + C + D + E is the sum of all cases in categories B2, B3, B4 and B5.

Similarly,

$$FP = \frac{((B1(bTf - B_{23}) + B1(bTf - B_{45})) + B_{23}(bTf - B_{45}))}{A + B + C} \quad (2)$$

where B1(bTf - B₂₃) means "Doctors' findings are category B1 but test T found the case to be B₂₃" and likewise. A+B+C is the sum of all cases in B1, B2 and B3.

Therefore, the accuracy of test T in question would then be:

$$\text{Accuracy} = \frac{A_{TN} + F_{TP} + G_{TP}}{N} \quad (3)$$

The specificity of test T would be:

$$\text{Specificity} = \frac{A_{TN}}{A} \times 100 \% \quad (4)$$

And the sensitivity of test T would then be modified to be:

$$\text{Sensitivity} = \frac{F_{TP} + G_{TP}}{B + C + D + E} \times 100\% \quad (5)$$

The modified calculation of the accuracy, sensitivity and specificity in the Two-by-Two Table as in the above equations were implemented on the results obtained from the execution of MAMMEX and SOUNDEX.

A summary of the accuracy, sensitivity, specificity, True Positive, True Negative, False Positive and False Negatives values for MAMMEX and SOUNDEX are as shown in Table 8.

CONCLUSION

A BI-RADS based mammographic and ultrasound expert system for breast diseases has been successfully developed in this study (MAMMEX and SOUNDEX). The Modified Two by Two Table results indicate that the expert systems developed have high performance and reliability with accuracies of 97% and 99%, sensitivities of 96% and 98% and specificities of 92% and 100% for MAMMEX and SOUNDEX respectively.

ACKNOWLEDGEMENTS

The authors wish to thank Universiti Sains Malaysia(USM) for making this work possible and the Ministry of Science, Technology and Innovations of Malaysia(MOSTI) for providing the necessary grants for this work.

REFERENCES

- Boyle, U. V. P., A. Goldhirsch, R. Orecchia, G. Viale, 2005. Breast Cancer. The Lancet. 365:1727-1741.
- Caulkin S. J., S. M. Astley 2000. Generating Realistic Spiculated Lesions in Digital Mammograms, 5th. International Workshop on Digital Mammography (IWDM), (ed. M. J. Yafe) pp. 713-720, Medical Physics, Wisconsin.
- Hadjiiski L., H. P. Chan, B. Sahiner, M. A. Helvie, M. A. Roubidoux, C. Blane, C. Paramagul, N. Petrick, J. Bailey, K. Klein, M. Foser, S. Paterson, D. Adler, A. Nees, J. Shen, 2004. Improvements in Radiologists' Characterization of Malignant and Benign Breast Masses on Serial Mammograms with Computer-aided Diagnosis: An ROC Study. Radiology. 233: 255-265.
- D'Orsi C. J., D. J. Getty, J. A. Swets, R. M. Pickett, S. E. Seltzer, B. J. McNeil, 1992. Reading and Decision Aids for Improved Standardization of Mammographic Diagnosis. Radiology. 184: 619-622.
- Elmore G., C. K. Wells, C. H. Lee, O. H. Howard, R. Alvan, A. R. Feinstein, 1994. Variability In Radiologists' Interpretations Of Mammograms. The New England Journal of Medicine. (Dec.): 1493-1439.

6. Baker J. A, P. J. Kornguth, M. S. Soo, R. Walsh, P. Mengoni, 1999. Sonography of solid breast lesions: observer variability of lesion description and assessment. *Am J. Roentgenol.* 172(6): 1621-1625.
7. Berg W.A., C. Campassi, P. Langenberg, M. J. Sexton, 2000. Breast Imaging Reporting and Data System: Inter- and Intraobserver Variability in Feature Analysis and Final Assessment.. 174: 1769-1777.
8. American College of Radiology, 1998. Breast imaging reporting and data system (BI-RADS), 3rd ed. Restem.
9. D'Orsi C. J., 1996. The American College of Radiology mammography lexicon: an initial attempt to standardize terminology. *Am. J. of Roentgenol.* 166: 779-780.
10. Orel S. G., N. Kay, C. Reynolds, D. C. Sullivan, 1999. BI-RADS categorization as a predictor of malignancy. *Radiology.* 211(3): 845-850.
11. Lehman C., S. Holt, S. Peacock, E. White, N. Urban, 2002. Use of the American College of Radiology BI-RADS Guidelines by Community Radiologists: Concordance of Assessments and Recommendations Assigned to Screening Mammograms. *Am. J. of Roentgenol.* 179: 15-20.
12. Cook H. M., M. D. Fox, 1987. Artificial Intelligence Applied to Mammographic Image Analysis. *Electronic Imaging.* 1154-1158.
13. Wu Y., M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, 1993. Artificial Neural in Mammography: Application to Decision Making in the Diagnosis of Breast Cancer. *Radiology.* 187: 81-67.
14. Floyd Jr. C. E., J. Y. Lo, A. J. Yun, D. C. Sullivan, P. J. Kornguth, 1994. Prediction of Breast Cancer Malignancy using an Artificial Neural Network. *Cancer.* 74: No. 11.
15. Baker J. A., P. J. Kornguth, J. Y. Lo, M. E. Williford, Jr. C. E. Floyd, 1995. Breast Cancer Prediction with Artificial Neural Network Based on B-RADS Standardized Lexicon. *Radiology.* 196: 817-822.
16. Lo, J. Y., J. A. Baker, , P. J. Kornguth, , J. D. Iglehart, C. E. Floyd, 1999. Predicting Breast Cancer Invasion with Artificial Neural Networks On The Basis of Mammographic Features. *Radiology.* 203: 159-163.
17. Chen D. R., R. F. Chang, Y. L. Huang, 1999. Computer-aided Diagnosis Applied to US of Solid Breast Nodules by Using Neural Networks. *Radiology.* 213: 407-412.
18. Floyd C. E., J. Y. Lo, G. D. Tourassi, 2000. Case-Based Reasoning Computer Algorithm that Uses Mammographic Findings for Breast Biopsy Decisions, *Am. J. of Roentgenol.* 175: 1347-1352.
19. Negnevitsky M., 2005. Artificial Intelligence – A Guide to Intelligent Systems, 2nd. Edition. Addison Wesley. Pg. 12.
20. Yahia M.E., R. Mahmod, N. Sulaiman, F. Ahmad, 2000. Rough Neural Expert System. *Expert System in Applications.* 18: pp. 87-99.
21. Moskowitz M., 1992. Positive Predictive Values for Mammographically Detected Carcinomas, *Letters. Am. J. of Roentgenol.* 158: pg. 688.
22. Langlotz, C. P., 2003. Fundamental Measures of Diagnostic Examination Performance: usefulness for Clinical Decision Making and Research. *Radiology.* 228: pp. 3-9.
23. Basic Test Interpretation, Dxtests, 2006. Available:<http://gim.unmc.edu/dxtests/Reviewof.htm> October.
24. Obuchowski, N., 2004. Fundamentals of Clinical Research for Radiologists. *Am. J. of Roentgenol.* 184: pp. 364-372.
25. Mat Isa, N. A., 2002. Early Diagnosis System for Cervical Cancer Based on Neural Networks. Phd. Thesis. Universiti Sains Malaysia.