A Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy-Neural and Feature Extraction Techniques

Brijesh Verma and John Zakos

Abstract—An intelligent computer-aided diagnosis system can be very helpful for radiologist in detecting and diagnosing microcalcifications' patterns earlier and faster than typical screening programs. In this paper, we present a system based on fuzzy-neural and feature extraction techniques for detecting and diagnosing microcalcifications' patterns in digital mammograms. We have investigated and analyzed a number of feature extraction techniques and found that a combination of three features, such as entropy, standard deviation, and number of pixels, is the best combination to distinguish a benign microcalcification pattern from one that is malignant. A fuzzy technique in conjunction with three features was used to detect a microcalcification pattern and a neural network to classify it into benign/malignant. The system was developed on a Windows platform. It is an easy to use intelligent system that gives the user options to diagnose, detect, enlarge, zoom, and measure distances of areas in digital mammograms.

Index Terms—Breast cancer diagnoser, classification, digital mammograms, fuzzy logic, microcalcification, neural networks.

I. INTRODUCTION

THE interpretation and analysis of medical images is arguably one of the most difficult and advanced applications of pattern recognition and computer vision. In the field of digital mammography, this is mainly because of the difficulty associated with discriminating between benign and malignant microcalcifications (MCCs). Computer-aided diagnosis (CAD) systems can help doctors in the diagnosis of breast cancer and serve as a useful "second opinion."

A. Introduction to Breast Cancer and Digital Mammography

1) Breast Cancer: Each year around the world, millions of women develop new cases of breast cancer. In Australia, approximately 1 in 13 women develop the disease. It is the leading cause of cancer for women between the ages of 35 and 54[1] Researchers are striving to find a cure by investigating lifestyles, environment, diet, and the possible roles of heredity. Until the cause or causes of this serious disease are fully understood, advances in this field will continue to struggle.

[1] NSW Breast Cancer Institute, Sydney, Australia.

It is estimated that current breast screening programs fail to detect approximately 25% of the cancers that are visible on retrospective review [1]. The screening programs involve problems such as high costs, inexperienced radiologists, and visual fatigue. Detection methods are based on clinical examination, mammography (breast X-ray image), ultrasound, and core biopsy. An early detection gives the patient a good chance of survival, whereas a late detection can be fatal and often ends in the death of the patient. A breast cancer once detected can be classified as benign or malignant. However, it is very difficult to distinguish a benign microcalcification from one that is malignant. Consequently, many unnecessary biopsies are often undertaken due to the high false positive rate.

2) Digital Mammogram: A digital mammogram is created when a conventional mammogram is digitized so it can be used by a computer. Digitization can be performed through the use of a specific mammogram digitizer [3] or a camera [4], [5]. 12 b of detection resolution are usually needed to produce a high-resolution digital mammogram without the loss of information from the original mammogram. Generally, most digital mammograms have 4096 gray levels per pixel over the whole area of the mammogram.

3) Digital Mammography: Digital mammography refers to the application of digital system techniques on digital mammograms. Digital systems have the capacity to bring revolutionary advantages to breast cancer detection [6]. Radiologists turn to digital mammography for an alternative diagnostic method due to the problems created by conventional screening programs. An automated system can overcome these problems by reducing the number of false positive and false negative readings from radiologists and increase the chance of detecting abnormalities early. This is a favorable prognosis for patients, as incorrect or late detections often result in mortality.

Currently, digital mammography is one of the most promising cancer control strategies since the cause of breast cancer is still unknown. A radiologist can refer to an automated system for a second opinion, as it is often difficult to distinguish microcalcifications from normal healthy breast tissue. An automated system can detect and diagnose microcalcifications in digital mammograms or point to suspicious regions for further evaluation.

4) Appearance of Microcalcifications in Digital Mammograms: It is very difficult to distinguish a benign microcalcification from an MCC because of the variability associated with the appearances of the cancer. The human breast varies considerably in composition, giving mammographic appearances ranging from relative uniformity to complex patterns of bright

streaks or blobs [7]. Abnormalities can vary in size, and density microcalcifications usually appear as small circular blobs that are usually marginally brighter than the background.

B. Introduction to Computational Intelligence and Digital Mammography

Computational Intelligence (CI) is relatively new term in the computer-science field that aims to mimic biological intelligence. Furthermore, it is a methodology involving computers that can learn and adapt to new situations. CI encompasses artificial neural networks [8], genetic algorithms, and fuzzy logic (FL). CI techniques have been applied to a wide range of fields from medical to engineering applications [9].

CI techniques can be applied to a digital mammogram to assist radiologists in the early diagnosis of breast cancer [10], [17]. In particular, FL techniques have introduced human reasoning to the analysis of digital mammograms [11], [12]. Digital mammograms are often fuzzy and uncertain and very rarely have clearly defined abnormalities.

Researchers in the medical field need to be aware of the medical technologies such as a CAD systems and be prepared to cooperate with scientists so advancements can be made. Attempts have been made to create CAD systems to provide doctors with an alternative method of analysis. The main aim of CAD is to diagnose and detect abnormalities and suspicious areas in digital mammograms by automating the segmentation, extraction, detection, and classification processes.

C. Previous Research in Microcalcification Detection and Classification

In the literature, various numbers of techniques are described to detect and classify the presence of microcalcifications in digital mammograms as benign or malignant: classical image processing techniques [3], [15], wavelet-based techniques [19], [20], statistical techniques [2],[2] [14], [21], neural networks [5], [10], [16], or a combination of techniques [3], [14].

Mascio et al. [3] developed a microcalcification detection algorithm, which operates on digital mammograms by combining morphological image processing with arithmetic processing. The algorithm begins by applying two high-frequency analyses to the original digital mammogram. The first analysis emphasizes any detail in the image that changes sharply in intensity and is larger than several pixels in size. The second analysis emphasizes any detail that is small and textured. Areas that are common to both analyses are segmented and kept for thresholding. This resulted in the detection of microcalcifications and suspicious areas.

Barman et al. [15] used a low-pass filter to detect microcalcification by analyzing digital mammogram. Although the system based on their algorithm is still under development, good preliminary results have been produced with further modifications still to be made.

Wavelet-based techniques are image-processing techniques that can be used in the detection of microcalcifications in digital mammograms. Yoshida et al. [19], [20] developed a system

based on the wavelet transform. In the wavelet transform, all of the wavelets are derived from scaling and the translation of a single function. Yoshida et al. used the least asymmetric Daubechies' wavelets in combination with a difference image technique. These methods are a powerful tool in separating microcalcifications from normal background tissue and achieve a detection rate of approximately 90%.

The k-nearest neighbor (KNN) algorithm is a very simple, but effective method of microcalcification classification. Classification is based on a pattern's similarity to other microcalcifications. A distance is computed from the sample pattern to every other k nearest microcalcification. The pattern is assigned to the class that has the most microcalcifications amongst the k nearest microcalcifications.

Woods et al. [16] modified the KNN algorithm, stating that an unknown test pattern is assigned to a particular class if at least k of the KNNs is in that class. The KNN rule will be more sensitive to microcalcification detection and less sensitive to non-microcalcifications.

Bayesian approaches to classification have been used successfully in their application to diagnosis of microcalcifications [7], [17]. Its decision-making process is based on choosing the most likely class given a particular feature vector. A probability of class membership is calculated and used to classify an area or object.

The following is a list of features identified in the literature: entropy [5], contrast [5], [14], [16], correlation [5], mean [4], [5], [10], [14], [16], deviation [5], area [3], [14], [16], root mean square (rms) [2], shape [16], [17], edge strength [3], [14], [16], smoothness [5], [14], Fourier transform [15], [17], and texture features [18].

The above features have been used by many researchers for the classification of microcalcification, however, not many attempts have been made to determine the importance of a single feature or a combination of features in conjunction with neural networks.

D. Statement of the Problem

In reviewing the literature, it is apparent that a lot of research has been undertaken in the field of digital mammography. However, there is a need for the further review of feature extraction and CI techniques and their application to digital mammography. The aim of the proposed research is to compare different feature extraction techniques for digital mammograms and develop a CAD system to detect and classify microcalcifications. The CI techniques that are used include FL and a neural network.

The remainder of this paper is organized as follows. An explanation of the methodology proposed is provided in Section II. Section III explains the implementation techniques used. Section IV discusses the results achieved, followed with a discussion in Section V. The conclusion of the research is presented in Section VI.

II. PROPOSED METHOD

The method proposed first extracts the suspicious MCC area from the image using the x, y coordinates and radius value al-

2[Online]. Available: http://canopus.llnl.gov/docum.documents/imaging/jmhspie93.html

Fig. 1.   Proposed technique.

ready annotated (provided by radiologists) in the database (or FL algorithm). Features are then extracted from this area and fed to the neural network for classification. The neural network outputs a value between 0–1. A benign MCC outputs a value between 0–0.5 and malignant outputs a value between 0.5–1. This threshold and most significant features in conjunction with FL and neural network were researched and analyzed in this project. The method is presented in Fig. 1.

The proposed method contains the following stages: mammographic database, fuzzy detection algorithm (or centers and radii provided by radiologists), feature extraction, selection of most significant features, and classification of features into benign or malignant (neural network).

Stage 1: Mammographic Database

The database of digital mammograms from University of Nijmegen, Nijmegen, The Netherlands, was used. All images (mammograms) were in raw format and were of size 2048 × 2048. They used 12 b (2 bytes) per pixel of grey-level information, which were recorded with a Kodak MIN-R/S0177 screen film combination using various types of equipment.

Stage 2: Fuzzy Detection Algorithm (or Centers and Radii Provided by Radiologists)

This algorithm aims at detecting microcalcifications and suspicious areas. In the process of detecting, it may detect other areas that look like a microcalcification. It is up to the user to decide whether the resulting detection is a microcalcification or some other area. The algorithm is simple and based on a fuzzy technique where the size of the microcalcification can be "about the size of a microcalcification." It uses a 16 × 16 window to scan over the entire digital mammogram and locate microcalcifications or other abnormalities:

WHILE entire image has not been examined by 16 × 16 window
    MOVE 16 × 16 window to next position
        RECORD x, y position and grey level value of pixel with largest grey level in window
    IF pixels **surrounding** the largest pixel are **as bright as** the largest pixel grey level value
        AND **outer** pixels are **darker** than the largest pixel grey level value
    THEN largest pixel x, y position is the center pixel of a microcalcification area
    END IF
END WHILE

The algorithm only detects the center pixel of a microcalcification area. It is up to the user to view the surrounding area to decide how big the microcalcification area actually is. Microcalcification areas vary greatly in size, shape, and grey color. Therefore, it is very difficult to develop a good algorithm that is effective at detecting microcalcifications based only on digital image processing and FL.

Stage 3: Feature Extraction

The main aim of the research method proposed was to identify the effectiveness of a feature or a combination of features when applied to a neural network. Thus, the choice of features to be extracted was important.

The following 14 features were used for the proposed method: average histogram, average grey level, energy, modified energy, entropy, modified entropy, number of pixels, standard deviation, modified standard deviation, skew, modified skew, average boundary grey level, difference, and contrast. The formulas for entropy, energy, skew, and standard deviation were modified so that the iterations started with the first pixel of the pattern and ended at the final pixel. Traditionally, the formulas for these features have iterations starting with the lowest grey level possible and ranging to the highest grey level possible. This modification was done in an attempt to achieve a better classification rate than its traditional version.

Stage 4: Selection of Most Significant Features

Initially, the method determined the ranking of single features from best to worst by using each feature as a single input to the neural network. After this was completed, a combination of features was tested and a best feature or a combination of features was determined.

1) First Feature Vector (Ten Features): Average histogram, average grey level, number of pixels, average boundary grey, difference, contrast , energy, entropy, standard deviation, and skew.

2) Second Feature Vector (14 Features): Average histogram, average grey level, number of pixels, average boundary grey, difference, contrast, modified energy, modified entropy, modified standard deviation, and modified skew.

The most significant feature or combination of features were selected based on neural-network classification. It was done as follows. We started with a single feature by feeding it to the neural network and analyzing the classification rate. If it was increased or unchanged by adding a particular feature, then we included this feature to the input vector. Otherwise, we removed this feature and added another feature to the existing input vector and repeated the whole process again. The total 8192 combinations were investigated and the combinations with the best classification rate were selected for the development of the our CAD system.

Stage 5: Classification of Features into Benign or Malignant

A back propagation neural network (BPNN) was used for classification of features into benign or malignant. A number of hidden units, learning rates, and momentums were investigated to achieve the best results.
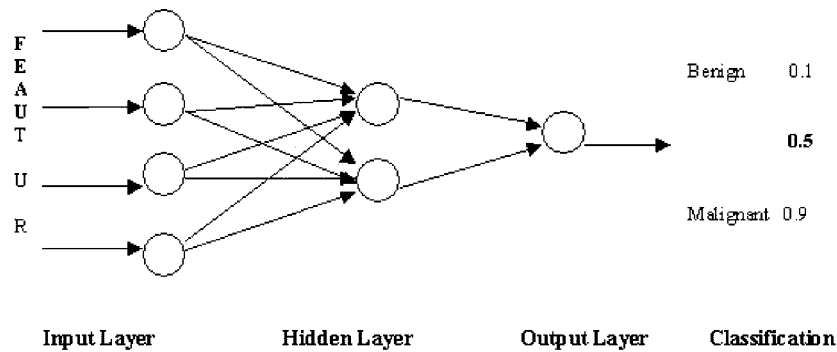
Fig. 2. BPNN structure.

1) Neural-Network Architecture: As shown in Fig. 2, the proposed research method used a feedforward BPNN for classification. The BPNN architecture used only one hidden layer. The number of input nodes and the number of hidden layer nodes must be specified. The number of hidden nodes in the hidden layer were continually adjusted in an attempt to achieve optimum classification rates. Usually, it is purely a matter of experimenting with various structures until a certain structure is recognized as the best structure for a particular set of inputs.

Possible input feature vector values include one feature or combination of features described in the following sections. The output value can be between 0–1. A value of less than 0.5 means that the neural network has classified the input features as belonging to a benign microcalcification. A value of more than 0.5 means that the neural network has classified the input features as belonging to an MCC. This is because in the training of the BPNN, the desired output was specified as 0.1 for benign and 0.9 for malignant.

III. IMPLEMENTATION

This section discusses the various implementation techniques used in developing the breast cancer diagnosis system (BCDS), microcalcification extraction program, experimental preparation program, and neural-network program. Particular emphasis will be placed on discussing the features and limitations of the BCDS.

A. Implementation Language and Platform

The BCDS was implemented using Borland Builder C++ on a Windows platform using a standard personal computer with a Pentium microprocessor. The other programs were implemented using the C and C++ programming languages on a UNIX platform. The microcalcification extraction program and experimental preparation program were both implemented and executed on the SP2 Supercomputer at the Queensland Parallel Supercomputing Facility (QPSF). The SP2 is an IBM machine consisting of eight RS/6000 390 machines and 14 RS/6000 590 machines. The SP2's operating system is UNIX, which supports the chosen programming languages C and C++.

B. Microcalcification Extraction Program

Before any experiments could be run, all microcalcification areas to be used as training or testing samples for the neural network had to be extracted from the Nijmegen mammographic database. This was the main aim of the microcalcification extraction program. The program uses the x, y and radius coordinates (or FL detection algorithm—Section II) for each microcalcification for each digital mammogram in the database to extract the area of interest. These microcalcification areas are then stored in separate files and used later in the experimental preparation program.

C. Experimental Preparation Program

Once all microcalcification areas have been extracted and stored in a file by the microcalcification extraction program, the experimental preparation program then extracts features and lets the user create training and testing files to be used on the neural network. Once all features have been extracted, the program then lets the user create training and testing files. It does this by prompting the user for the features to be used in the feature vector along with the neural-network parameters. Once all this information has been received from the user, the program then automatically creates the testing and training files that are now ready to be used on the neural network.

The main feature of this program is that it allows the user to quickly, easily, and efficiently create training and testing files for experimentation. This is of paramount importance because of the large number of experiments with varying settings that had to be run. Every experiment had different training settings to any other experiment in terms of features selected or the neural-network parameters. This program was vital in allowing continuous experimentation to occur with ease.

D. Explanation of the BCDS

The BCDS was developed to give a demonstration of a possible diagnosis system that could be used in hospitals. The main screen of the BCDS is shown in Fig. 3. BCDS is an easy-to-use system with a main purpose of detecting and diagnosing microcalcification areas in digital mammograms.

1) Features of the BCDS: Although the main purpose of the system is to detect and diagnose microcalcifications, it has a number of other features that assist the user in their diagnosis and evaluation of the digital mammogram. Once the system is loaded, the user has the choice of the following options: open file, grid, zoom, distance, detect, region of interest, diagnose, etc.
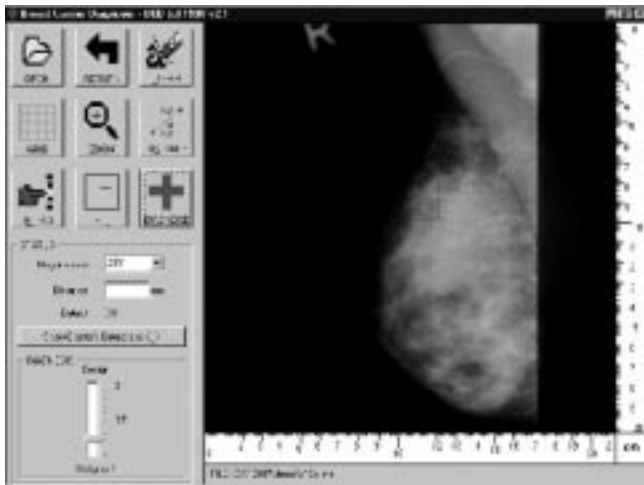
Fig. 3.   Main screen of the BCDS.

The BCDS is a fully working demonstration system that can detect and diagnose microcalcifications effectively in a matter of seconds. This is an enormous bonus to doctors who need immediate responses to any doubt they may have on a particular microcalcification area. The BCDS can be used as a second opinion to the already established conventional methods of diagnosis.

E. Neural-Network Program

The BPNN [25] used for the running of the experiments was modified and used in the BCDS. Originally, the neural network was implemented on a UNIX platform with experimental intentions. Modifications had to be made to the code so the neural network could run under the Windows environment and Borland Builder C++. It was important to do this, as the BCDS needs to use the neural network for the classification of a suspicious area. By incorporating the neural network into the BCDS, the user has an instantaneous diagnosis on a suspicious area. It is easy to maintain changes to the neural-network parameter settings. All settings are held in a text file that the neural-network program reads when it runs. If at any time during further experimentation better neural-network settings are determined, then the text file has to be updated with the new settings. This is a very easy, quick, and practical method of improving the diagnostic accuracy of BCDS without recompiling the entire system.

IV. EXPERIMENTAL RESULTS

Many experiments were run with varying feature vectors and neural-network parameters. Feature vector values were based on the 14 features that were extracted from the microcalcification areas in the Nijmegen mammographic database. A total of 40 microcalcification (20 benign, 20 malignant) areas were used as training samples. The microcalcification areas (centers and radii) for training were provided by two expert radiologists. The value of 0.1 for benign and 0.9 for MCC was used to train the neural network for classification, however, during the recognition (testing) process, the value less or equal than 0.5 was used for benign and greater than 0.5 for malignant. This threshold

TABLE I
DETECTION RATES FOR BENIGN AND MALIGNANT CASES

Detection Rates	
Benign Cases	Malignant Cases
83.30%	77.80%

(0.5) was also investigated in this research and thresholds between 0.1 and 1 to distinguish benign and MCC were investigated to find the best threshold, and the results are presented in Table VII. A total of 18 microcalcification (nine benign, nine malignant) areas were used as testing samples.

The experiments have been conducted to find a best feature or a combination of features, neural-network settings such as the number (#) of hidden units, learning rates, threshold, etc. We are unable to include all experimental results here because more than 20 tables contain all the results. In this paper, we are only presenting the results of experiments with combination of the best four features, threshold, and some results from FL detection algorithm. Our final BCDS uses a combination of best features (entropy, standard deviation, number of pixels—Table VI), neural-network settings (hidden units = 10, learning rate = 0.9, momentum = 0.5—Table VI) and threshold (= 0.3—Table VII).

A. Results of Experiments Using the FL Detection Algorithm

The experiments run using the FL detection algorithm used ten images, five that contained benign cases and five malignant that contained malignant cases. The algorithm achieved good detection rates, as shown below in Table I.

Although the algorithm detected most microcalcification areas with detection rates of 83.3% and 77.8%, it also detected other areas that were not microcalcification areas. This is because the entire digital mammogram is full of areas with varying contrast, size, and shape. The fuzzy detection algorithm alone is not intelligent enough to detect only microcalcifications. A neuro-fuzzy method would probably be more useful, as a neural network could be used to classify a detected area as a microcalcification or nonmicrocalcification. This would help in eliminating detections of nonmicrocalcification areas.

B. Results of Experiments with Combinations of Best Four Features as Input to the Neural Network

All experiments presented in this section relate to Table II. The top four features from many initially tested are shown in Table II. The possible combinations from these top four ranking features are shown in Table III. All these combinations were used for experimentation and, finally, the best combination was used to implement our CAD system. The results from these experiments are shown in Tables IV–VI.

Table IV shows the results of the experiments run using all 11 possible combinations of the top four ranking normal features. All experiments were set with the same neural-network parameter settings so there was consistency across all 11 experiments.

77.8% was the best classification rate achieved in Table IV experiments #5, #9, and #11. The top three best classifying experiments were experiments #5, #9, and #11. Experiments #5, #9, and #11 were run with a different number of hidden units,

TABLE II
RANKING OF NORMAL AND MODIFIED FEATURES

Ranking	Features	
	Normal	Modified
1	Skew	Entropy
2	Number of pixels	Number of pixels
3	Histogram	Histogram
4	Standard Deviation	Standard Deviation

TABLE III
POSSIBLE COMBINATIONS OF THE TOP FOUR RANKING FEATURES

Experiment #	Combinations of Rankings
1	1, 2
2	1, 3
3	1, 4
4	1, 2, 3
5	1, 2, 4
6	1, 2, 3, 4
7	2, 3
8	2, 4
9	2, 3, 4
10	3, 4
11	3, 4, 1

learning rates, and momentum in an attempt to further improvement of the classification rate. The results for experiment #5 are shown in Table V.

No experiments shown in Table V improved on the previous best classification rate of 77.8%. Experiment #6 achieved the same as the previous best classification rate of 77.8%.

Table VI shows the results of experiments with modified combination #5 as the feature vector.

83.3% was the best classification rate achieved in Table VI experiments #7 and #9. Both these experiments had ten hidden units and 0.5 as the momentum value. Although experiments #7 and #8 both had the same number of hidden units, learning rate, and momentum values, the classification rate dropped from 83.3% to 77.8%. This is because experiment #8 was trained for 120 000 iterations and experiment #7 was trained for 60 000. Just because an experiment is trained for a large number of experiments, it does not necessarily mean the classification rate will improve.

C. Results of Experiments with Threshold Value 0.1–0.9

In an attempt to improve the classification rate even further, it was necessary to experiment with the threshold value that determines what is a benign microcalcification and what is MCC. After a testing pair is presented to the neural network, it computes an output value ($0 <$ output value < 1) that is classified according to a threshold value. All results displayed in the previous sections were based on a set threshold value of 0.5. Any output values smaller than or equal to 0.5 were classified as benign and any values greater or equal to 0.5 were classified as malignant.

0.5 may not necessarily be the ideal threshold to classify benign microcalcification and MCCs, although it is situated exactly halfway between desired benign (0.1) and malignant (0.9) values. The experiments shown in Table VII were run in an attempt to determine the best threshold value with the same

training and testing pairs used in earlier sections. The experiments shown in both Table VII used modified entropy, modified standard deviation, and number of pixels as the feature vector. This feature vector best distinguishes a benign microcalcification from MCC as determined in the previous sections.

Of the experiments shown in Table VII, experiment #3 achieved the best classification rate of 88.9%. Experiment #4, however, achieve the best training classification rate of 82.5% and a good testing classification rate of 83%.

V. DISCUSSION AND FUTURE RESEARCH

A. Comparison of Classification Rates

Various results were achieved using various features, neural-network settings, and threshold values. The results of the all conducted experiments (some of them are shown in previous sections) can be grouped into the following five different sections:

1) results of experiments with single features as input to the neural network;
2) results of experiments with combined features randomly selected as input to the neural network;
3) results of experiments with all ten features as input to the neural network;
4) results of experiments with combinations of best four features as input to the neural network;
5) results of experiments with threshold value 0.1–0.9.

The best classification rates from each section are shown in Table VIII.

Overall, 88.9% was the best classification rate achieved in section #5 when the threshold value was set to 0.4. Sections #1, #2, and #4 achieved a good testing classification rate of 83.3%. Although the training classification of sections #1, #2, #4, and #5 were also good, they did not achieve as well as the training classification rate of 100% in section #3.

Section #1 determined which single feature was the best feature to describe a microcalcification area. Out of all 14 features, modified entropy achieved the best classification rate of 83.3%. Experiments achieved the best results when they were run for 30 000 iterations and had learning rate and momentum values set to 0.5 or smaller.

In section #2, the best classification rates were achieved when number of pixels and contrast features were combined together as the feature vector. Once again 83.3% was the best classification rate achieved with 16 hidden units and learning rate and momentum values set to 0.5. Generally, the better results in section #2 were achieved when experiments were run for 20 000 or 30000 iterations. Most experiments that were run for over 30 000 iterations achieved a poor testing classification rate, but a good training classification rate. This is because the neural network overtrained on the training data. Its ability to recognize data it had been trained with was obviously good and its ability to recognize data it had not been trained with was poor.

Unlike sections #1, #2, #4, and #5, section #3 achieved a 100% training classification rate in most experiments. A modest 72.2% classification rate was achieved when all features were combined and run on the neural network with 12 hidden units and 20 000 iterations. When the number of hidden units were increased or decreased, the classification rate dropped.

TABLE IV
CLASSIFICATION RATES FOR NORMAL COMBINATIONS AS THE FEATURE VECTOR

#	# Hidden Units	Learning Rate	Momentum	# Iterations	Classification Rate	
					Training Set	Test Set
1	16	0.5	0.5	10000	70.0% [28/40]	72.2% [13/18]
2	16	0.5	0.5	10000	70.0% [28/40]	72.2% [13/18]
3	16	0.5	0.5	10000	65.0% [26/40]	61.1% [11/18]
4	16	0.5	0.5	10000	70.0% [28/40]	66.7% [12/18]
5	16	0.5	0.5	10000	85.0% [34/40]	77.8% [14/18]
6	16	0.5	0.5	10000	80.0% [32/40]	61.1% [11/18]
7	16	0.5	0.5	10000	62.5% [25/40]	72.2% [13/18]
8	16	0.5	0.5	10000	65.0% [26/40]	66.7% [12/18]
9	16	0.5	0.5	10000	75.0% [30/40]	77.8% [14/18]
10	16	0.5	0.5	10000	65.0% [26/40]	66.7% [12/18]
11	16	0.5	0.5	10000	85.0% [34/40]	77.8% [14/18]

TABLE V
CLASSIFICATION RATES FOR NORMAL COMBINATION #5 AS THE FEATURE VECTOR

#	# Hidden Units	Learning Rate	Momentum	# Iterations	Classification Rate	
					Training Set	Test Set
1	16	0.5	0.5	30000	85.0% [34/40]	61.1% [11/18]
2	16	0.5	0.5	20000	82.5% [33/40]	61.1% [11/18]
3	16	0.5	0.5	5000	67.5% [27/40]	66.7% [12/18]
4	16	0.9	0.9	10000	75.0% [30/40]	61.1% [11/18]
5	14	0.5	0.5	10000	82.5% [33/40]	55.6% [10/18]
6	12	0.5	0.5	10000	85.0% [34/40]	77.8% [14/18]
7	12	0.1	0.1	10000	67.5% [27/40]	66.7% [12/18]
8	12	0.5	0.5	30000	85.0% [34/40]	61.1% [11/18]
9	10	0.5	0.5	10000	80.0% [32/40]	72.2% [13/18]
10	8	0.9	0.9	10000	75.0% [30/40]	61.1% [11/18]

TABLE VI
CLASSIFICATION RATES FOR MODIFIED COMBINATION #5 AS THE FEATURE VECTOR

#	# Hidden Units	Learning Rate	Momentum	# Iterations	Classification Rate	
					Training Set	Test Set
1	16	0.9	0.5	10000	72.5% [29/40]	77.8% [14/18]
2	16	0.9	0.9	10000	72.5% [29/40]	77.8% [14/18]
3	16	0.1	0.1	10000	70.0% [28/40]	66.7% [12/18]
4	12	0.5	0.5	10000	75.0% [30/40]	77.8% [14/18]
5	12	0.9	0.1	10000	72.5% [29/40]	77.8% [14/18]
6	10	0.9	0.5	7000	67.5% [27/40]	77.8% [14/18]
7	10	0.9	0.5	60000	77.5% [31/40]	83.3% [15/18]
8	10	0.9	0.5	120000	80.0% [32/40]	77.8% [14/18]
9	10	0.1	0.5	20000	70.0% [28/40]	83.3% [15/18]
10	6	0.9	0.9	30000	77.5% [31/40]	66.7% [12/18]

Section #4 achieved consistently high classification rates when compared with the other four sections. The best classification rate achieved was 83.3% in a number of experiments. Although section #4 did not improve on the previous best classification rate, overall the results were better. Modified combination #5 (modified entropy, number of pixels, and modified standard deviation) achieved 83.3% when run for 60 000 iterations and had the learning rate set to 0.9 and momentum set to 0.5. This experiment achieved a training classification rate of 77.5%, which was the best training result out of all the experiments that achieved a top testing classification rate of 83.3%. For this reason, these features were decided as the best combination of features to describe a microcalcification and used for the threshold analysis experiments in section #5.

Section #5 determined that 0.3 was the best threshold value to classify microcalcifications as either benign or malignant when using modified combination #5 as the feature vector. A top classification rate of 88.9% was achieved. The classification rate dropped with the threshold value was increased or decreased. Thus, 0.3 will be used as the threshold value in the BCDS with

TABLE VII
CLASSIFICATION RATES OF TRAINING AND TESTING PAIRS FOR THRESHOLD
0.1–0.9 USING MODIFIED ENTROPY, MODIFIED STANDARD DEVIATION, AND
NUMBER OF PIXELS AS THE FEATURE VECTOR

#	Threshold	Classification Rate	
		Training	Testing
1	0.1	57.5% [23/40]	72.2% [13/18]
2	0.2	57.5% [23/40]	77.8% [14/18]
3	0.3	70.0% [28/40]	88.9% [16/18]
4	0.4	82.5% [33/40]	83.3% [15/18]
5	0.5	77.5% [31/40]	83.3% [15/18]
6	0.6	77.5% [31/40]	72.2% [13/18]
7	0.7	70.0% [28/40]	72.2% [13/18]
8	0.8	65.0% [26/40]	72.2% [13/18]
9	0.9	62.5% [25/40]	55.6% [10/18]

TABLE VIII
CLASSIFICATION RATES FOR SECTIONS #1–#5

Section #	Classification Rate	
	Training	Testing
1	70.0%	83.3%
2	85.0%	83.3%
3	100%	66.7%
4	85.50%	83.3%
5	82.50%	88.9%

modified entropy, number of pixels, and modified standard deviation as the feature vector.

This research has achieved very commendable results when compared with results obtained by other leading researchers. Woods *et al.* [16] achieved a classification rate of 93.6%. This, however, was only for the classification of areas as microcalcifications or nonmicrocalcifications. Woods *et al.* experimented with a number of classifiers including a BPNN, a linear classifier, the KNN classifier, and a binary decision tree. The results from each of the classifiers did not vary too much, with the best being 93.6% and the worst being 88.8%.

Nishikawa *et al.* [17] achieved a classification rate of 85%. Chitre *et al.* achieved a true-positive classification rate of 72.4%. Shen *et al.* achieved a 94% classification rate for benign cases and an 87% classification rate for malignant cases. Woods *et al.*, Nishikawa *et al.*, Chitre *et al.*, and Shen *et al.* used feature extraction techniques and neural networks for classification.

It is difficult to justify and compare how successful a method is as there has been no benchmark database testing. Meaningful comparisons require methods to be tested by the same database. Alternatively, databases containing similar cases of a similar degree of difficulty could be used. Both options would then allow for meaningful comparisons of results and truly distinguish the best methods.

B. BPNN as a Classifier

The neural network used for classification performed extremely well. It achieved good classification rates with a top rate of 88.9%. This result emphasizes the potential of the BPNN to be used as a microcalcification classifier. The good classification rate that the BPNN did achieve meant that there was no need to use another classifier to try and improve the classification rate. It was of more interest to focus the

experimentation on trying to improve the classification rate by comparing feature vectors and neural-network settings. Consequently, a very high classification rate of 88.9% was achieved. Future research may improve this classification rate and achieve over 90%.

There was no consistency in the trend of results when the learning rate and momentum values were altered. Some experiments achieved good results with high learning rate and momentum values and others performed well with lower learning rate and momentum values. This made it difficult to establish a strategic approach to the experimentation, as each set of features responded differently when run with the same neural-network settings.

C. Future Research

1) Improving the Classification Rate: There will always be a need to continue researching until a method is developed that classifies with 100% accuracy. Obviously, it is arguable whether this will ever eventuate. However, since the motivation to save human lives is inspiring researchers to develop accurate and efficient methods of detection and diagnosis, research will and should continue well into the future.

88.9% is a very good classification rate achieved by the experiments using the proposed method. Plans have already been made to continue testing the method with a database from a local breast screening clinic. This experimentation will thoroughly test the stability of the proposed method. To build a reputation as a good solid diagnoser of breast cancer, a classification rate above 80% needs to be achieved when tested with foreign samples. Even if the proposed method does not classify with such a high accuracy when tested on a foreign database, it will provide valuable ideas as to how the method can be improved and how it can be made more accurate, efficient, and effective.

2) Developing the BCDS Further: Although the BCDS has impressive capabilities, it does have limitations due to time constraints and the scope of this research. These limitations include the following: only detects and diagnoses benign and MCC and no other abnormalities or lesions, does not have a classification rate of 100%, and best displays digital mammograms that are 2048 × 2048 pixels in resolution.

The BCDS developed is a demonstration of what a real-life diagnosis system could be. To become viable as a real-life system, the BCDS will need to overcome limitations. It has the potential to do this with some further research and development, as the foundation work has now been completed.

The first two limitations will only be overcome with more research. It is a difficult process to diagnose benign microcalcifications and MCCs, let alone any other types of abnormalities or lesions. Being able to diagnose other abnormalities or lesions would be a very useful feature that could be added in the future. Whether anyone will ever develop a system that diagnoses with a classification rate of close to 100% is arguable, but not very probable. The BCDS, however, is a very effective and useful tool for detecting microcalcifications and diagnosing them as benign or malignant.

The last limitation mentioned is not very difficult to overcome. With more time, this option would have been implemented accordingly. Since digital mammograms can be

stored in a number of different file formats, it would be very time consuming to consider every one of these file formats and implement the system to be capable of handling each format. The Nijmegen database that was used for this research only contained raw format files of 2048 × 2048, thus, the choice to only implement the system for this file format.

VI. CONCLUSION

In this paper, we have presented a CAD system for detection and classification of microcalcification in digital mammograms. We have investigated and analyzed 14 feature extraction techniques, neural-network settings, and a FL detection algorithm. We have modified some traditional features and found that a combination of our three modified features, such as entropy, standard deviation, and number of pixels, is the best combination of features to distinguish a benign microcalcification pattern from one that is malignant. We have also found that a threshold of 0.3 could further improve our system. Our system has achieved promising results, with a top result of 88.9%. This result emphasizes the potential of the system to be used to detect and diagnose microcalcifications' patterns in digital mammograms.

Our system was tested by local radiologists, and they were very impressed. In their opinion, 88.9% is a very good classification rate achieved by the experiments using our system. Plans have already been made to continue testing the method with a database from a local breast screening clinic.

REFERENCES

[1] N. Karssemeijer, *Computer-Assisted Reading Mammograms, European Radiol.*, vol. 7, pp. 743–748, 1997.

[2] L. Mascio, M. Hernandez, and L. Clinton, "Automated analysis for microcalcifications in high resolution mammograms," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 1898, pp. 472–479, 1993.

[3] L. Shen, R. Rangayyan, and J. Desaultels, *Detection and Classification Mammographic Calcifications, International Journal of Pattern Recognition and Artificial Intelligence*. Singapore: World Scientific, 1994, pp. 1403–1416.

[4] F. Aghdasi, R. Ward, and B. Palcic, "Restoration of mammographic images in the presence of signal-dependent noise," in *State of the Art in Digital Mammographic Image Analysis*. Singapore: World Scientific, 1994, vol. 7, pp. 42–63.

[5] Y. Chitre, A. Dhawan, and M. Moskowtz, "Artificial neural network based classification of mammographic microcalcifications using image structure features," in *State of the Art of Digital Mammographic Image Analysis*. Singapore: World Scientific, 1994, vol. 7, pp. 167–197.

[6] E. Pisano and F. Shtern, "Image processing and computer-aided diagnosis in digital mammography," in *State of the Art of Digital Mammographic Image Analysis*. Singapore: World Scientific, 1994, vol. 7, pp. 280–291.

[7] K. Bowyer and S. Astley, *The Art of Digital Mammographic Image*. Singapore: World Scientific, 1994, vol. 7, p. v.

[8] Y. Pao, *Adaptive Pattern Recognition and Neural Networks*. Reading, MA: Addison-Wesley, 1989, ch. 2, pp. 25–56.

[9] S. Palaniswami, Y. Attikiouzel, R. Marks, D. Fogel, and T. Fukuda, *Computational Intelligence*. Piscataway, NJ: IEEE Press, 1995, sec. 6, pp. 277–325.

[10] B. Zheng, W. Qian, and L. Clarke, "Multistage neural network for pattern recognition in mammogram screening," in *IEEE Int. Neural Networks Conf.*, 1994, pp. 3437–3447.

[11] A. Drijarkara, G. Naghdy, and F. Naghdy, "Fuzzy detection of microcalcifications in mammograms," in *Proc. 2nd IEEE Int. Intelligent Processing Syst. Conf.*, 1998, pp. 314–317.

[12] M. Sutton and J. Bezdek, "Enhancement and analysis of digital mammograms using fuzzy models," in *Proc. SPIE*, 1997, pp. 345–352.

[13] N. Karssemeijer, "Adaptive noise equalization and recognition of microcalcification clusters in mammograms," in *State of the Art of Digital Mammographic Image Analysis*. Singapore: World Scientific, 1994, vol. 7, pp. 148–166.

[14] J. Solka, W. Poston, C. Priebe, G. Rogers, L. Richard, D. Marchette, K. Woods, and K. Bowyer, "The detection of microcalcifications in mammographic images using high dimensional features," in *7th Annu. IEEE Computer-Based Med. Syst. Symp.*, 1993, pp. 139–145.

[15] H. Barman, G. Granlund, and L. Haglund, "Feature extraction for computer-aided analysis of mammograms," in *State of the Art of Digital Mammographic Image Analysis*. Singapore: World Scientific, 1994, vol. 7, pp. 128–147.

[16] K. Woods, C. Doss, K. Bowyer, J. Solka, C. Priebe, and W. Kegelmeyer, "Comparative evaluation of pattern recognition techniques for detection of microcalcifications in mammography," *Int. J. Pattern Recog. Artif. Intell.*, vol. 7, no. 6, pp. 1417–1436, 1994.

[17] R. Nishikawa, M. Giger, K. Doi, C. Vyborny, and R. Schmidt, "Computer-aided detection and diagnosis of masses and clustered microcalcifications from digital mammograms," in *State of the Art of Digital Mammographic Image Analysis*. Singapore: World Scientific, 1994, vol. 7, pp. 82–102.

[18] C. Priebe, R. Lorey, D. Marchette, J. Solka, and G. Rogers, "Nonparametric spatio-temporal change point analysis for early detection in mammography," in *Proc. 2nd Int. Digital Mammography Workshop*, 1994, pp. 111–120.

[19] H. Yoshida, R. Nishikawa, K. Muto, K. Doi, and M. Tsuda, "Application of the wavelet transform to automated detection of clustered microcalcifications in digital mammograms," Tokyo Inst. Polytech., Tokyo, Japan, Academic Rep., vol. 16, 1994.

[20] H. Yoshida, R. Nishikawa, G. Maryellen, and K. Doi, "Computer-aided diagnosis in mammography: Detection of clustered microcalcifications based on multiscale edge representation," in *Computer Assisted Radiology*. Amsterdam, The Netherlands: Elsevier, 1996.

[21] N. Karssemeijer, "Stochastic model for automated detection of calcifications in digital mammograms," *Image Vision Comput.*, vol. 10, no. 6, pp. 369–375, 1992.

[22] H. Wechsler, *Neural Networks for Perception*. New York: Academic, 1992.

[23] J. Hertz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computation*. Reading, MA: Addison-Wesley, 1991.

[24] D. Brzakovic and M. Neskovic, "Mammogram screening using multi-resolution-based image segmentation," in *State of the Art of Digital Mammographic Image Analysis*. Singapore: World Scientific, 1994, vol. 7, pp. 103–127.

[25] B. Verma, *Comparative Evaluation of Two Neural Network-Based Techniques for Classification of Microcalcifications in Digital Mammograms, Knowledge and Information Systems: An International Journal*. Berlin, Germany: Springer-Verlag, 1999, vol. 1, pp. 107–117.

Brijesh Verma received the B.S., M.S., and Ph.D. degrees in information technology from the Warsaw University of Technology, Warsaw, Poland.

He is currently with the School of Information Technology, Griffith University, Gold Coast, Queens., Australia. His research interests include artificial neural networks, pattern recognition, FL, content-based image retrieval, segmentation and recognition of cursive handwriting, and digital mammography. He has authored or co-authored over 50 technical papers, one book, one book chapter and has co-edited six books. He is Co-Editor-in-Chief of the *International Journal of Computational Intelligence and Applications* (IJCIA).

Dr. Verma is a member of the IEEE Computer Society. He was a general co-chair of ICCIMA'99, local arrangement chairs of IEEE ICIPS'98, and IPMM'97, program chair of ICCIMA'98, and a program committee member of many international conferences.

John Zakos received the B.S. degree (with honors) in information technology from the School of Information Technology, Griffith University, Gold Coast, Queens., Australia.

He is currently a Software Developer at IXLA, Melbourne, Australia. His research interests include neural networks, FL, breast cancer detection, digital mammography, and image processing.