MICROCALCIFICATION ENHANCEMENT IN DIGITIZED MAMMOGRAM IMAGES USING FUZZY LOGIC TECHNIQUE

Dr. Ayman AbuBaker
Asst. Prof. at Electrical and Computer Engineering Dept, Applied Science Private University, Jordan
a_abubaker@asu.edu.jo

ABSTRACT
This paper proposes a novel fuzzy logic method which aims to enhance the performance of Radiologists by automatically enhancing the microcalcifications in the mammogram images. In this approach, the image intensities are fuzzified using three linguistic labels. Then 6561 activation rules are generated as a inference engine based on investigation the eight connected neighbors of a mask size 3x3 pixels. Finally, the value of the local mask center is enhanced after defuzzification the input sets. The proposed method was tested and evaluated on two different types of resources which are Mammographic Image Analysis Society database (MIAS) and University of South Florida (USF) database. As a result, this technique can enhance the microcalcifications regions in mammogram with acceptable number of false positive regions.

Keywords: Medical Imaging; Mammograms; Microcalcifications; Enhancing MCs, Fuzzy logic.

1 INTRODUCTION
Breast cancer is the second leading cause of cancer deaths in women today after lung cancer and is the most common cancer among women, excluding nonmelanoma skin cancers [13][18]. The breast contains a variable amount of connective, glandular and fatty tissue. These are organized in structures of different shapes, densities and scales. In cases involving a large amount of glandular tissue (so-called glandular breasts) mammograms are very bright, which significantly decreases the visibility of subtle microcalcifications (MCs). In situations involving the examination of large numbers of mammograms, the efficiency of visual inspection also drops considerably [2], [3], [4]. It is thought that approximately 10% to 30% of breast cancer cases are missed by radiologists [11], [14], [15], [21]. This might be caused by very small and non-palpable MC clusters and/or masses.

Early detection is the key to improving breast cancer prognosis. Mammography has been shown to be one of the most reliable methods for early detection of breast carcinomas. Most of the work in mammography aims at detecting one or more of the three abnormal structures in mammograms [9][18]: microcalcifications, circumscribed masses and speculated lesions. Other methods depend on classifying the breast lesions as benign or malignant [9]. There are problems with the subjective analysis of mammographic images by radiologist. Subjective analysis depends mainly of the experience of the human operator, but it is also affected by fatigue and other human-related factors. Since, the interpretation is a repetitive task that requires lot of attention to minute details, it requires lot of staff time and effort, which results in increasing diagnosis time. On the other hand, the objective analysis of mammograms, which is carried out by automated systems, provides consistent performance but its accuracy is usually lower. Due to the sensitivity of this problem, we believe that radiologists should be involved and computers should not replace them completely. However, computer systems can help them perform better by enhancing the quality of images, highlighting the suspicious regions and providing better analysis tools.

For these reasons, computer-aided diagnoses (CAD) are exciting a great deal of attention from the radiologist community [16], [17]. CAD is defined as a diagnosis made by a physician taking into account the computer output as a second opinion. The goal of applying CAD is to support radiologists’ image interpretation and improve the diagnostic accuracy and consistency [16] since automated interpretations of microcalcifications and masses are very difficult due to the regions of interests are usually of low contrast, especially in the case of young women. The difference between the suspicious areas and the normal tissues can be quite slim.

The Mammographic feature enhancement will be essential and critical for automated mammogram analysis [6]. It is performed to emphasize the suspicious regions in the mammogram images and also to the radiologist in detection the tumors easily.

This paper is organized as follows: a brief survey...
of previous work is presented in Section 2. The mammogram image databases used in the work are described in section 3. The fuzzy logic approach is introduced in Section 4. The algorithm evaluations are presented in Section 5, while concluding remarks are given in Section 6.

2 CURRENT APPROACHES FOR ENHANCING AND DETECTING MCS IN MAMMOGRAM IMAGES

Several authors had investigated the implementation of mathematical and artificial intelligence methods for enhancing and detecting MCs in mammogram images. One of these techniques, described by Cheng [5], used Fuzzy logic and scale-space approaches. The first stage of the procedure was to apply a fuzzy-based image enhancement to images from the Nijmegen database. The regions of interest in each image were manually located as rectangular sub-images that contained a maximum number of MCs. The MC clusters in the enhanced images were then detected using a Laplacian of Gaussian (LoG) filter. The resulting approach was very effective in locating the MCs achieving a TP rate of about 90% with a FP rate of about 1%. Netsch and Peitgen [8] proposed an approach for automatic detection of MCs utilizing multi-scale analysis based on the LoG filter and a mathematical model that describes MCs as bright spots of different sizes. This model was effectively detecting the MCs by 84% TP with 1 FP cluster per image. An NN approach was also investigated in order to reduce the FP rate in the mammogram images by Gurcan et al. [7]. Their method employed convolution neural network (CNN), from which the potential MC locations were determined using global and local thresholds. The FP percentage was then reduced using a first rule based classification that employs size contrast and signal-to-noise ratio (SNR) information. A trained CNN classifier was finally used to recognize the abnormal patterns. When they used CNN, they had a high MCs detection rate which was 93.3% TP with 0.7 FP cluster per image. Papadopoulos et al. [12] also used an artificial intelligence method using a hierarchical pyramid neural network for detecting the suspicious locations in mammogram images. This method can successfully detect the MCs and masses in mammogram images, and can also reduce the FP predictions by 50%. Since MCs are difficult to detect because of the variability in their shape, supervised learning using support vector machines (SVMs) was proposed by El-Naqa et al. [3]. The sensitivity of El-Naqa’s algorithm in detecting the MCs was 94% with one false positive cluster per image. The number of features was reduced through a principal component analysis (PCA). As a result, the detected TP was 92% with 1.15 clusters per image.

3 DATABASE RESOURCES

In this work, the MC enhancement algorithm is trained and tested on 190 mammographic images from the University of South Florida (USF) and MIAS databases (140 from USF and the remainder from MIAS). The USF database is a publicly available digital database for mammography screening. Its images are collected from different medical schools and hospitals across the USA. These images all have the same specification (3000 pixel × 4500 pixel and 16-bit pixel depth). This database is divided into four volumes representing the different types of diagnosis: normal, cancer, benign, and benign without call back. Normal images are from patients with normal examination results that have had normal examinations in the previous four years. A normal screening examination is one in which no further “work-up” is required. Cancer images are from patients with screening examinations in which at least one pathology proven cancer is found. Benign cases are from patients with screening examinations in which something suspicious was found, which turned out to be non-malignant (by pathology, ultrasound or some other means). The term benign without callback is used to identify benign cases in which no additional X-rays or biopsies were done. In this paper 70 MCs mammogram images are collected from seven cancer volumes and 70 normal mammogram images are collected from four normal volumes. The cancer volumes are: cancer_01, cancer_05, cancer_06, cancer_07, cancer_13, cancer_14, and cancer_15. The normal volumes are: normal_02, normal_05, normal_07 and normal_09.

The MIAS mammograms have been carefully selected from the United Kingdom National Breast Screening Program. The 322 images represent 161 patients in the MIAS database. These images have been expertly diagnosed and the positions of the MCs in each image are recorded. In this paper, 25 MC and 25 normal additional mammogram images were selected from the MIAS database. The mammograms in this database were obtained using the medio-lateral oblique (MLO) view and were digitized at a spatial resolution of 0.05 mm pixel size.
with 8-bit density resolution. Four image sizes, corresponding to different breast sizes, are included in the 322 images from 161 patients: small (4320 pixel × 1600 pixel), medium (4320 pixel × 2048 pixel), large (4320 pixel × 2600 pixel) and extra-large (5200 pixel × 4000 pixel). Digitization was performed on a Joyce-Loeble scanning microdensitometer (SCANDIG-3) which has a linear response in the range 0.0 to 3.2 optical densities.

4 PROPOSED METHOD

In this section, a complete computer platform for the automated enhancement of MCs in mammograms is introduced. The proposed system consists of 4 major stages as shown in Fig. 1.

Figure 1: Enhancement flowchart

4.1 Pre-processing Stage

Artifacts exist in the mammogram images because of the process of capturing images on X-ray film and the process used to digitize the images. These artifacts are removed based on the fact that MCs areas in mammogram images are hazy regions [18]. I have conducted extensive analysis of 190 mammogram images from the USF and MIAS databases and concluded that all MCs have gray scale values in the range from 40 to 240. In accordance with these observations, each mammogram is analyzed to determine more precisely: an upper limit threshold (UT) and a lower limit threshold (LT). These thresholds are used to exclude the artifacts within the breast region and are determined based on the statistical characteristics of every mammogram. An upper threshold value (UT), is set to 240 to eliminate the bright artifact regions. The lower threshold value (LT) is calculated from the mean \( \mu_0 \) and standard deviation \( \sigma_0 \) values of all non-zero pixel values \( x_{i,j}^0 \), as shown in Eq. (1). Since I have different types of breasts such as normal, fatty and dense breasts, I have found that using this equation provides better performance compared to using the mean value only, which will eliminate MC clusters from dense breasts in particular.

\[
LT = \mu_0 - \sigma_0 \quad (1)
\]

Applying a lower threshold LT helps to eliminate the background regions and low intensity pixel artifacts, and to eliminate the low-level boundary regions of the breast, hence reducing the size of the regions of interest. These two limits are applied as the algorithm scans the breast region to enhance the potential MCs in the mammogram images.

4.2 Fuzzification the Mammogram Image

In Fuzzy Logic, the fuzzification stage means that each input is related to a fuzzy set that has number of sub-sets with different shapes. The crisp (Image Intensity) input is translated into fuzzy linguistic labels (L, M,…) with membership values \( (\mu_L(x), \mu_M(x),...) \) of the input crisp value of that set. So in the proposed approach, the input image intensities were represented as fuzzy sets with three linguistic labels (Low, Medium, High) within a universe of discourse of \([LT \text{ to } UT]\) grey level as shown in Fig. 2.

Figure 2: Fuzzification input image intensities

The membership function of a fuzzy set maps all input intensities of the set into real numbers in [0, 1]. The larger values of the membership represent the higher degrees of the belongings. In this paper two commonly used membership functions are used for a gray level image which are trapezoid and triangular membership function as presented in Eq. (2) and (3) respectively.

\[
M(X,a,b,c,d) = \begin{cases} 
0 & X \leq a \\
\frac{(X-a)}{b-a} & a \leq X \leq b \\
1 & b \leq X \leq c \\
\frac{(d-X)}{c-b} & c \leq X \leq d \\
0 & X \geq d 
\end{cases} \quad (2)
\]

\[
M(X,a,b,c) = \begin{cases} 
0 & X \leq a \\
\frac{(X-a)}{b-a} & a \leq X \leq b \\
\frac{(d-X)}{c-b} & b \leq X \leq c \\
0 & X \geq c 
\end{cases} \quad (3)
\]
The parameters $a$, $b$, $c$ which determine the shape of the trapezoid and triangular membership function are dynamically set based on the LT and UT for each mammogram image. These parameters are changes for different mammogram images that have different topologies. The step $(S)$ in the fuzzification engine is calculated based on the Eq. (4).

$$S = \frac{UT - LT}{4} \quad (4)$$

### 4.3 Fuzzy Inference Engine

In the inference stage, the first step is to construct a fuzzy rule base engine. The construction of this fuzzy rule requires the knowledge of the actions that the MCs need to be enhanced. The number of the rules to construct the rule-base depends on the number of inputs ($n$) and the number of fuzzy sets allocated to each input ($m$). Consequently, the number of rules required is $m^n$. As shown in Fig. 3, a mask of size $3 \times 3$ is used to select number of inputs for the inference engine. The location of the central pixel of the mask is chosen as a reference point and the eight connected neighbors will be inputs for the inference engine. The mask convolution for the whole mammogram pixels in the range between the LT and UT are processed in inference engine as eight inputs for each convolution step. Therefore number of rules in the inference engine is $38 \times 6561$ activation rule. The rules were of the form:

**IF $N1$ is Low And $N2$ is Low And $N3$ is Low And $N4$ is Medium And $N5$ is Medium And $N6$ is High And $N7$ is High And $N8$ is High THEN Enhance is 20E**

Figure 3: Eight connected neighbors

Where, the variable $Enhance$ is the output action from the inference block. The output from this block was called the enhancement for MCs in the mammogram image. The output variable $Enhance$ was represented by five linguistic labels (-40E, -20E, NE, 20E, 40E) within a universe of discourse of (-50% to +50% of the intensity value of mask center) as will be explained in the next section.

### 4.4 Defuzzification Engine

The Defuzzification stage receives the output of the inference engine and converts them into crisp (intensity) values. In this paper, the defuzzification is represented as five linguistic labels (-40E, -20E, NE, 20E, 40E). The universe of discourse for the fuzzy set is dynamic based on the intensity value of local mask center. The universe of discourse is set to be (-50% to +50% of the intensity value of mask center) as shown in Fig. 4.

Figure 4: Defuzzification output sets

After defuzzification is carried out some rules are set to have the final intensity value of the local mask center as:

1. If the output of the defuzzification is less than LT then the LT is set as an output for this pixel.
2. If the output of the defuzzification is greater than UT then the UT is set as an output for this pixel.
3. If the output of the defuzzification is fraction such as 102.4 then an approximation to up is carried out so the output will be 103 grey levels.

### 5 EVALUATION THE PROPOSED ALGORITHM

Different mammogram images from the USF and MIAS databases are used in this evaluation process. The processed images are later subjectively compared with pre-diagnosis cases for the mammogram images from the USF and MIAS databases in order to classify the detected regions into TP and FP clusters. The processing time needed to enhance the MCs regions in the mammogram images was monitored and recorded for each image. Here, I present a few of the experimental results to demonstrate the performance of the proposed method and also the image size and processing time is proposed for each image as shown in Fig. 5, 6, 7, 8.

The main advantage gained by utilizing the fuzzy logic approach that enhancing the MCs regions in the mammogram images. The recorded processing time, is the time needed to process the whole mammogram image which is high as shown in the previous figures. On the other hand, the false positive regions that are enhanced in the mammogram image are acceptable.
Figure 5: (A) Original Image of Size 1656×2559 pixels (B) processed Image, 2.34 minute processing time.

Figure 6: (A) Original Image of Size 914×2486 pixels (B) processed Image, 1.42 minute processing time.
Figure 7: A) Original Image of Size 1134× 2513 pixels (B) processed Image, 2.01 minute processing time.

Figure 8: (A) Original Image of Size 1513× 3000 pixels (B) processed Image, 2.52 minute processing time.
6 CONCLUSION
In this work, a new method was designed and implemented to enhance the MCs in the mammogram images. The new fuzzy logic approach includes four main stages: (1) Preprocessing the mammogram image, (2) Fuzzification mammogram image, (3) Inference Engine, (4) Defuzzification the mammogram image. In the first stage, the artifacts are accurately extracted from the breast region using LT and UT. Then, all the mammogram intensities are represented as fuzzy set. The inference engine is also generated based on using 6561 activation rules. In the final stage, the MC is enhanced after defuzzification the mammogram images. This algorithm was tested on 190 mammogram images from both USF and MIAS database. As a result, this algorithm can enhance the MCs in mammogram images with an accepted number of false positive regions but in high processing time.
In the near future work, this work will be modifying this algorithm to accurately enhance the MCs regions by using different mask sizes. Also the hybrid neuro-fuzzy approach will be implemented to reduce the processing time and to increase the performance of the algorithm by accurately enhancing the MCs regions with minimum number of false positive regions.

7 ACKNOWLEDGEMENT
I wish to express our gratitude to the Applied Science Private University (ASU), Department of Electrical and Computer Engineering, Faculty of Engineering, Jordan, for supporting this research.

8 REFERENCES