

Detection of partially overlapped masses in mammograms

Mustafa H. Alhabib¹, Omar H. Alhabib²

¹Communication & Computer Engineering Department, Cihan University-Erbil, KRG, Iraq

²Software Engineering Department, Mosul University, Iraq

Article Info

Article history:

Received Jul 10, 2019

Revised Oct 1, 2019

Accepted Oct 21, 2019

Keywords:

Clahe

Euclidean distance

Mammography

Overlapping

Segmentation

Thresholding

ABSTRACT

Breast cancer remains one of the major causes of cancer deaths among women. For decades, screening mammography has been one of the most common methods for early cancer detection and diagnosis. Digital mammography images are created by applying a small burst of x-rays that pass through the breast to a solid-state detector, which transmits the electronic signals to a computer to form a digital image. However, due to projection, some mass areas may be partially covered, which makes them difficult to be interpreted. This paper addresses the issue of potential mass regions being distorted by other normal breast tissues, which will negatively affect some of the features being extracted from the mass and in turn deteriorate the classification accuracy. The goal was to estimate the overlapped parts of the mass border using Euclidean distance in order to give more accurate results in next stages. The presented method achieved 95.744% region sensitivity at 0.333 False Positive per Image (FPI), outperforming other researches in this branch of mammography analysis.

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Mustafa H. Alhabib,
Communication & Computer Engineering Department,
Cihan University-Erbil, KRG, Iraq.
Email: mustafa.alhabib@cihanuniversity.edu.iq

1. INTRODUCTION

Breast cancer remains a leading cause of cancer deaths among women in many parts of the world [1]. It is identified as a chronic disease, contributing to female mortality across the globe [2]. According to recent reports, an estimated 268,600 new cases of invasive breast cancer are to be diagnosed in women in the US during 2019, in addition to 2,670 new cases to be diagnosed in men. From these cases, an estimated 42,260 breast cancer deaths (41,760 women, 500 men) will occur in 2019. Excluding lung cancer, breast cancer is the most frequently diagnosed cancer in women [3].

Early detection and appropriate treatment of breast cancer can significantly increase the chances of survival. Since it enhances the chances of successful treatment and completes recovery of the patient [4]. It is also shown that early detection of small lesions boosts prognosis and leads to a significant reduction in mortality rate [5].

In general, mammography is the best diagnostic technique for screening. It is one of the most reliable and effective ways of detecting breast cancer in early stage [6]. However, there is a difficulty to detect masses because sometimes masses (areas of interest) seemed to be similar to normal breast tissues on mammograms [7]. Therefore, various computer aided diagnosis (CAD) systems have been developed for assisting doctors to identify the symptoms in early stage by using mammography [6].

There is a growing demand for accurate and fast computing algorithms to segment the diseased regions from the mammogram [2]. Still, finding an accurate and efficient breast region segmentation technique remains a challenging problem in digital mammography [4]. Among these techniques, the detection of masses in mammography is the most challenging because of the mass poor contrast, ambiguous shape margins, complex shapes and being indistinguishable from surrounding parenchyma tissues [6].

Breast abnormalities are defined with wide range of features & can be easily missed or misinterpreted by radiologists while reading large amount of mammographic images provided in screening programs [7]. One standard solution is to assign two radiologists for interpreting the same mammogram. However, this would be costly in terms of time, money and effort. Another approach is to rely on Computer Aided Detection (CAD) algorithms to replace the second radiologist, acting as a "second opinion" to assist the radiologist in making more accurate decisions.

Over the past two decades, many attempts have been developed to assist the radiologists in the detection and diagnosis of masses by designing computer-aided tools for mammography interpretation. Image processing and intelligent systems are two mainstreams of computer technologies that have been constantly explored in the development of computer-aided mammography systems [8]. Some of these studies in the field of mass detection and segmentation are listed below:

Campanini R. et al. [1] performed a multiresolution overcomplete wavelet representation, codifying the image with redundancy of information. The vectors of the very-large space obtained were applied to the first SVM classifier. False candidates were then eliminated with a second cascaded SVM classifier. The sensitivity of the presented system was nearly 80% with a false-positive rate of 1.1 marks per image, estimated on images taken from the Digital Database for Screening Mammography (DDSM) database.

Cordeiro F. et al. [9] presented a semi-supervised segmentation algorithm based on a modification of an algorithm known as GrowCut to perform automatic segmentation once a region of interest is selected by a specialist. An automatic point selection process is designed based on the simulated annealing optimization method, avoiding the need of human intervention. In order to validate their proposal, the authors built an image classifier using a classical multilayer perceptron with Zernike moments to extract the segmented image features. This analysis employed 685 mammograms from IRMA breast cancer database. The authors claimed that the proposed technique could achieve a classification rate of 91.28% for images with fat tissues.

Wang H. et al. [10] proposed a framework to detect breast masses in digitized mammograms. In their research, the authors analyzed the progress of radiologist's mammography screening, then a series of visual rules based on the morphological characteristics of breast masses are presented and quantified by mathematical methods. The experiments are performed on Mammographic Image Analysis Society (MIAS) and Digital Database for Screening Mammography (DDSM) data sets. The authors claimed that the sensitivity reached to 92% at 1.94 false positive per image (FPI) on MIAS and 93.84% at 2.21 FPI on DDSM.

ISA N. and SIONG T. [6] presented an automated system for mass segmentation and detection in mammograms. Constraint region growing based on local statistical texture analysis was applied to detect and segment out the mass from the mammograms. The system was developed and evaluated with 322 mammograms from the MIAS Database. It is stated that the proposed technique had a sensitivity of 94.59% and the number of false positive per image of 3.90.

Junior G. et al. [2] proposed a method for the detection of mass regions on digitized mammograms through diversity analysis, geostatistical and concave geometry (Alpha Shapes). The detection rate for each feature extraction were evaluated using Support Vector Machine in both MIAS and DDSM databases, with 74 and 621 mammograms, respectively, all containing at least one mass region. The obtained results were 91.63% of detection rate and 0.86 false positive per image for the DDSM database. This paper is organized as follows: Section 2 will present the methodology used for employing the proposed algorithm, Section 3 discusses the implementation results and Section 4 concludes the article.

2. METHODOLOGY OF THE PROPOSED SCHEME

In general, the radiologist can recognize most of tumors in the mammograms especially in the case of fatty tissues. Unfortunately, other types with high density are elusive and hard to be detected [11-13]. The difference between fatty and dense tissues mammograms can be easily viewed as shown in Figure 1.

Accurate segmentation of breast masses has a great effect on the later classification stage. Therefore, it is crucial that the required ROI is extracted carefully to obtain precise feature values, which leads to proper classification [14]. In this research, an approach is proposed to solve the problem of overlapped tissues in mammograms to facilitate the processes of features extraction and classification.

Although the final goal is to segment the tumor from the overlapped tissues, several steps are needed for achieving this goal. The first stage of the applied methodology is data acquisition. The used dataset here is the Digital Database for Screening Mammography (DDSM) [15], which is widely used by researchers in the field of breast cancer images analysis. First, the acquired images are cropped in order to remove any extra-irrelevant regions that may interfere with process of segmentation. Then, the cropped images are enhanced to improve the accuracy of detection by applying histogram adjustment and adaptive

histogram equalization. Next, the resulted images are pre-segmented to locate any neighborhoods of pixels that may present potential Regions-of Interest (ROIs).

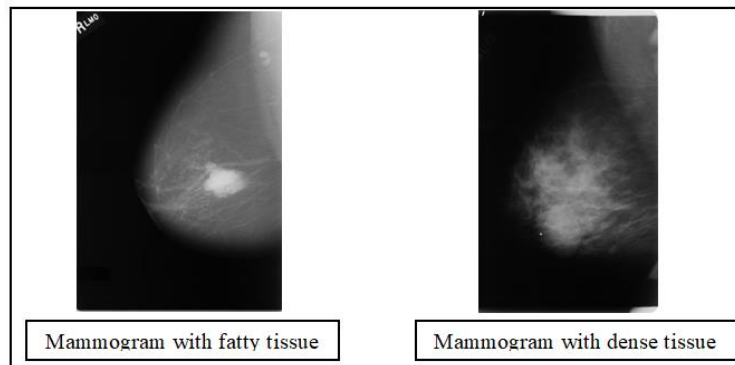


Figure 1. Fatty versus dense mammogram examples

In this stage, morphological operations are a useful for the refinement of any groups that are not adequate as mass candidates. This can be done throughout the process of morphological opening. Then, the suggested algorithm is applied on each pixel neighborhood to discover any bottleneck areas that may exist using the Euclidean distance among perimeter pixels. Finally, the region is divided at these areas into two or more regions in order to extract the mass candidate pixels from the rest of the neighborhood and present it as a potential ROI. These steps can be summarized by the Figure 2.

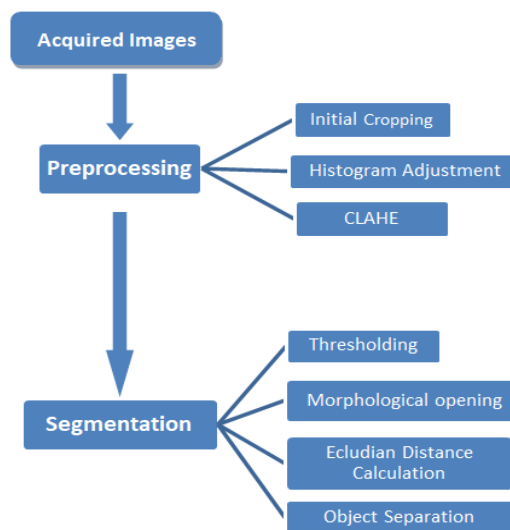


Figure 2. Steps required for implementation

3. SYSTEM IMPLEMENTATION

The steps illustrated in the previous section are further explained in this one. Such approaches usually include the two main stages shown in Figure 2 with minor steps to be conducted in each stage. These stages are given in more detail as follows:

3.1. Preprocessing

Before implementing the segmentation phase, some preprocessing steps should be applied on the image to facilitate the process of detecting the tumor's borders and extracting features. The first stage is initial cropping to delete any irrelevant areas in the image that has no contribution to the detection process. Then histogram adjustment [11, 16, 17] is used to utilize the entire histogram range (if not already utilized),

which increases the distance among histogram bins and hence the overall image contrast. The final preprocessing step is applying image histogram equalization (HE) to improve the contrast of the image, which increase separation between foreground and background gray level distribution.

Contrast-Limited Adaptive Histogram Equalization (CLAHE) [18-20] is utilized for the enhancement process, which instead of equalizing the entire histogram of the image like HE, works on parts of the image called tiles. CLAHE computes and equalizes the histogram of each tile apart then uses bilinear interpolation to restore the resulted image from these tiles in order to remove any artificial boundaries that may result from putting these tiles together.

The benefits of contrast enhancement are removing pixels out of intensity range, enhance the readability of areas with low contrast and produce images that can be easily analyzed by the system. All the steps of the preprocessing stage can be viewed in Figure 3.

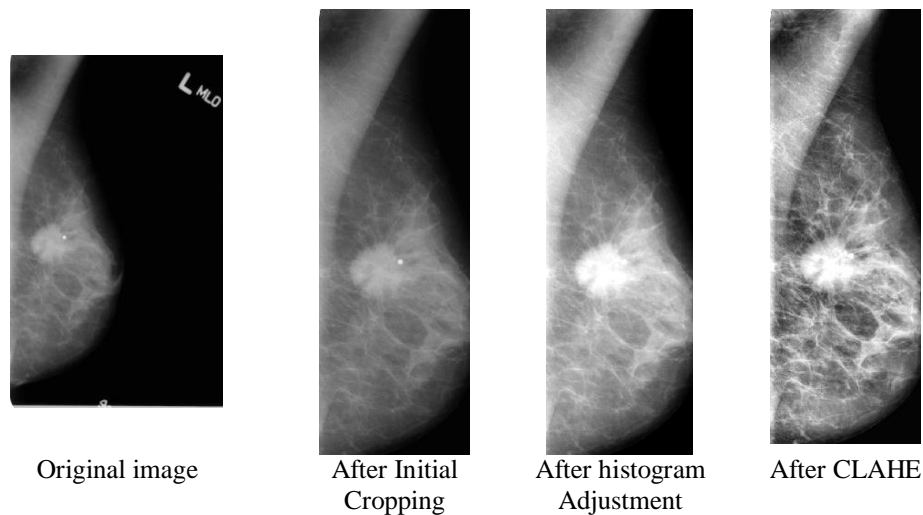


Figure 3. An example of the preprocessing stage

3.2. Segmentation

The images resulted from preprocessing are now ready to be processed for segmenting possible mass regions from the rest of the image. Segmentation requires multiple minor steps to be performed as to expose the overlapping sections which are the main focus of this research. The proposed image segmentation can be summarized by the following steps:

- a) Pre-segmentation: At first, a threshold is applied on the images resulted from preprocessing. Then mathematical morphologically [21-23] [is used to remove any small objects that do not qualify as potential mass regions. Morphological operations are used to analyze the shapes and textures in images. Suppose $I(s, t)$ be a gray scale image and S be a structuring element then Erosion (\ominus) and Dilation (\oplus) operations are defined as [21]:

$$\text{Erosion: } [I \ominus S](s,t) = \min_{(u,v) \in S} I(s+u, t+v) \quad (1)$$

$$\text{Dilation: } [I \oplus S](s,t) = \max_{(u,v) \in S} I(s-u, t-v) \quad (2)$$

From the previous equations, the opening morphological operation (\circ) is $I \circ S = (I \ominus S) \oplus S$. Similarly, the closing operation (\bullet) is $I \bullet S = (I \oplus S) \ominus S$. The resulted binary mask represent pixel neighborhoods of high intensity, since masses in mammograms appear to have higher intensity levels than the surrounding tissue [9, 19, 24].

- b) Now the image is converted into a mask with several objects only, each object is inspected individually to find the areas where overlapping is present. Perimeter pixels are arranged consecutively and the direct distances among these pixels are computed. Pixels at narrow areas (bottlenecks) are sequentially far from each other along the perimeter line, yet the direct displacement between them is small. Taking this into account, bottleneck areas can be automatically recognized, which represent the part of the mass region being overlapped by other tissues.

A criteria can be set for the detection of such spots that if the direct displacement between any two pixels is smaller than a given percentage of the successive distance on the perimeter line, a bottleneck may be present in that area. Figure 4 shows an example of bottleneck areas detection on one of the pixel neighborhoods.

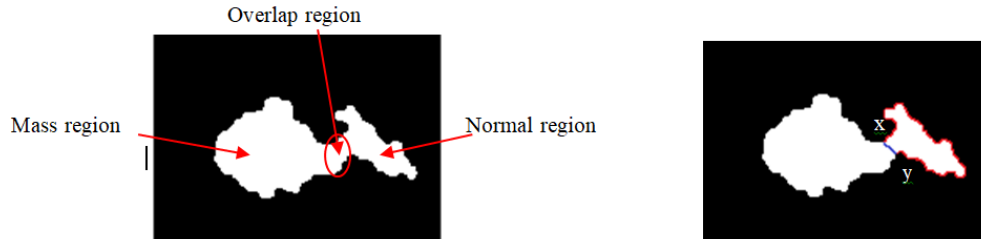


Figure 4. An example of a mass region being separated from normal tissue at the narrow region (a) shows overlapping location in the neighborhood (b) compares direct displacement line (in blue) with the long border line (in red)

From the figure above, it can be seen that in (b) the two points (x and y) are far from each other on the border (red) line, but they are close in terms of displacement (blue line). The algorithm automatically finds and selects the shortest displacement path in the area in terms of Euclidean Distance [9, 25] [9, and divides the region based on it. The resulted ROIs can resemble a mass region candidate more accurately than before the separation, which gives more accurate features next stage and thus improves the classification.

- c) The final step of this approach is to separate the neighborhood into two or more smaller regions, based on the line in narrow region determined from the previous step. The steps implemented for the segmentation stage can be viewed in Figure 5.

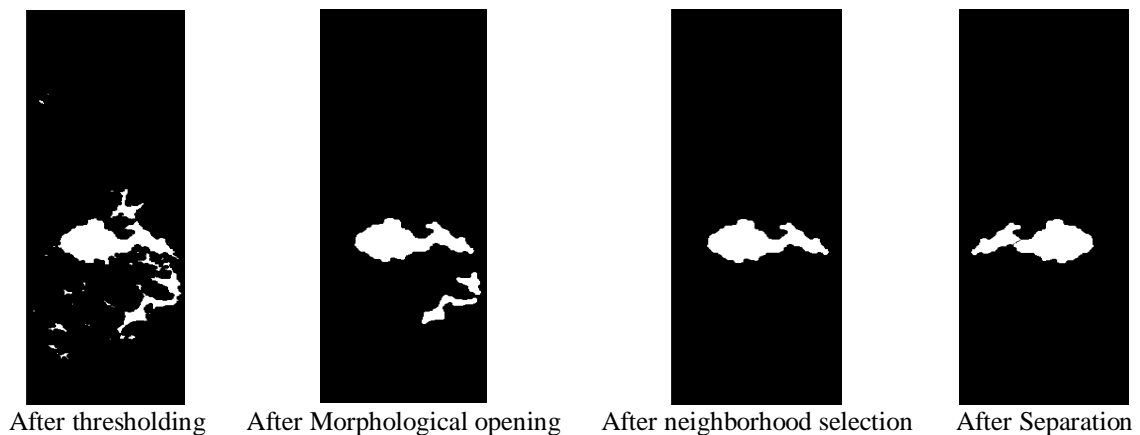


Figure 5. An example of the steps applied for segmentation

The intent of this separation is to isolate any possible mass region from other areas, which are marked as false positives. The resulted mass region (if there is any) may have a higher chance of correct classification than before separation due to the fact that the features extracted from such area will be more precise and will better describe the ROI's characteristics prior to classification.

4. RESULTS AND ANALYSIS

The procedure presented in this work is applied on 42 mammography images downloaded from the DDSM website. Among these images, 47 ROIs are formally confirmed by the database, which are to be considered as the ground truth over which the results will be compared against. The algorithm managed to correctly detect and segment 45 regions located in 40 images. This can be summarized as:

True Positive (TP) Images =40,

True Positive (TP) regions =45,

False Positive (FP) =14,

False Negative (FN) =2

From these results, Image Sensitivity (SN_{Image}), region Sensitivity (SN_{region}) and False Positive per Image (FPI) can be calculated as:

$$\text{SN Image} = \frac{TP(\text{Images})}{TP(\text{Images}) + FN} = \frac{40}{40+2} * 100\% = 95.238\%$$

$$\text{SN region} = \frac{TP(\text{regions})}{TP(\text{regions}) + FN} = \frac{45}{45+2} * 100\% = 95.744\%$$

$$\text{FPI} = \frac{FP}{\text{No.Images}} = \frac{14}{42} = 0.333$$

The regions missed by the algorithm are of rare unusual properties (having much dimmer levels of gray than other regions or being almost fully overlapped in extremely dense cases). On the other hand, false positives were resulted from the fact that the separation produces regions other than the main ROI. These extra regions have poor features that don't relate to a mass and can be filtered out in later stages.

5. CONCLUSION

In this work, various mammography images with overlapping cases were analyzed and processed by the presented algorithm. It is noticed that applying CLAHE in preprocessing gives better results compared to the general histogram equalization. This is due to the diverse nature of mammography images, which is more consistent with CLAHE. Implementation also showed that geometrical analysis of the tumor's shape have a positive influence on the detection accuracy, since many normal tissues may interfere with or be projected upon the mass region by the time of mammography image creation, distorting its shape and giving false parameters that deteriorate results of the later classification stage. On the other hand, parameters extracted from the isolated regions have more accurate values that can truthfully describe the mass regions and facilitate the classification process.

REFERENCES

- [1] Campanini R. *et al.*; "A novel featureless approach to mass detection in digital mammograms based on Support Vector Machines", *Physics in Medicine and Biology*, Vol. 49, pp. 961-975, 2004.
- [2] Junior G. *et al.*; "Breast cancer detection in mammography using spatial diversity, geostatistics, and concave geometry", *Multimedia Tools and Applications*, Vol.78, No.10, June 2018.
- [3] Cancer Facts & Figures 2019 by Corporate Center: American Cancer Society Inc.
- [4] Patel B. and Sinha G.; "Reliable Computer-aided Diagnosis System using Region based Segmentation of Mammographic Breast Cancer Images", *Mathematical Methods and Systems in Science and Engineering, Proceedings of the 17th International Conference on Mathematical Methods, Computational Techniques and Intelligent Systems (MAMECTIS '15)*, Tenerife, Canary Islands, Spain, 2015.
- [5] Sundaravinayagam S. and Sankari B.; "Detection and Classification of Masses in Mammograms Using a Hybrid GA-PSO-KNN Approach", *International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)*, Vol.2, No. 5, pp.36-40, 2015.
- [6] Isa N. and Siong T.; "Automatic Segmentation and Detection of Mass in Digital Mammograms", *Recent Researches in Communications, Signals and Information Technology, Proceedings of the 11th International Conference on Telecommunications and Informatics (TELE-INFO '12) Proceedings of the 11th International Conference on Signal Processing (SIP '12)*, Saint Malo & Mont Saint-Michel, France April, 2012.
- [7] Gaikwad V.; "Detection of Breast Cancer in Mammogram using Support Vector Machine", *International Journal of Scientific Engineering and Research (IJSER)*, Vol. 3, No.2, pp.26-30, 2015.
- [8] Martins L. *et al.*; "Detection of Masses in Digital Mammograms using K-means and Support Vector Machine", *Electronic Letters on Computer Vision and Image Analysis*, Vol. 8, No. 2, pp. 39-50, 2009.
- [9] Cordeiro F. *et al.*; "A Semi-Supervised Fuzzy GrowCut Algorithm to Segment and Classify Regions of Interest of Mammographic Images", *Expert Systems with Applications*, Vol.65, pp. 116-126, December 2016.
- [10] Wang H. *et al.*; "Breast Mass Detection in Digital Mammogram Based on Gestalt Psychology", *Journal of Healthcare Engineering*, Vol. 2018, Article ID 4015613, May 2018.
- [11] Ogiela M and Krzyworzeka N.; "Heuristic approach for computer-aided lesion detection in mammograms", *Soft Computing*, Vol.20, No.10, pp. 4193-4202, October 2016.
- [12] Radovic M. *et al.*; "Parameter optimization of a computer-aided diagnosis system for detection of masses on digitized mammograms", *Technol Health Care*, Vol. 23, No.6, pp. 757-774, 2015.

- [13] Carneiro P. *et al.*; "Breast density pattern characterization by histogram features and texture descriptors", *Research on Biomedical Engineering*, Vol. 33, No.1, pp. 69-77, 2017.
- [14] Ciecholewski M. ; "Malignant and Benign Mass Segmentation in Mammograms Using Active Contour Methods", *Symmetry*, Vol. 9, No.277, November 2017.
- [15] M. Heath *et al.*; "*The digital database for screening mammography*", in Proceedings of the 5th International Workshop on Digital Mammography, pp. 212–218, Toronto, Canada, June 2000.
- [16] Cruz C.; "Automatic Analysis of Mammography Images: Enhancement and Segmentation Techniques", M.Sc. thesis, Engineering Faculty, Porto University, Porto, 2011.
- [17] Dong M. *et al.*; "An Efficient Approach for Automated Mass Segmentation and Classification in Mammograms", *Journal of Digital Imaging*, Vol.28, No.5, October 2015.
- [18] Alhaj A. and Hussein E.; "Detection and Recognition of Breast Cancer using Neural Network", *International Journal of Engineering, Applied and Management Sciences Paradigms*, Vol. 45, No. 1, pp. 116-121, 2017.
- [19] Al-Juboori R.; "Contrast Enhancement of the Mammographic Image Using Retinex with CLAHE methods", *Iraqi Journal of Science*, Vol. 58, No.1B, pp. 327-336, 2017.
- [20] Zhang Y. *et al.*; "Smart detection on abnormal breasts in digital mammography based on contrast-limited adaptive histogram equalization and chaotic adaptive real-coded biogeography-based optimization", *Simulation*, Vol. 92, No. 9, September 2016.
- [21] Thawkar S. and Ingolikar R.; "Automatic Detection and Classification of Masses in Digital Mammograms", *International Journal of Intelligent Engineering and Systems*, Vol.10, No.1, 2017.
- [22] Vikhe P. and Thool V.; "Morphological operation and scaled Rényi entropy based approach for masses detection in mammograms", *Multimedia Tools and Applications*, Vol.77, No.18, pp 23777–23802, 2018.
- [23] Ancy A. and Nair L.; "*An Efficient CAD for Detection of Tumour in Mammograms using SVM*", International Conference on Communication and Signal Processing (ICCSP), 2017.
- [24] Li Y. *et al.*; "A bilateral analysis scheme for false positive reduction in mammogram mass detection", *Computers in Biology and Medicine*, Vol. 57, No.1, pp. 84-95, February 2015.
- [25] Angayarkanni N. and Kumar D.; "Mammogram Breast Cancer Detection Using Pre-processing Steps of Binarization and Thinning Technique", *Pensee Journal*, Vol 76, No. 5, May 2014.

BIOGRAPHIES OF AUTHORS



Mustafa H. Mohammed Alhabib completed BSc. degree from the Department of Computer Engineering \ College of Engineering \ university of Mosul in 2008, obtained MSc degree in Technical Computer Engineering from the Technical Engineering College\ Mosul in 2013. Specialized in Image Processing and Artificial Intelligence fields. Currently lecturing at the department of Communications and Computer Engineering at cihan University, Erbil, KRG of Iraq. Published 5 papers in the fields of Medical Image Processing and pattern recognition.



Omar H. Al-Habeeb received his master degree from faculty of computer science, UTM University, Malaysia. He received his B.S. degree from faculty of computer science in Mosul University, Iraq. He is currently a lecturer in the faculty of software engineering at Mosul University, Mosul, Iraq. His research interests include computer networks, security in mobile communications and image processing.