Texture Analysis of Mammogram Using Histogram of Oriented Gradients Method

Athraa H. Farhan and Mohammed Y. Kamil *
College of Sciences, Mustansiriyah University, Baghdad, Iraq.
* Email: m80y98@uomustansiriyah.edu.iq, ORCID: orcid.org/0000-0001-5709-2549

Abstract. The second foremost reason for dying ladies all across the world is breast cancer. The possibilities of survival can be raises when cancer detects earlier; therefore, the mortality reduction. The radiologist used mammograms to recognize breast tumors at an early level. Since the mammograms have little contrast, hence, it is unclear to the radiologist to distinguish small tumors. A computer-aided detection system contributes to explaining mammograms and helps the radiologist to indicate the appearance of breast mass and discriminate among normal and abnormal tissue. In this research, we introduce a histogram of oriented gradients as a method of feature extraction and identify mass regions in mammograms. The features extraction from this method classified by a support vector machine. To improve the diagnosis ability, contrast limited adaptive histogram equalization pre-processing system is utilized. Mini-MIAS database used to estimate this approach. The top accuracy, sensitivity, and specificity obtained are 90%, 69%, 100%, respectively.

1. Introduction

Breast cancer is an uncontrolled increase in irregular cells of breast tissue. It is the popular reason for death among women in their 40s[1]. Due to medication developments, earlier detection, and consciousness, the fatality decreased amongst women during the latest two decades[2]. The detection for masses in breast tissue in the beginning stages is the most suitable instance to raise the possibilities of survival. Accordingly, women in the 40s or earlier are advised to take mammograms frequently [3]. Mammography represents a particularly significant part in cancer detection. It is a special kind of imaging that employs a low-dose of X-ray, and it differs in four regions of the grayscale: background, fat tissue, breast parenchyma, and masses with increasing gray scale[4]. Great experience and large practices of radiologists are wanted when examining breast tissue in a mammogram. Therefore, to improve the accuracy of cancer detection, the CAD was produced as another reader. CAD system assists in determining the mass in breast tissue in mammograms. After discovering of masses, this system illuminating suspicious regions in the initial image with the inclusion of more interpretation by the radiologist[5]. The most reliable method to recognize breast cancer is to extract features from mammograms. Features are labeled as color, shape, and texture[6]. Here, we are focused on the texture features and to extract features; HOG is used. There are many types of research in the field of mammogram analysis for identifying the tumors including the feature extraction method[7]. S. Krishnaveni et al. (2014)[8] introduced (HOG) as a Feature extraction approach and Naive Bayes classifier are employed to recognize microcalcification in the mammogram. The examination of the suggested method achieved 96.5% mass diagnosis mammograms. V. Pomponiu et al. (2014) [9] offered an uncomplicated method applied to filter the output of the automated tumor discovery systems, (HOG) that is a method used for filtering the tumor and whole tissue areas. The sensitivity of this method is high. M. Abdel-Nasser et al. (2016)[10] investigated the behavior of different texture analysis techniques to lessen FP in breast tumor diagnosis; histogram of oriented gradients is one of
these methods. The sensitivity obtained from Hog 81.80 ±0.0214 and Specificity 92.45 ±0.0473. K. C. Tatikonda, et al. (2018)[11] connected algorithm of simplistic (HOG) and (GLCM) technique for breast tumor diagnosis and classification. The suggested approach achieved an accuracy of 99.11%. In this paper, we transact with the field of extract features from mini MIAS database images. We aim to implement features that provide important information about images. Histogram of oriented gradients is the method used to lessen the false positive diagnosis and develop the performance of the computer-aided detection systems. ROI extracted manually from the region inside the ground truth which specified automatically in the mini MIAS images and then we apply hog to ROIs. After generating the HOG features, we classify them into exact mass and healthy tissue.

2. Methodology
The levels of breast cancer diagnosis in the CAD system are involved in preprocessing, feature extraction, and classification[12]. The following steps are used in this research to determine normal and abnormal tissue.

2.1. Determining ROI
The ground truth (GT), as provided by mini- MIAS database[13] is according to the x-axis and y-axis and the radius that included the mass region, ground truth represents as a circular region. ROI determined inside GT as a square region, and this region is needed for extract features. Select ROI has shown in Figure 1.

2.2. Preprocessing.
The mammograms include various types of noise and undesired data that affect to the result of classification. Accordingly, the pre-processing techniques are important to enhance the nature of the mammogram and perform the feature extraction method more positively[14]. In this study, the Contrast Limited Adaptive Histogram Equalization (CLAHE) method used as the preprocessing technique, which is a successful method for image enhancement because of its integrity and relatively more reliable performance for any kind of images [15]. The CLAHE method produces high-grade flexibility in determining the social histogram mapping function. This method splits mammograms into proper quarters and implements him to them. In this method, the grayscale values are transformed by applying a nonlinear methodology to increase the contrast of grayscale of a mammogram[16]. Figure 2 shows the mammogram before and after applying the CLAHE method.

2.3. Feature Extraction, Histogram of Oriented Gradients (HOG).
The technique of forming a collection of characteristics that illustrates the details of the image and development of its description is the feature extraction technique. The feature extraction process reduces the data in a mammogram and modifies an image to a compressed description assortment of characteristics[17]. In this research, we used the histogram of oriented gradient (HOG) as a feature extraction technique.
HOG is an object descriptor introduced by Dalal and Triggs[18] focalize on the structure or appearance of an object in an image. Histogram of Oriented gradients provides distinguishing features when lighting variation and background noise, so it is an effective descriptor. HOG technique calculated the occurrence of gradient directions in ROI[19]. This technique describes the structural shape within ROI by capturing the local grey levels gradient. The directional variation in gray levels is a gradient. Because the value of gradients throughout edges is high, it is utilized to derive meaningful information from ROI. In the HOG descriptor, histograms do not create for entire ROI, ROI is divided into small related regions called a cell and the histogram of gradients is calculated for every cell[20].

Figure 2. Sample of: (a) image before enhancement (b) image after enhancement.

To implement the HOG descriptor, starting with dividing ROI into cells and obtain gradient for each cell in x and y-direction[21]. The gradient is defined as:

$$G_x = I * M_x , \quad G_y = I * M_y$$

(1)

Where I is ROI, $M_X$ is the mask in the x-direction, $M_Y$ is the mask in the y-direction.

The final gradient value of ROI is:

$$|G| = \sqrt{G_X^2 + G_Y^2}$$

(2)

The gradient orientation can be defined as:

$$\theta = \arccos \left( \frac{G_Y}{|G|} \right)$$

(3)

According to the gradient is ‘unsigned’ or ‘signed’, the histogram bins are equally separated to 0–π or 0–2π. The summation of the gradient value of the gray level, which has edge orientation within the frontiers of the bin is the magnitude of the bin [22].the grayscale value is computed for larger overlapping regions denominated blocks. These blocks are produced by accumulated cells in ROI. The cells inside the blocks are normalized by computed the grayscale value for each block and applying the contrast normalization process. Next, the Hog features derived from the normalize ROI[23]. The implementation of the HOG technique is shown in Figure3.

2.4. Classification.

In this investigation, a support vector machine (SVM) employed as a classifier. SVM is an assortment of an associated administered training technique that examines data and identifies exemplars employed in the analysis. This classifier uses input data and foretells two probable forms for which input is part[24]. SVM performs high efficiency compared to different varieties of systems.

3. Results and discussion

The offered approach is examined by the mini-MIAS database, which available on the internet [13] that consists of 322 images divided into 209 normal, 51 malignant, and 62 benign. The size of every image in the database is 1024×1024 pixels. Mammograms are classified by the character of background tissue (Fatty, Fatty-glandular, Dense-glandular) and the kind of exact mass (ill-define, normal, asymmetry, architectural distortion, speculated, circumscribed, and calcification)[25]. In mini-MIAS, mass regions are selected automatically
(ground truth), these regions are manually cropped to get ROIs inside the ground truth in three sizes (10×10, 20×20, 30×30) pixels according to the dimensions of mass in mammograms. Normal cases of mammograms, ROIs are determined manually from arbitrary regions with the same sizes in abnormal cases. After that, we enhance selecting ROI by applying the CLAHE method to increase the accuracy of the diagnosis of breast tumor then applied the HOG descriptor to all ROIs of all images. The HOG feature representation is illustrated in Figure 4, the direction of all cells is reached. This viewing to edge directions can be better to understand the contours and shape encoding.

![Figure 3. HOG descriptor for a region of interest of (10×10) pixel.](image)

- architectural (AR)
- asymmetry (AS)
- calcification (CA)
Figure 4. Samples HOG representation for kinds of tumors.

Four features we obtained when applying the HOG technique are contrast, correlation, energy, and homogeneity. The input ROI is categorized to normal or abnormal by applying the SVM classifier. The SVM is practiced with the derived features and using the practiced magnitudes. Testing features distinguish the normal tissue and the mass. Tables 1, 2, and 3 display the performance of the suggested method with SVM and 5-fold cross-validation.

The accuracy ranged between growing up and growing down around ±20 for different breast tumors. This extension is due to the different types of lesions. The highest accuracy value is observed 90% in the region of interest of (30×30) for calcification tumors. The method is given high results in the diagnosis of abnormality tissue is a disease. Table 3 shows the highest specificity value is observed 100% for calcification tumors too. Nevertheless, it obtained intermediate results in the diagnosis of normal tissue is healthy. Table 2 illustrates the highest sensitivity value is observed 69% for architectural tumors.

Table 1. The accuracy produced by using SVM classifier.

| Type ROI | AR  | AS  | CA  | CI  | MI  | SP  |
|----------|-----|-----|-----|-----|-----|-----|
| 10 x 10  | 70.6| 71.4| 72.9| 70.0| 76.6| 71.7|
| 20 x 20  | 78.0| 80.0| 80.4| 75.5| 83.3| 71.4|
| 30 x 30  | 81.1| 78.0| 90.0| 71.4| 79.6| 78.0|

Table 2. The Sensitivity produced by using SVM classifier.

| Type ROI | AR  | AS  | CA  | CI  | MI  | SP  |
|----------|-----|-----|-----|-----|-----|-----|
| 10 x 10  | 29.0| 13.0| 21.0| 25.0| 46.0| 53.0|
| 20 x 20  | 59.0| 33.0| 62.0| 50.0| 44.0| 38.0|
| 30 x 30  | 69.0| 31.0| 62.0| 47.0| 25.0| 31.0|
Table 3. the Specificity produced by using SVM classifier.

| Type | Specificity % |
|------|---------------|
| ROI  | AR | AS | CA | CI | MI | SP |
| 10 x 10 | 91.0 | 97.0 | 94.0 | 91.0 | 88.0 | 82.0 |
| 20 x 20 | 88.0 | 97.0 | 88.0 | 88.0 | 94.0 | 73.0 |
| 30 x 30 | 84.0 | 95.0 | 100 | 84.0 | 97.0 | 95.0 |

4. Conclusion
The histogram of the oriented gradient is an extremely successful technique employed in this study to extract features from ROIs and improve the diagnosis of breast tumors. To enhance the ROIs and improve its character, we applied CLAHE as a pre-processing technique. After this stage, mass tissue can be analyzed by using the feature extraction system. We derived contrast, correlation, energy, and homogeneity as the features from ROIs by using the HOG technique. Better results can get when the normalized histogram of grade direction and thus the false positive frequencies decrease. To identify breast tissue as normal or abnormal, SVM is used. This classifier gets the best results where the accuracy, sensitivity, specificity is 90%, 69%, 100%, respectively. The accuracy of the diagnosis of breast cancer in the CAD system can be improved by the combination Hog method with other feature extraction methods.

References
[1] Nabeel F. Lattoofi, Israa F. Al-sharuee, Mohammed Y. Kamil, Ayoob H. Obaid, Aya A. Mahidi, Ammar A. Omar and Saleh A k 2019 Melanoma Skin Cancer Detection Based on ABCD Rule First International Conference of Computer and Applied Sciences (1st CAS2019)
[2] Kamil M Y 2016 Morphological gradient in brain magnetic resonance imaging based on intuitionistic fuzzy approach. In: 2016 Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA): IEEE) pp 1-3
[3] Mousa D S A, Mello-Thoms C, Ryan E A, Lee W B, Pietrzyk M W, Reed W M, Heard R, Poulos A, Tan J and Li Y J A r 2014 Mammographic density and cancer detection: does digital imaging challenge our current understanding? 21 1377-85
[4] Silverstein M J, Lagios M D, Recht A, Allred D C, Harms S E, Holland R, Holmes D R, Hughes L L, Jackman R J and Julian T B J o t A C o S 2005 Image-detected breast cancer: state of the art diagnosis and treatment 201 586-97
[5] Padmanabhan S and i Sundararajan R Enhanced Accuracy of Breast Cancer Detection in Digital Mammograms using Wavelet Analysis, fN 47907. USA 978-1-4673-2322-2112/$31.00, 2012 IEEE)
[6] Pathak B, Barooah D J I J o A R i E, Electronics and Engineering I 2013 Texture analysis based on the gray-level co-occurrence matrix considering possible orientations 2 4206-12
[7] Saleh A M and Kamil M Y 2018 Mammographic Image Segmentation Based on Fuzzy Morphological Operations. In: 2018 1st Annual International Conference on Information and Sciences (AiCIS): IEEE) pp 40-4
[8] Krishnaveni S, Bhanumathi R, Pugazharasan T J I J o A R i E, Electronics and Engineering I 2014 Study of Mammogram Microcalcification to aid tumour detection using Naive Bayes Classifier 3
[9] Pomponiu V, Hariharan H, Zheng B and Gur D 2014 Improving breast mass detection using histogram of oriented gradients. In: Medical Imaging 2014: Computer-Aided Diagnosis: International Society for Optics and Photonics) p 90351R
[10] Abdel-Nasser M, Moreno A, Puig D J J o E and Intelligence T A 2016 Towards cost reduction of breast cancer diagnosis using mammography texture analysis 28 385-402
[11] Tatikonda K C, Bhuma C M and Samayamantula S K 2018 The Analysis of Digital Mammograms Using HOG and GLCM Features. In: 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT): IEEE) pp 1-7
[12] Rouhi R, Jafari M, Kasaeei S and Keshavarzian P J E S w A 2015 Benign and malignant breast tumors classification based on region growing and CNN segmentation 42 990-1002
[13] SUCKLING J P J D M 1994 The mammographic image analysis society digital mammogram database 375-86
[14] AlQoud A and Jaffar M A J I J o C S N S 2016 Hybrid gabor based local binary patterns texture features for classification of breast mammograms 16 16
[15] Al-Juboori R A L J I J o S 2017 Contrast enhancement of the mammographic image using retinex with CLAHE methods 58 327-36
[16] Mohan S and Ravishankar M 2012 Modified contrast limited adaptive histogram equalization based on local contrast enhancement for mammogram images. In: International Conference on Advances in Information Technology and Mobile Communication: Springer) pp 397-403
[17] Hussain M, Khan S, Muhammad G, Berbar M and Bebis G 2012 Mass detection in digital mammograms using gabor filter bank
[18] Dalal N and Triggs B 2005 Histograms of oriented gradients for human detection.
[19] Abdel-Nasser M, Moreno A, Abdelwahab M A, Saleh A, Abdulwahab S, Singh V K and Puig D 2019 Matching Tumour Candidate Points in Multiple Mammographic Views for Breast Cancer Detection. In: 2019 International Conference on Innovative Trends in Computer Engineering (ITCE): IEEE) pp 202-7
[20] Skibbe H and Reisert M 2012 Circular Fourier-HOG features for rotation invariant object detection in biomedical images. In: ISBI, pp 450-3
[21] Tasdemir S B Y, Tasdemir K and Aydin Z 2018 ROI Detection in Mammogram Images Using Wavelet-Based Haralick and HOG Features. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA): IEEE) pp 105-9
[22] Ren H and Li Z-N 2014 Object detection using edge histogram of oriented gradient. In: 2014 IEEE International Conference on Image Processing (ICIP): IEEE) pp 4057-61
[23] Christobel J S and Kumar C V J I J o S R 2017 Feature based breast cancer detection using mammographic images 257-62
[24] Wang Z, Yu G, Kang Y, Zhao Y and Qu Q J N 2014 Breast tumor detection in digital mammography based on extreme learning machine 128 175-84
[25] Kamil M Y, Salih A M J I J o I E and Systems 2019 Mammography Images Segmentation via Fuzzy C-mean and K-mean 12 22-9