

Hybrid Gabor based Local Binary Patterns Texture Features for classification of Breast Mammograms

Amal AlQoud^{1†} and M. Arfan Jaffar^{††},

College of Computer and Information Sciences, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

Summary

Breast cancer is one of the most widespread cancers in the world. A digital mammogram is an X-ray picture of the breast, which used for early detection of breast cancer and other abnormalities. The early detection of the cancer can reduce mortality. Mammograms help radiologists to detect the malignant masses precisely at their early stage. But, the mammograms are usually characterized by low contrast therefore it is not easy to read by the radiologist to detect small masses and micro-calcifications, which are indirect signs of malignancy, precisely and in their early stage. So we need a Computer aided diagnosis (CAD) system for mammography, which simulates the process of the radiologist and used to interpret mammography image and check for the presence of breast cancer and distinguish between malignant and benign tumors of breast cancer. In this paper, we introduce an approach for mammogram breast cancer detection system. This system involves three phases. : preprocessing of mammogram image, features extraction, and classification. Texture features based upon Gabor filters with Local Binary Patterns has been extracted for classification.

Key words:

Hybrid Features, LBP, Gabor, Classification.

1. Introduction

Breast cancer is one of the most common cancer in women worldwide. It is a complex disease that will affect one in nine women. Breast cancer can also infect men, but it isn't common. However, early detection of breast cancer improves the chances of survival with successful treatment of breast cancer. A digital mammogram is used for early detection of breast cancer and other masses.. Computer aided diagnosis (CAD) for mammography process is the same radiologist process. It is used to interpret the mammogram image and check for the exits of breast cancer.

Mammograms are poor quality images with low-contrast. Therefore, it is difficult to detect small masses and calcification. The performance of classification depends upon the effectiveness of features the features that characterize a specific region are calculated and the important features amongst those are selected for the classification of the mass as normal or abnormal and benign or malignant. Feature extraction has attracted the attention of many researchers, and many techniques have

been proposed so far. Texture properties can be used to collect information about micro-patterns like edges, lines, spots and flat areas, which help to detect normal and lesions region in mammograms.

Feature extraction has attracted the attention of many researchers. Timp et al. [1] created two kinds of features: similar features and different features. Similar features measured the ability of the two regions to compare in appearance. Different features measured amount changing in feature values between corresponding prior regions and the current regions. Bellotti et al. [2] used means of textural features computed from the gray level co-occurrence matrix (GLCM) to characterize ROI. Varela et al. [3] used filter output, texture, gray level, contour-related and morphological features for extracting features. Rangayyan et al. [4] obtained shape features from the turning angle functions of contours. Wei et al. [5] distinguished between masses from normal breast tissues through extraction texture features from wavelet coefficients. Lad et al. [6] used Local Binary Pattern (LBP) descriptor to represent textural properties. Sampaio et al. [7] extracted texture using geo-statistic functions and used SVM for classification. Lahmiri and Boukadoum [8] used Gabor filters along with discrete wavelet transform (DWT) for mass detection and extracted statistical features (mean and standard deviation) from the Gabor images. Hussain et al [9]. Proposed a method based on textural properties. They used texture proprieties to represent masses by using a Gabor filter bank on different scales and orientations. They observed the outperform performance of their approach. Ojala et al [10] used LBP histogram to achieve rotation texture classification. Their approach of description pattern achieved efficiently performance. Ojala also used LBP in difference representation of a local pattern. In [11] Gargouri et al used using signed gray level difference (SGLLD), (CGLLD), and (MGLLD) to describe the texture features. Their approach leads to the best performance comparison with other methods.

2. Proposed Method

In this section, we discuss all stages of our method for breast cancer detection. The first step involves an image

enhancement, which is used to improve an image. The next step involves a feature extraction procedure. This step is an intensive task in our CAD system. Feature extraction includes: 1) application of Gabor filters for representat image to different scales and orientations, 2) application of LBP each image, separately. The next stage involves the classification using ANN classifier.

2.1 Preprocessing

Mammogram images contain different kind of information which is unwanted for the algorithm and may cause to misguide the algorithm of classification. In addition, mammograms are unclear images and low contrast. Therefore, preprocessing phase is necessary to improve the quality of the image and make the feature extraction phase more reliable. Preprocessing may compose of the two steps: - Background removal, and image enhancement by CLAHE. Background removal minimizes the computational cost and also finds the ROI (region of interest) while image enhancement increases the quality of the image. The two preprocessing steps are applied to original image. At first, the background removal is applied shown in fig 1-b, then image enhancement is applied to improve the quality of the image shown in fig 1-c.

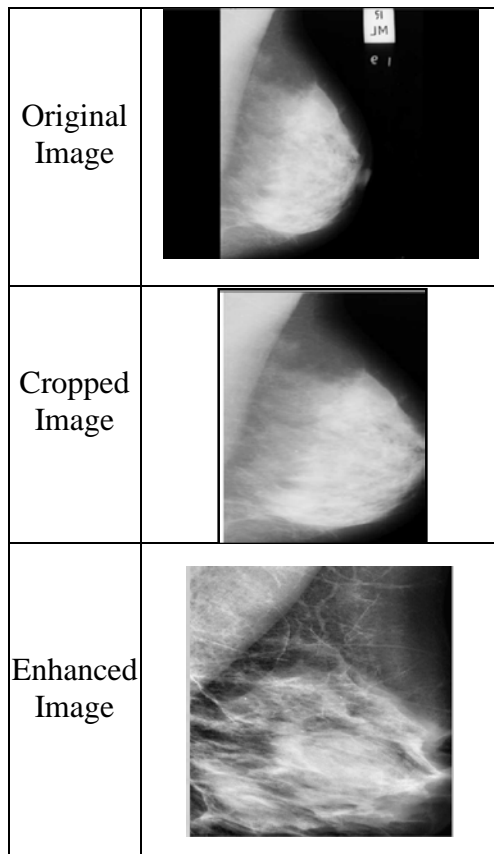


Fig. 1 Proposed beam former

2.2 Features Extraction

Feature extraction is the process of generating a set of features (commonly called vectors) which describes the contents of the image and constitutes representation of the image. The feature set is important for classification tasks. Mammogram images contain a huge amount of heterogeneous information that represents different tissues. The feature extraction phase will reduce the dimensionality in mammogram images. It transforms the mammogram images into a reduced representation set of features [12]. Texture analysis is used widely in image processing. In this study, we use texture analysis to extract features from the mammogram in order to classify it into (normal and abnormal) and classify abnormal into (benign and malignant). Texture is a type of features described spatial relationship. Mammogram contains local spatial pattern at different scales and orientations. This pattern is useful to distinguish between mass and non-mass, and between benign and malignant masses. Feature extraction techniques embody features of the image as much discriminative information as possible. Recent studies have been used Gabor and LBP individually for feature extraction processing. In our procedure, we propose an innovative method for feature extraction. We use two techniques for feature extraction which include Gabor filter and LBP techniques in order to increase features extraction accuracy and get better results. The following sections include a description of each technique.

2.2.1 Gabor Filter Based Features

Gabor filter is used to extract texture features of the image. The procedure of the Gabor feature extraction includes three steps:

- Segmentation the ROI into several sub-windows
- Implementation the Gabor filter bank on each window
- Moments calculation such (means, standard deviation, and skewness) based features from the Gabor filter bank.

The Fig 2 shows the description of Gabor feature extraction

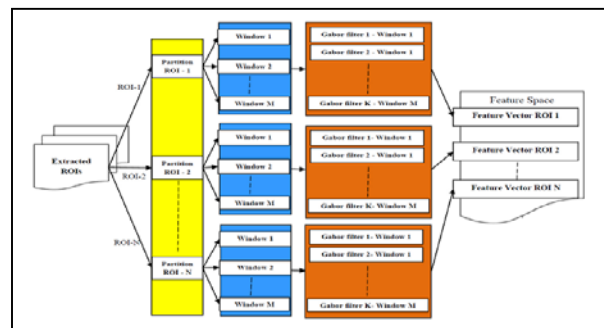


Fig 2 Gabor feature extraction

The characteristics of texture can be described by spatial frequencies and their orientations. Gabor filters have been successfully applied to a different of applications in the fields of image processing [9]. They provide a powerful tool to represent texture in different orientations and scales in order to describe textural properties of the image. The 2-D Gabor filter is a Gaussian kernel modulated by a sinusoidal wave of particular frequencies and orientation. The following equation describes the 2-D Gabor filter:

$$g(x, y) = \frac{1}{2\pi \sigma_x \sigma_y} e^{-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right)} e(2\pi jW \bar{x}) \quad (1)$$

$$\bar{x} = x \cdot \cos \theta + y \cdot \sin \theta \quad \text{and} \quad \bar{y} = x \cdot \sin \theta + y \cdot \cos \theta \quad (2)$$

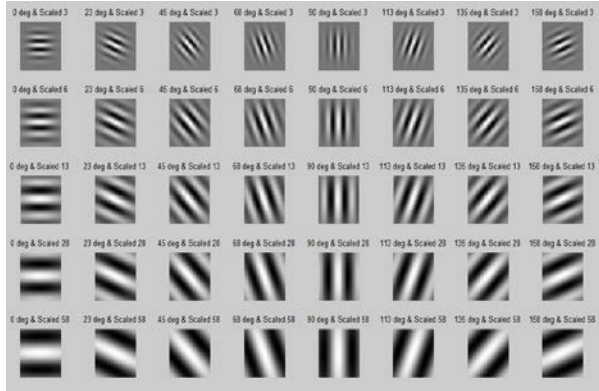


Fig 3 Gabor Filter with different scales and, different orientations

W is the central frequency of the complex wave and the orientation of the normal of the parallel stripes of the Gabor function. Different scales and, different orientations. We can describe all variable in this equation as the following: X, and Y are the spatial variables, σ_x and σ_y are the scaling parameters of the filter described the neighborhood of a pixel where the weighted summation takes place,

Gabor filters with different scales, orientation, and frequencies are called Gabor filter bank. Each filter of Gabor filter bank is described with orientations, scales, and frequency. Gabor filter is initialized with different frequencies in order to extract the pattern might be present at different scales and orientation in the image. [13]. Fig 3 shows 40 Gabor Filters with The filter bank orientations and frequencies are calculated by using the following equations:

$$\text{orientation}(i) = \frac{(i-1)\pi}{m} \quad (3)$$

$$\text{frequency}(i) = \frac{f_{\max} = 0.2}{(\sqrt{2})^{i-1}} \quad (4)$$

Where m is the total number of orientations, n is the total number of frequencies and I is a filter bank [13]. In our method we use Gabor bank with 8 filters (2 frequency \times 2 scales \times 2 orientations) as shown in Fig 4.

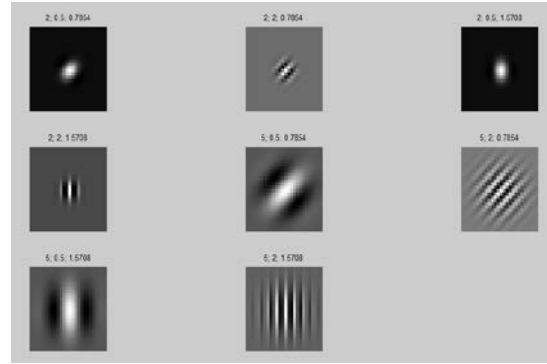


Fig 4 Gabor Filters with 2 orientations and 4 scales

We use a Gabor filter technique with different frequencies, scales, and orientations. We select frequencies= {2, 5}, scaling= {0.5, 2}, and orientation= {45, 90} in order to tune the filter to a particular pattern. Gabor filters are implemented for each window of an ROI to extract features. These extracted features are represented by three statistical measures (mean, standard deviation and the skewness) [9]. Fig 5 shows sub-windows of ROI. Each Gabor filter bank convolves with the original mammogram image to get eight different representations of same mammogram image where each image gives features represented by the three statistical measures. The features vector is useful for expression recognition because it contains all information extracted from different frequencies and orientations of the image as shown in fig 6

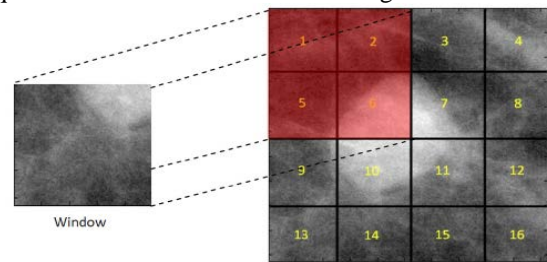


Fig 5 sub-windows of ROI

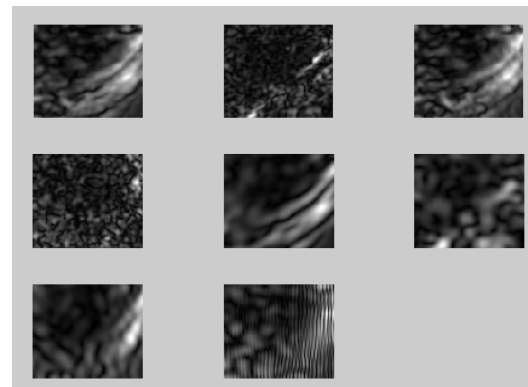


Fig 6 Gabor implantation on original image

The magnitude values of the Gabor transform change very slowly with displacement, so they can be further encoded. We encode the magnitude values with LBP operator in order to enhance the information in the Gabor filtered images. Applying local binary pattern on Gabor is actually a representation approach based on multi-resolution spatial histogram combining local intensity distribution with the spatial information, therefore, it is robust to noise and local image transformations. Additionally, instead of directly using the intensity to compute the spatial histogram, multi-scale and multi-orientation Gabor filters are used for the decomposition of an image, followed by the local binary patterns (LBP) operator. The Gabor and LBP combination further improves the representational power of the spatial histogram greatly. Note that, to make the GBLBP model, one does not need a training stage necessarily, which has obviously avoided the generalizability problem.

2.2.2 Local Binary Patterns based Features

After applying the Gabor filter on mammogram image, we use local binary pattern to describe local image patterns on different representations of the original mammogram image. Local binary pattern LBP is a type of feature transforms an image into an array. We apply the LBP operator on each pixel of the image to get LBP coded image. Fig 7 illustrates the work of LBP.

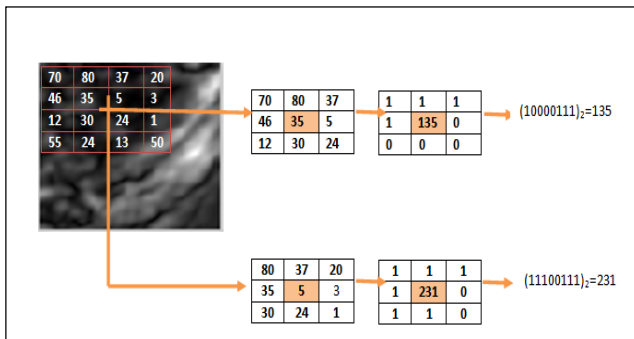


Fig 7 Example of basic LBP operator

As shown in fig 7 the LBP operator is done as the following:

- It uses eight neighbor pixels and considers the center value as a threshold.
- Then, it compares the center value with its neighbors. If the center value is greater than the neighbor's value, the operator generates 1. Otherwise, it generates 0.
- The output gives 8 binary numbers that represent the feature vector (Which is usually converted to decimal)

Fig 8 shows the result of implementation LBP on Gabor filter

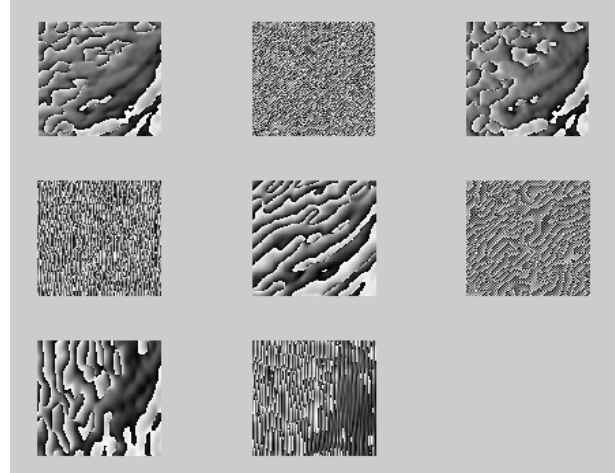


Fig 8 implementation of LBP on Gabor filter

2.3 Artificial Neural Network Based Classification

The features vector extracted in third phase can be used for classification tasks. The goal of image classification is to predict the categories of the input images using their features. In this study, we use Artificial Neural Network (ANN) as a classifier because it interprets image quickly and effectively. ANN classifier has the ability to detect and classify complex patterns from the given data. All features MIAS images which include normal and abnormal are used for training purpose. In addition, all abnormal images of MIAS which include benign and malignant are used for training. The backpropagation neural network model could distinguish normal mammogram from abnormal and benign from malign.

Artificial neural network ANN is a process of learning to separate samples into different classes by finding common features between samples of known classes. The features are presented to the network through the input layer. Therefore, the number of neural network nodes or nerous, which communicate with hidden layer, is equal to the number of features. The input data will be normalized before presented to ANN. In each node, a log sigmoid transfer function will be implemented. The ANN training processor is executed in hidden layer by a priori algorithm to discover association rules from features and to categorize these features. In other words, priori association rule algorithms describe frequent set of features per each classification (normal or abnormal, and benign or malignant).

The ANN training processor is executed until accept the maximum sum square error. The classification of each object is determined after training. First, we use all MIA's features as input of ANN and we use 1, and 0 as a target (0 for normal and 1 for an abnormal). Then we use just features of abnormal of MIAS and we use 1, and 0 as a

target (0 for benign and 1 from malignant). Fig 9 shows the ANN input and output.

3. Results and Discussion

The proposed method has been tested by using Mammographic Institute Society Analysis MIAS mini [14]; the mammograms have a size of 1024x1024 pixels, and resolution of 200 micron. This database is composed of 322 mammograms of right and left breast, from 161 patients, where 54 were diagnosed as being malignant, 69 benign and 207 normal. This database contains lists the mammograms in the MIAS database and provides appropriate details, for example, the class of abnormality, xy image-coordinates of centre of abnormality, and approximate radius (in pixels) of a circle enclosing the abnormality. MIAS is commonly used as a benchmark for testing new proposals dealing with processing and analysis of mammograms for breast cancer detection. Each case in this database is annotated by expert radiologists; the complete information is provided as an overlay file.

We have tested the performance of the SVM classifier by calculating and analysis of accuracy, sensitivity and specificity for malignancy detection. These are defined as follows:

Accuracy: number of classified mass/number of total mass

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

Sensitivity: number of correct classified malignant mass/number of total malignant mass

$$\text{Sensitivity} = (\text{TP}) / (\text{TP} + \text{FN})$$

Specificity: number of correct classified benign mass/number of total benign mass

$$\text{Specificity} = (\text{TN}) / (\text{TN} + \text{FP})$$

First preprocessing has been performed to remove the unwanted part of mammograms. These unwanted parts does not play any role in classification and this area is considered as a background. Thus first process is to remove background and results has been shown in Figure 1. In the second step, features extraction has been performed. For this purpose, texture features has been extracted. First Gabor filter has been applied to extract Gabor features from mammograms. To incorporate rotation invariant and scale invariance, Gabor features has been applied on different angles and scales as shown in Figure 5 (b). Thus it plays a major role in classification. Gabor filter has been already discussed in detail in previous section. Gabor has been applied at different angles and scales so it returns different images. Like in this case we have applied at 2 orientation and 4 scales so it returns eight different images. For Gabor features, we have calculated the average of all these eight different images and used for classification. The results has been shown in table 1 as a first row and it shows that it provide good accuracy as well as specificity. But still there is an error

for accuracy, sensitivity as well as specificity. To improve its performance, some more texture features required. For texture feature, LBP also shows promising role for classification. So we also tested LBP on all those eight Gabor images and then calculate the histogram of all LBP codes calculated at eight different images. These LBP codes can be used for classification. We also tested separate LBP on mammograms and test its performance and results has been shown in table 1.

Thus first LBP has been extracted directly from mammograms after preprocessing and results has been shown in Table 1 as a second row. But to combine LBP with Gabor, LBP has been applied on all eight Gabor images and calculate the histogram of all LABO codes for fused texture features. These fused features has been used for further classification. Results has been shown in Table 1 as a third row.

The whole dataset of features is randomly divided three times: at first, it is divided into 30-70. 30% is for training and 70% for hold on testing. At the second time, it is divided into 50-50. 50% is for training and 50% for hold on testing. At the third time, it divided into 70-30. 70% is for training and 30% for hold on testing.

The ROC curves are also drawn for each set of features. As in the Fig 9, 10, 11.

Table 1. Results of Gabor, LBP and Fused Features

	Training-Testing	Accuracy (%)	Sensitivity (%)	Specificity (%)
Gabor Features	30-70	94.77±3.22	93.17±2.46	94.04±1.93
	50-50	95.23±2.57	93.39±1.85	94.91±1.66
	70-30	95.97±3.64	94.97±3.64	95.36±3.56
LBP Features	30-70	96.14±2.47	95.04±2.39	95.37±2.69
	50-50	96.95±3.02	95.47±2.49	96.81±1.88
	70-30	96.82±2.96	96.73±2.32	96.03±2.47
Gabor + LBP Features	30-70	98.32±1.44	97.17±1.78	97.39±2.37
	50-50	98.75±1.17	97.05±1.49	97.45±1.52
	70-30	98.72±1.22	97.38±1.78	97.68±1.67

It can be seen in figure 11, the area under the ROC curve Gabor + LBP Features is the best.

4. Conclusion and Future Work

The major contributions of this paper is to fused Gabor and LBP texture features. Gabor filters are used along with discrete wavelet transform (DWT) for mass detection. Then, the local binary pattern features are extracted of

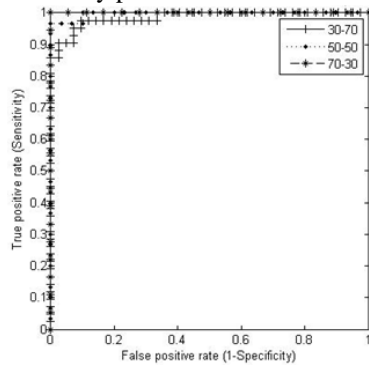


Fig ROC for Gabor Features

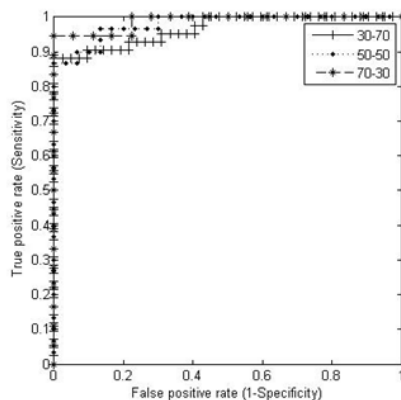


Fig 10 ROC for LBP Features

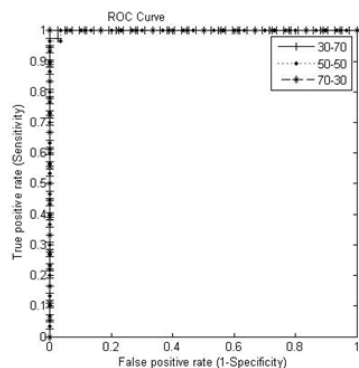


Fig 11 ROC for Gabor + LBP Features

Each filter. These texture features are robust due to rotation invariant as well as scale invariant. We use also an artificial neural network technique as a high tolerance

classifier which ability to classify patterns. Thus, our approach achieved more accurate and effective results. In future, we will extend our work with deep neural network.

References

- [1]. Timp, S., Varela, C., & Karssemeijer, N. (2007). Temporal change analysis for characterization of mass lesions in mammography. *IEEE Transactions on Medical Imaging*, 26(7), 945–953.
- [2]. Bellotti, R., De Carlo, F., Tangaro, S., Gargano, G., Maggipinto, G., Castellano, M., De Nunzio, G. (2006). A completely automated CAD system for mass detection in a large mammographic database. *Medical Physics*, 33(8), 3066–3075.
- [3]. Varela, C., Tahoces, P. G., Méndez, A. J., Souto, M., & Vidal, J. J. (2007). Computerized detection of breast masses in digitized mammograms. *Computers in Biology and Medicine*, 37(2), 214–226.
- [4]. Guliato, D., De Carvalho, J. D., Rangayyan, R. M., & Santiago, S. a. (2008). Feature extraction from a signature based on the turning angle function for the classification of breast tumors. *Journal of Digital Imaging*, 21(2), 129–144.
- [5]. Wei, D., Chan, H. P., Helvie, M. a, Sahiner, B., Petrick, N., Adler, D. D., & Goodsitt, M. M. (1995). Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis. *Medical Physics*.
- [6]. Lladó, X., Oliver, A., Freixenet, J., Martí, R., & Martí, J. (2009). A textural approach for mass false positive reduction in mammography. *Computerized Medical Imaging and Graphics*, 33(6), 415–422.
- [7]. Borges Sampaio, W., Moraes Diniz, E., Corrêa Silva, A., Cardoso de Paiva, A., & Gattass, M. (2011). Detection of masses in mammogram images using CNN, geostatistic functions and SVM. *Computers in Biology and Medicine*, 41(8), 653–664.
- [8]. Ahmiri, S., & Boukadoum, M. (2011). Hybrid discrete wavelet transform and Gabor filter banks processing for mammogram features extraction. In *2011 IEEE 9th International New Circuits and Systems Conference, NEWCAS 2011* (pp. 53–56).
- [9]. Hussain, M., Khan, S., Muhammad, G., Ahmad, I., & Bebis, G. (2012). Effective Extraction of Gabor Features for False Positive Reduction and Mass Classification in Mammography, 33(1), 29–33.
- [10]. Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971–987.
- [11]. Gargouri, N., Dammak Masmoudi, a., Sellami Masmoudi, D., & Abid, R. (2012). A New GLLD Operator for Mass Detection in Digital Mammograms. *International Journal of Biomedical Imaging*, 2012, 1–13. [12] Kom, G., Tiedeu, A., & Kom, M. (2007). Automated detection of masses in mammograms by local adaptive thresholding. *Computers in Biology and Medicine*, 37(1), 37–48.
- [12]. Hussain, M., Khan, S., Muhammad, G., Berbar, M., & Bebis, G. (2012). Mass Detection in Digital Mammograms Using, 1–6.
- [13]. <http://skye.icr.ac.uk/miasdb/miasdb.html>.