Research Article

The Impact of Pixel Resolution, Integration Scale, Preprocessing, and Feature Normalization on Texture Analysis for Mass Classification in Mammograms

Mohamed Abdel-Nasser,[1] Jaime Melendez,[2] Antonio Moreno,[1] and Domenec Puig[1]

[1]Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Avinguda Paisos Catalans 26, 43007 Tarragona, Spain
[2]Department of Radiology, Radboud University Medical Center, 6525 GA Nijmegen, Netherlands

Correspondence should be addressed to Mohamed Abdel-Nasser; egnaser@gmail.com

Texture analysis methods are widely used to characterize breast masses in mammograms. Texture gives information about the spatial arrangement of the intensities in the region of interest. This information has been used in mammogram analysis applications such as mass detection, mass classification, and breast density estimation. In this paper, we study the effect of factors such as pixel resolution, integration scale, preprocessing, and feature normalization on the performance of those texture methods for mass classification. The classification performance was assessed considering linear and nonlinear support vector machine classifiers. To find the best combination among the studied factors, we used three approaches: greedy, sequential forward selection (SFS), and exhaustive search. On the basis of our study, we conclude that the factors studied affect the performance of texture methods, so the best combination of these factors should be determined to achieve the best performance with each texture method. SFS can be an appropriate way to approach the factor combination problem because it is less computationally intensive than the other methods.

1. Introduction

Breast cancer was responsible for the largest number of cancer deaths among the EU females in 2014 [1]. Mammography is considered, in general, the most effective method for early detection of breast cancer and thus has been adopted for breast cancer screening. Computer-aided detection (CAD) systems are typically used to analyze mammograms in screening. While radiologists are generally pleased with the performance of CAD for clustered microcalcification detection, they have little confidence in CAD for mass detection. The most common complaint of radiologists is that CAD systems lead to a large number of false positives [2].

A breast cancer CAD system consists of three main stages: *segmentation* of a region of interest (ROI) from the mammogram, *feature extraction* from the ROI, and *classification*. Although mammography is a highly sensitive method for early detection of breast cancer, low specificity has been achieved in the classification of benign and malignant masses.

Texture analysis methods constitute one of the options for improving the specificity of classification algorithms applied to mammography. These methods may provide additional information in distinguishing benign and malignant masses. Although several feature extraction methods have been proposed for analyzing mammograms, improving the classification performance remains a challenging problem.

Texture analysis methods have been widely used to analyze mammographic images because they produce information about the spatial arrangement of intensities in the mammogram. Texture is one of the major mammographic characteristics for mass classification. For instance, several studies have used texture analysis methods to distinguish between normal and abnormal tissue [3–8] or to discriminate between benign and malignant masses [9–11]. Table 1 briefly summarizes some of this previous work. In addition, other studies have used texture analysis methods to estimate breast density [12] or to segment masses from mammograms [13].

TABLE 1: Summary of texture analysis methods that have been used to analyze mammograms.

Method	Extracted features	Utilized classifiers	Purpose
[6]	Local binary pattern (LBP)	Support vector machines (SVMs)	Classification of ROIs into mass/normal
[7]	Histogram of oriented gradients (HOG)	SVM	Classification of ROIs into mass/normal
[11]	Haralick's features (HAR)	k-nearest neighbour (k-NN)	Microcalcification classification
[8]	Gabor filters (GF)	Threshold-based approach	Breast cancer detection
[29]	Grey levels, texture, and features related to independent component analysis	Neural network (NN)	Classifying ROIs into normal/abnormal Classifying ROIs into benign/malignant
[30]	A set of texture features	SVM	Mass detection
[31]	Ripley's K function texture measures	SVM	Detection of breast masses
[9]	Texture features derived from concurrence matrix	NN	Microcalcification classification
[32]	A set of texture features	k-NN, SVM, random forests, logistic model trees, and Naive Bayes	Lesion classification
[10]	HAR	Bayesian classifier Fisher linear discriminant	Study the effect of pixel resolution on the performance of texture methods
[3]	LBP, robust LBP, centre symmetric LBP, fuzzy LBP, local grey level appearance, LDN, HOG, HAR, and GF	k-NN, linear SVM, nonlinear SVM random forest, and Fisher linear discriminant analysis (FLDA)	Finding the best combination among the texture methods to classify ROIs into mass/normal
[33]	Local ternary pattern and local phase quantization	SVM	Classifying tumors into benign/malignant
[34]	Novel sets of texture descriptors extracted from the cooccurrence matrix	SVM	Six medical datasets were used for validation, one of them for breast cancer
[35]	Texture analysis techniques based on the cooccurrence matrix and region-based approaches	SVM	15 datasets were used for validation, one of them for breast cancer
[36]	HOG, dense scale invariant feature transform, and local configuration pattern	SVM, k-NN, FLDA, and decision tree	Classifying ROIs into normal/abnormal Classifying ROIs into benign/malignant
[37]	Curvelet moments	k-NN	Classifying ROIs into normal/abnormal Classifying ROIs into benign/malignant

CAD systems usually focus on a ROI to study breast masses. The texture of this ROI describes the pattern of spatial variation of gray levels in a neighbourhood that is small compared to the breast area but big enough to include the masses. In other words, texture must be analyzed in a region, and the size of this region should be tuned. Thus, we should answer the question: *what is the optimal neighbourhood size (integration scale) for texture analysis*? In addition, the size of a mammogram is usually in the range of thousands of pixels. Consequently, several works have reduced the original resolution of a mammogram to reduce the computational complexity and the execution time of their algorithms [14], or to save resources (e.g., memory and storage space). However, image downsampling may also affect the performance of the texture analysis methods. Therefore, we should answer the question: *how far can we downsample the image while keeping the performance of the texture methods*?

In breast cancer CAD systems, several preprocessing operations such as image filtering or enhancement are usually applied to mammograms. Pisano et al. show that the contrast-limited adaptive histogram equalization (CLAHE) applied to a mammogram before it is displayed can make the indicative

structures of breast cancer more visible [15]. Sharpening (SH) is used to improve the detection of clustered calcifications [16]. The median filter (MF) is used to remove the noise from the mammograms [17]. *Preprocessing may affect the performance of texture analysis methods* because it effectively changes the gray levels of the images. This effect should be assessed. After extracting the texture features from a given mammogram, they are usually normalized before proceeding to the classification stage. *The utilized normalization method may also affect the final classification results*.

In this paper, we study the effect of pixel resolution, integration scale, preprocessing, and feature normalization on the performance of texture analysis methods when used to classify masses in mammograms. For that purpose, we have chosen five widely/recently used texture methods: local binary pattern (LBP), local directional number (LDN), histogram of oriented gradients (HOG), Haralick's features (HAR), and Gabor filters (GF). In order to evaluate the performance of the aforementioned methods, we extracted a set of regions of interest (ROIs) containing lesions from the mini-MIAS database [18], and we used each texture analysis method to classify the ROIs into benign or

malignant. The performance of each texture method is evaluated with five pixel resolutions (200 μm, 400 μm, 600 μm, 800 μm, and 1000 μm), six integration scales (25×25, 32×32, 50×50, 64×64, 75×75, and 100×100 pixels), three preprocessing steps (CLAHE, MF, and SH), and five feature normalization methods. In addition, linear and nonlinear SVM classifiers are used.

To the best of our knowledge, only one previous study has conducted a similar evaluation. Rangayyan et al. studied the effect of pixel resolution on texture features of breast masses in mammograms [10]. However, only pixel resolution and Haralick's features were considered. In contrast, the current study takes into account a wider range of factors such as pixel resolution, integration scale, preprocessing, and feature normalization, and it considers a larger number and more powerful texture descriptors that have been successfully applied in recent relevant work. Moreover, we include linear and nonlinear SVMs; thus, both relatively simple and complex classification approaches can be assessed. Lastly, we analyze the combination of the best options for those factors using three approaches: greedy, sequential forward selection (SFS), and exhaustive search (ExS).

The rest of this paper is organized as follows. Section 2 describes the database and the methods used in this study. Section 3 shows our experimental results, which are then discussed in Section 4. Finally, Section 5 concludes our study.

2. Materials and Methods

In this study, we assess the performance of five texture analysis methods (LBP, LDN, HOG, HAR, and GF) while varying the pixel resolution, integration scale, image preprocessing algorithm, and data normalization method. To that end, we extracted a set of ROIs containing either benign or malignant masses from the mini-MIAS database. Given a certain texture analysis method, a feature vector is extracted from each ROI to be fed into a linear support vector machine (LSVM) or a nonlinear support vector machine (NLSVM). The trained models are used to determine if an unseen ROI contains a benign or a malignant mass.

2.1. Materials. The mini-MIAS database, consisting of 322 mediolateral oblique images of 161 cases, is used in our experiments. It was created from the original MIAS database by downsampling the images from 50 μm to 200 μm per pixel and clipping/padding to a fixed size of 1024×1024 pixels. A ground truth was prepared by experienced radiologists and confirmed using a biopsy procedure. The dataset is available at http://peipa.essex.ac.uk/info/mias.html. In this study 109 ROIs, 60 containing a benign mass and 49 containing a malignant mass, were used. Figure 1 shows examples of the extracted ROIs. Interested researchers can request the ROIs from the corresponding author of the paper.

The authors of the mini-MIAS database reported that they reduced the pixel resolution of the original MIAS database (digitized at 50 μm) to 200 μm by popular request. Moreover, several studies have used the pixel resolution 200 μm as a

baseline resolution in their applications [14, 19]. We do the same in this work.

2.2. Texture Analysis Methods. This section explains the utilized texture analysis methods including the parameters selected for each of them.

2.2.1. Local Binary Pattern. The LBP labels the pixels of an image by comparing a 3×3-pixel neighbourhood with the value of the central pixel [20]. Pixels in this neighbourhood with a value greater than the central pixel are labelled as 1 and the rest as 0; thus, each pixel is represented by 8 bits. The size of the neighbourhood may vary on different applications (e.g., 3×3 and 5×5). A *uniform LBP* is an extension of the original LBP in which only patterns that contain at most two transitions from 0 to 1 (or vice versa) are considered. In uniform LBP mapping, there is a separate output label for each uniform pattern and all the nonuniform patterns are assigned to a single label. In this study, a 3×3 neighbourhood is used to generate the histogram of uniform LBPs for each ROI. The uniform mapping produces 59 output labels (59 dimensions) for neighbourhoods of 8 pixels. The implementation of LBP descriptor is available at http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab.

2.2.2. Local Directional Number. In the LDN [21], the edge responses are computed in eight different directions by convoluting the Kirsch compass masks [22] with the ROIs. The locations of the top positive and negative edge responses are used to generate a 6-bit code for each pixel. Finally, the histogram of the LDN codes is calculated in the given ROI (64 dimensions). The implementation of LDN descriptor is available at https://gitlab.com/my-research/local-directional-number-pattern.git.

2.2.3. Histogram of Oriented Gradients. In the HOG method [23], the occurrences of edge orientations in a ROI are counted. The image is divided into blocks (small groups of cells) and then a weighted histogram is computed for each of them. The combination of the histograms of all blocks represents the final HOG descriptors. In order to get the best performance of HOG, its parameters have been empirically tuned. In this study, we used a 3×3 cell size, 8×8 cells for the block size, and a 9-bit histogram. The implementation of HOG descriptor is available at http://www.vlfeat.org/overview/hog.html.

2.2.4. Haralick's Features. The HAR features are computed from the gray level cooccurrence matrix (GLCM). In the GLCM, the distribution of cooccurring gray level values at a given offset (direction and distance) is computed [24]. A GLCM is computed from each ROI, and then 14 texture features are calculated: *angular second moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, information measure of correlation 1, information measure of correlation 2*, and *maximal correlation coefficient* [10]. The mathematical expression of each feature can be
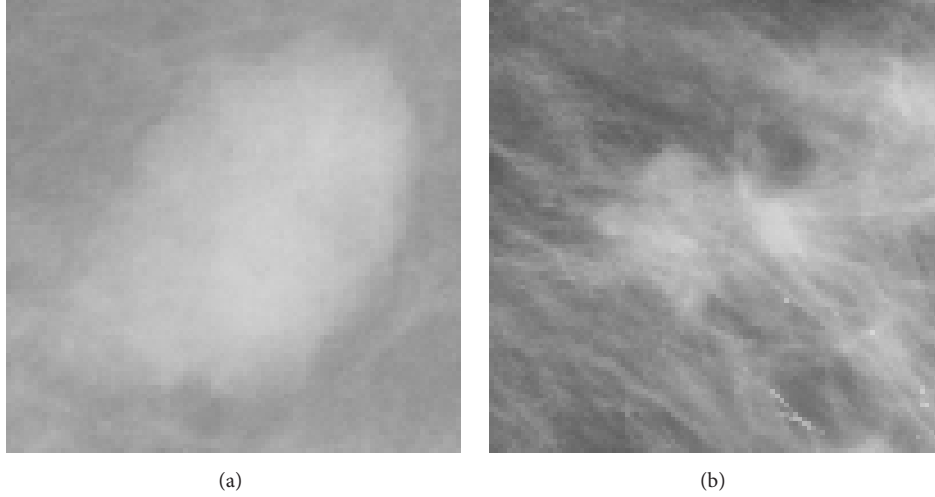
(a)

(b)

FIGURE 1: ROIs extracted from the mini-MIAS breast cancer database. A ROI containing (a) a benign mass and (b) a malignant mass.

found in the relevant previous work [25, 26]. The implementation of HAR descriptors is available at https://github.com/nutsiepully/spiff/blob/master/src/haralick.m.

2.2.5. Gabor Filters.
A two-dimensional Gabor filter $g(x, y)$ can be expressed as a sinusoid with a particular frequency and orientation, modulated by a Gaussian envelope

$$g(x, y) = \exp^{(-1/2)(x^2/\sigma_x^2 + y^2/\sigma_y^2)} \exp^{-j2\pi(u_0 x + v_0 y)}, \quad (1)$$

where (u_0, v_0) is the centre of a sinusoidal function and σ_x and σ_y are the standard deviations along two orthogonal directions (which determine the width of the Gaussian envelope along the x- and y-axes in the spatial domain). Given a ROI $I(x, y)$, the filtered ROI $f(x, y)$ is the result of convoluting $I(x, y)$ and $g(x, y)$. Tuning GF to specific frequencies and directions can lead them to detect both local orientation and frequency information from an image [27]. In this study, we used 4 scales and 6 orientations to obtain these filtered ROIs. This design produces 24 responses. For each ROI, the energies of the 24 responses are calculated, and then they are aggregated in order to form the feature vector. The implementation of Gabor filters is available at https://github.com/mhaghighat/gabor.

2.3. Preprocessing.
The performance of the texture analysis methods is evaluated with three preprocessing algorithms: CLAHE, median filter (MF), and sharpening (SH).

(i) *CLAHE*: it works on small regions of the input ROI (known as tiles). The contrast of each tile is enhanced; consequently the histogram of the output region approximately matches a predefined distribution [28]. In this study, the *Rayleigh* distribution is used [15].

(ii) *MF*: each pixel in the filtered ROI contains the median value of the $m \times n$ neighbourhood around the corresponding pixel in the input ROI [17]. In this study, a 3×3 neighbourhood is used.

(iii) *SH*: in order to sharpen a ROI, it is first blurred; edges are detected in the blurred ROI and added to it to produce a sharper image [16].

The preprocessing operations can be carried out using the following MATLAB functions: CLAHE (adapthisteq.m), median filter (medfilt2.m), and sharpening (imsharpen.m). Figure 2 shows examples for MF, SH, and CLAHE when they are applied to benign and malignant masses.

2.4. Feature Normalization Methods.
Feature vectors are normalized in order to prevent attributes with higher numeric ranges from dominating those with lower numeric ranges. Given a feature vector $x = [x_1, x_2, x_3, \ldots, x_N]$, the normalized feature vector x_{new} is calculated using five normalization methods as follows [38, 39]:

(i) The zero mean unit variance (zs) method: $x_{new} = (x - \mu)/\sigma$, where μ and σ are the mean and the variance of x.

(ii) The maximum-minimum (mn) method: $x_{new} = (x - x_{min})/(x_{max} - x_{min})$, where x_{max} and x_{min} are the maximum and minimum of x.

(iii) The ℓ^1 method scales x to unit length using the ℓ^1-norm, $x_{new} = x/\sum_{n=1}^{N} |x_n|$.

(iv) The ℓ^2 method scales x to unit length using the ℓ^2-norm, $x_{new} = x/\sqrt{\sum_{n=1}^{N} |x_n|^2}$.

(v) The nh method scales x to unit length as follows: $x_{new} = x/\sum_{n=1}^{N} x$.

The normalization methods can be easily implemented in MATLAB. ℓ^1- and ℓ^2-norm can be carried out using the MATLAB function *norm.m*.

2.5. Classification.
Given a labelled training set of the form (x_i, y_i), $i = 1, 2, \ldots, k$, where $x_i \in \mathbb{R}^n$ are the feature values, $y_i \in \{1, -1\}$ is the class of x_i, n is the number of
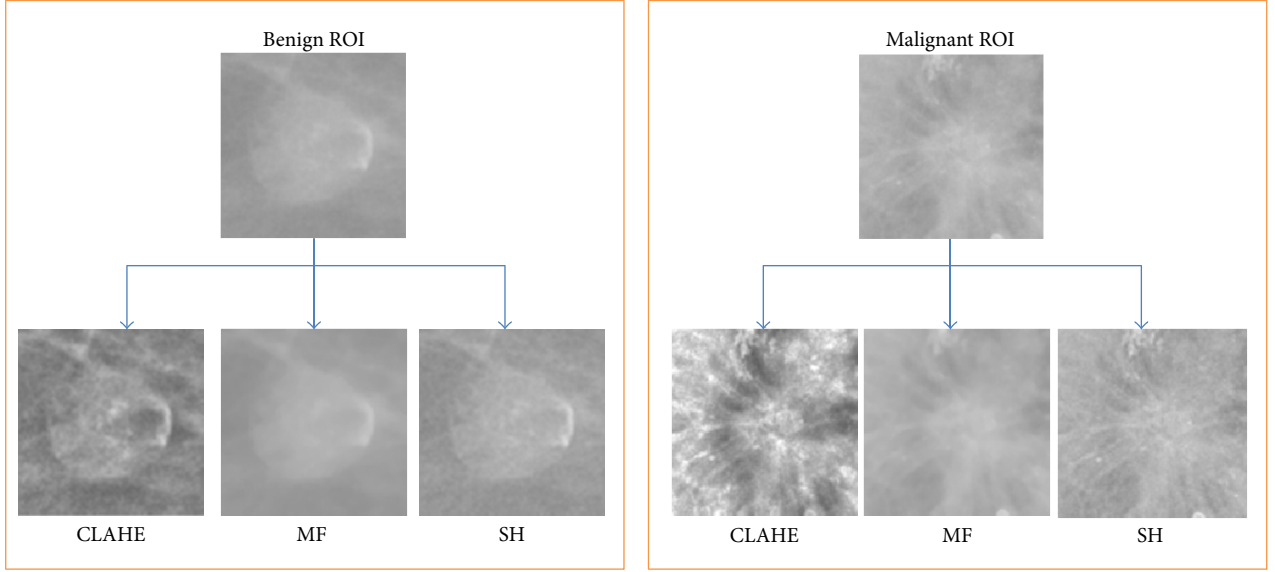
FIGURE 2: Examples of ROI preprocessing.

features, and k is the number of samples, an SVM attempts to discriminate between positive and negative classes by finding a hyperplane that separates them [40]. The SVM classifier solves the following optimization problem:

$$\|\omega\|_{\omega,\xi}^2 + C\sum_{i=1}^{k}\xi_i,$$

$$\text{s.t.} \quad y_i\left(\omega^T\phi\left(x_i\right)+b\right) \geq 1-\xi_i, \quad \xi_i \geq 0, \tag{2}$$

where the soft margin parameter C controls the trade-off between the training error and the complexity of the SVM's model in order to fit the training data and to avoid overfitting. The weight vector ω is normal to the separating hyperplane. The parameter ξ is used to give a degree of flexibility for the algorithm when fitting the data and b represents the bias.

The SVM uses a kernel function to make the data linearly separable. It projects the training data x_i to a higher dimensional space as follows: $K(x_i, x_j) = (\phi^T(x_i) \cdot \phi(x_j))$. The SVM algorithm attempts to find the hyperplane with maximum margin of separation between the classes in the new higher dimensional space. In the case of a LSVM classifier, ϕ refers to a dot product. In the case of a NLSVM, the classifier function is formed by nonlinearly projecting the training data in the input space to a feature space of higher dimension by using a kernel function. In this study, we use a radial basis function (RBF) as a mapping kernel, which is defined as follows:

$$K\left(x_i, x_j\right) = \exp\left(-\gamma\left\|x_i - x_j\right\|_2^2\right), \tag{3}$$

where $\gamma = 1/2\sigma^2$, $\|x_i - x_j\|_2^2$ is the squared Euclidean distance between the two feature vectors x_i and x_j, and σ is a free parameter. In this work, we use LIBSVM [41] to implement SVM classifiers. LIBSVM is available at https://www.csie.ntu.edu.tw/~cjlin/libsvm/. A grid search algorithm is performed to find the optimal parameter of the

RBF kernel, γ, and the regularization parameter, C. For each training set, we estimated the parameters used by SVM in the classification as done in [42].

2.6. Evaluation. The performance of each texture analysis method is measured in terms of the area under the curve (AUC) of the receiver operating characteristics (ROC) curve [43]. The SVM classifier provides decision values related to the membership of each class. To generate a ROC curve, we vary a threshold over the decision values. We also use the k-fold cross validation technique to generate the training and testing data. In this procedure, the data are partitioned into k folds; thus $1/k$ of ROIs are used for testing and the rest of ROIs are used for training. In this study, $k = 10$. The mean AUC value is calculated over the cross validation process.

3. Experiments

In this section, we present the effect of pixel resolution, integration scale, preprocessing steps, and normalization methods on the performance of the texture analysis methods when they are applied to *benign/malignant* mass classification in mammograms. Moreover, we study the effect of different combinations of the aforementioned factors.

3.1. Effect of Pixel Resolution and Integration Scale. As we commented in Section 2.1, the pixel resolution 200 μm has been widely used in several studies [14, 19]. So, in this experiment we start with this pixel resolution and then the mammograms are downsampled to generate different pixel resolutions. The downsampling step includes antialiasing filtering and a bicubic interpolation. Five pixel resolutions are generated (200 μm, 400 μm, 600 μm, 800 μm, and 1000 μm), and then we use six integration scales (25 × 25, 32 × 32, 50 × 50, 64 × 64, 75 × 75, and 100 × 100 pixels) to analyze

TABLE 2: Summary of the ANOVA results of pixel resolution and integration scale with the LSVM (the value in each cell is a p value).

Method	Res	IS	Res∗IS
LBP	**0.0024**	**0.001**	0.908
LDN	0.1174	0.4035	0.8037
HOG	0.3905	0.6515	0.4636
HAR	0.7846	0.0962	0.2895
GF	0.083	0.8259	0.9864

TABLE 3: Summary of the ANOVA results of pixel resolution and integration scale with the NLSVM (the value in each cell is a p value).

Method	Res	IS	Res∗IS
LBP	0.9332	**0.0101**	**0.0095**
LDN	0.2387	0.0772	0.6451
HOG	**0.0448**	**0.0103**	0.5138
HAR	0.4253	**0.004**	0.0847
GF	0.6552	0.3109	0.2024

the texture of each ROI. In this experiment, no preprocessing is applied, and the standard zs normalization method is used to normalize the extracted feature vectors. The effect of pixel resolution and integration scale in the performance of LBP, LDN, HOG, HAR, and GF with the LSVM and the NLSVM is shown in Figure 3.

As shown in Figure 3, each texture method achieves its best AUC value at a certain pixel resolution and integration scale. Among all texture methods, LBP achieves the best AUC value (0.78) at pixel resolution 800 μm, integration scale 75 × 75.

The analysis of variance (ANOVA) test [44] has been used to examine the interaction between pixel resolutions and integration scales. The experimental design of ANOVA includes two factors: pixel resolution (Res) and integration scale (IS). Res includes five levels (200 μm, 400 μm, 600 μm, 800 μm, and 1000 μm), whereas IS includes six levels (25 × 25, 32 × 32, 50 × 50, 64 × 64, 75 × 75, and 100 × 100 pixels). Each combination of the levels of Res and IS produces an AUC value (response). The confidence level is set to 0.05. The results are shown in Tables 2 and 3.

As shown in Table 2, with LBP and the LSVM, the mean responses for the levels of pixel resolution are significantly different ($p = 0.0024$). Similarly, the mean responses for the levels of integration scale are significantly different. In the case of LDN, HOG, HAR, and GF, the mean responses for the levels of pixel resolution and integration scale are not significantly different. The p values indicate that the interactions between the levels of pixel resolution and integration scale (Res ∗ IS) are not significant.

As shown in Table 3, the mean responses for the levels of pixel resolution are significantly different in the case of HOG with the NLSVM. In the case of LBP, LDN, HAR, and GF, the mean responses for the levels of pixel resolution are not significantly different. The mean responses for the levels of integration scale are significantly different in the case of LBP, HOG, and HAR. With LBP and the NLSVM, the interaction

between pixel resolution and integration scale (Res ∗ IS) is significant.

3.2. Effect of Preprocessing. In this experiment, the integration scale that obtained the highest AUC value with each texture analysis method at the baseline pixel resolution of 200 μm and the standard zs normalization method are used. The effect of no preprocessing (NP), CLAHE, MF, and SH on the performance of each texture analysis method is shown in Figure 4. As can be seen, each texture method produces the highest AUC value with a certain preprocessing algorithm. In this experiment, LBP achieves the highest AUC value with SH and the NLSVM, while LDN and HAR achieve the highest AUC value with NP and the LSVM. HOG achieves the highest AUC value with CLAHE and the LSVM. In turn, GF achieves the highest AUC value with CLAHE and the NLSVM.

3.3. Effect of Feature Normalization Methods. In this experiment, we study the effect of five normalization methods (zs, mn, ℓ^1, ℓ^2, and nh) on the performance of each texture analysis method. For each texture analysis method, we use the integration scale that produces the highest AUC value at pixel resolution 200 μm. No preprocessing method is used. The effect of the normalization methods is shown in Figure 5. With the LSVM, zs normalization has led LBP and LDN to AUC values better than other normalization methods, while GF achieves its highest AUC value with ℓ^1 normalization and the NLSVM. As shown in the figure, each texture analysis method achieves its highest AUC value with a certain normalization method.

3.4. Summary of the Results. The best AUC values of each texture analysis method considering the experiments in Sections 3.1, 3.2, and 3.3 are summarized in Table 4. LBP produces the best AUC value (0.78) at pixel resolution 800 μm, integration scale 75 × 75, no preprocessing, zs normalization method, and the LSVM. In turn, HAR produces the lowest AUC value (0.61). LBP, LDN, HOG, and HAR achieve their best values with the LSVM, whereas GF achieves its best AUC value with the NLSVM.

3.5. Combining the Levels of All Factors. To find the best combination among the levels of all factors, we use three approaches: greedy, sequential forward selection (SFS), and exhaustive search (ExS). In the greedy approach, we try to combine the best options of the aforementioned factors. For each texture analysis method, we summarize the best levels of pixel resolution, integration scale, and normalization methods in Table 5.

Table 6 shows that combining the best levels of pixel resolution, integration scale, preprocessing, and feature normalization does not yield improvement on the AUC values of the texture analysis methods reported in Table 4. In fact, LBP, HOG, and GF produced substantially lower AUC values. The LSVM yields higher AUC values than the NLSVM.

Secondly, we use a SFS approach to find the best combination. It consists of two sequential steps: finding the normalization method that improves the current performance
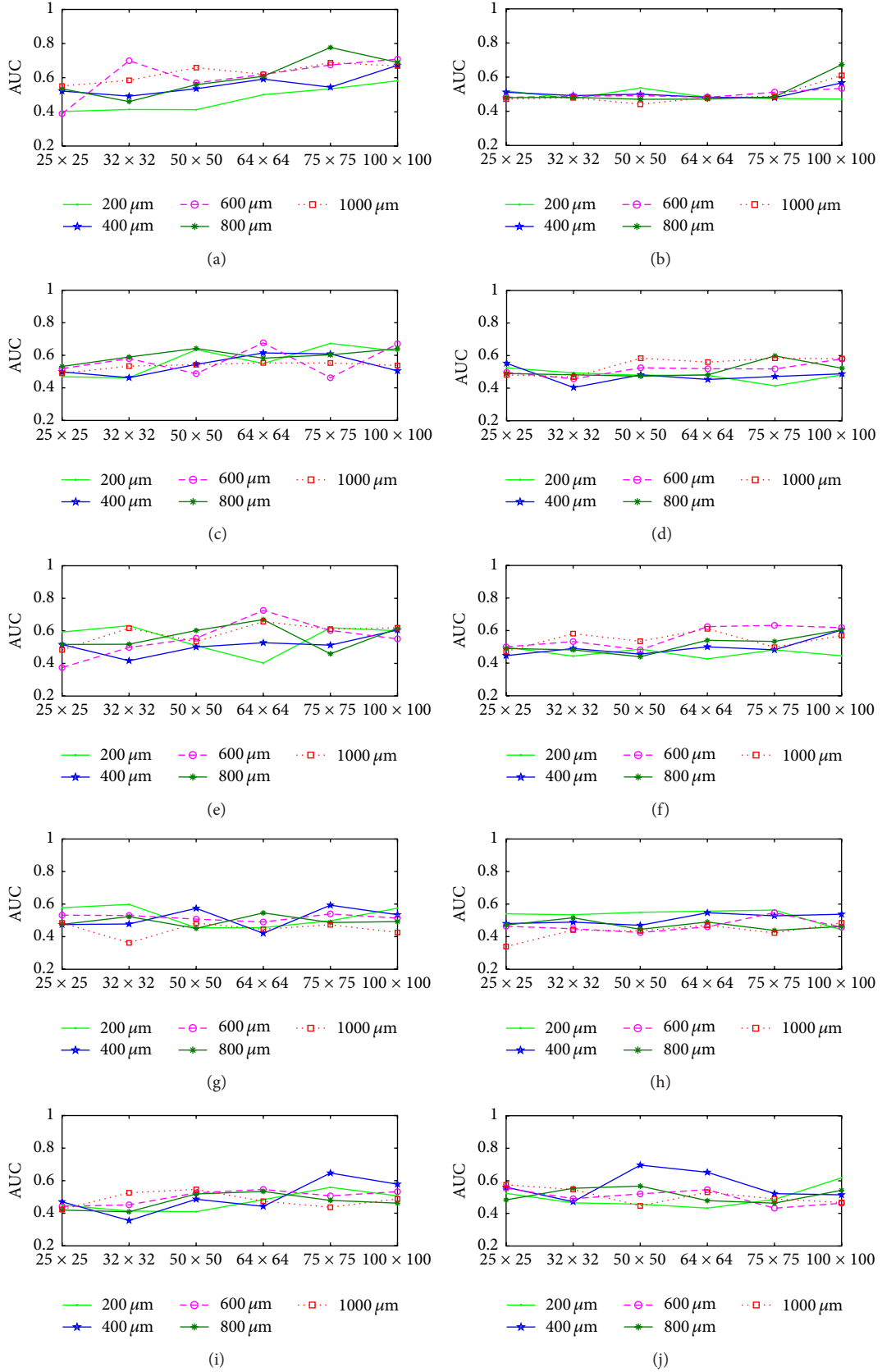
FIGURE 3: The effect of pixel resolution and integration scale on the performance of the texture methods with the LSVM (a, c, e, g, i), the NLSVM (b, d, f, h, j), (a)-(b) LBP, (c)-(d) LDN, (e)-(f) HOG, (g)-(h) HAR, and (i)-(j) GF.
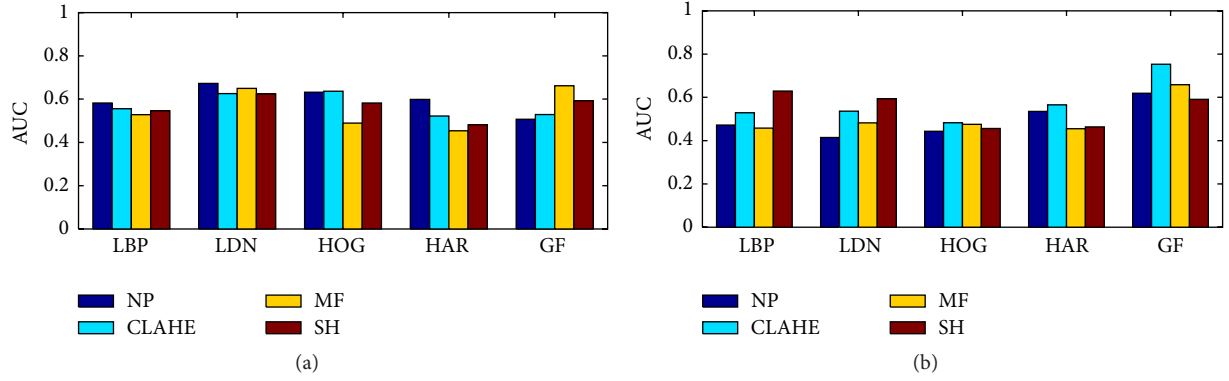
FIGURE 4: The performance of the texture analysis methods with NP, CLAHE, MF, and SH using (a) the LSVM and (b) the NLSVM.
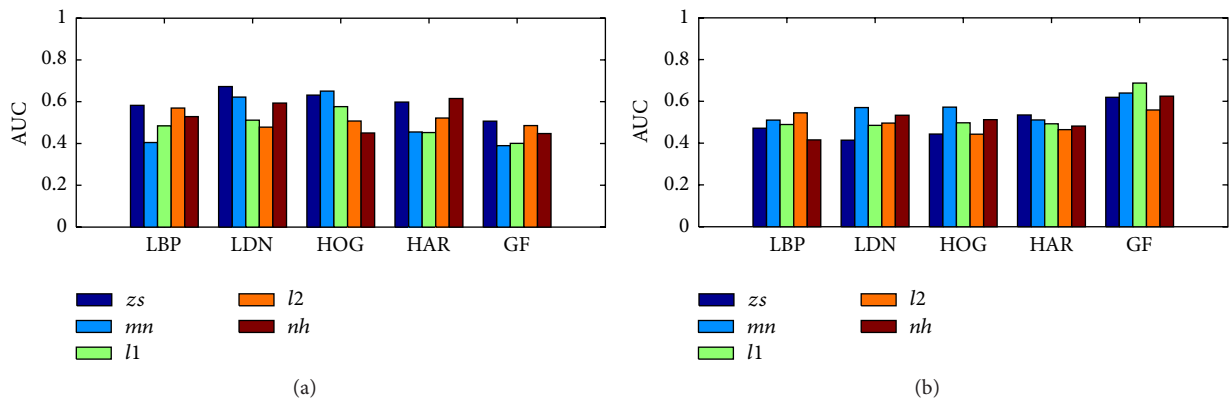

FIGURE 5: The performance of the texture analysis methods with different feature normalization methods using (a) the LSVM and (b) the NLSVM.

the most and then finding the preprocessing method that keeps improving this performance. For each texture method, in the first step, we start with the best pixel resolution and integration scale summarized in Table 5. Then, with no preprocessing, the extracted features are separately normalized by each normalization method. Then, the one that improves the performance in combination with the previous two factors is added. In the second step, we apply each preprocessing option to the ROIs (NP, CLAHE, MF, and SH). Then we extract the texture features and normalize them using the best normalization method obtained in the previous step. Both LSVM and NLSVM are used to classify the ROIs. Table 7 shows that the SFS does not improve the AUC value of GF achieved in Table 4. LBP, LDN, HOG, and HAR achieve AUC values close to the ones listed in Table 4. With all texture methods, the SFS approach achieves AUC values better than the greedy approach.

Lastly, we use an ExS algorithm, which is looking for the best combination among five pixel resolutions, six integration scales, and four preprocessing (NP, CLAHE, MF, and SH) and five data normalization methods, resulting in 600 combinations. In the previous experiments, we found that the LSVM usually achieves the best results except with GF. The NLSVM has two parameters that need to be optimized to achieve the best classification results. Adding NLSVM's

parameters optimization to the ExS substantially increases its complexity. So we decided to only use the LSVM in this final test.

As shown in Table 8, the ExS approach improves the AUC values of LDN, HOG, and HAR. The GF achieves an AUC value lower than the one listed in Table 4 because the LSVM can not perfectly separate the GF features.

4. Discussion

Many factors affect the performance of texture analysis methods when applied to benign/malignant mass classification. In this work, we study the effect of factors such as pixel resolution, integration scale, preprocessing, and feature normalization. We use the well-known mini-MIAS database in this study. We start with the original pixel resolution of the mini-MIAS database (200 μm); then we downsample the mammograms in order to generate the pixel resolutions 400 μm, 600 μm, 800 μm, and 1000 μm. In addition, six integration scales are used (25 × 25, 32 × 32, 50 × 50, 64 × 64, 75 × 75, and 100 × 100 pixels). These integration scales cover most of the sizes of the masses in the mini-MIAS database, which range from a few pixels to tens of pixels (the mean diameter of the circle containing the masses is about 49 pixels). Several previous studies have

TABLE 4: Best AUC value for each texture analysis method and the configuration that yields it considering the experiments in Sections 3.1, 3.2, and 3.3.

Method	Best value	Res (μm)	IS	Classifier	Preprocessing	Nor.
LBP	0.78	800	75×75	LSVM	NP	*zs*
LDN	0.68	600	64×64	LSVM	NP	*zs*
HOG	0.72	600	64×64	LSVM	NP	*zs*
HAR	0.61	200	32×32	LSVM	NP	*nh*
GF	0.75	200	100×100	NLSVM	CLAHE	*zs*

TABLE 5: The best option of pixel resolution, integration scale, preprocessing, and normalization methods with each texture method.

Method	Res (μm)	IS (pixels)	Preprocessing	Normalization
LBP	800	75×75	SH	*zs*
LDN	600	64×64	NP	*zs*
HOG	600	64×64	CLAHE	*mn*
HAR	200	32×32	NP	*nh*
GF	400	50×50	CLAHE	ℓ^1

TABLE 6: Results of the greedy approach (AUC).

Method	LSVM	NLSVM
LBP	0.46	0.40
LDN	0.68	0.52
HOG	0.44	0.44
HAR	0.61	0.48
GF	0.58	0.54

TABLE 7: Results of the SFS approach.

Method	Best AUC	Best parameters
LBP	0.780	*zs*, NP, and LSVM
LDN	0.679	*zs*, NP, and LSVM
HOG	0.716	*zs*, NP, and LSVM
HAR	0.605	*nh*, NP, and LSVM
GF	0.720	*zs*, CLAHE, and NLSVM

TABLE 8: Results of the ExS approach.

Method	Best AUC	Best parameters
LBP	0.78	800, 75×75, NP, and *zs*
LDN	0.70	600, 75×75, MF, and *zs*
HOG	0.737	1000, 50×50, SH, and *mn*
HAR	0.666	800, 32×32, CLAHE, and *nh*
GF	0.691	600, 32×32, NP, and ℓ^1

used one of these integration scales to analyze the texture of mammograms [3, 6, 7]. Thus, we hypothesize that the aforementioned integration scales are able to deal with all the masses appearing in the mini-MIAS database.

The shape of breast masses is one of the powerful features that can be used to discriminate between benign and malignant masses. The boundaries of malignant masses usually have irregular shapes, while the boundaries of benign masses have regular ones. In the case of breast mass analysis, pixel resolution may be a critical factor because image downsampling may remove some fine detail from the image. However, as our results indicate, it would be possible to decrease the resolution far beyond 200 μm and obtain good classification results. A notable example is LBP, which actually achieved its best performance at 800 μm. A possible explanation is that core information such as that contained in the boundary of masses may still be preserved even after downsampling and become more useful for methods such as LBP that operate over higher order statistics of gray intensity values. Obviously, when the resolution is far too low, the classification performance degrades, as the shape of the boundaries of benign and malignant masses will be very similar. Another important factor is the integration scale, as it should be big enough to cover the masses and their boundaries and small enough to exclude other tissues. The effect of pixel resolution and integration scale on the performance of texture methods should be jointly studied.

As summarized in Table 5, each texture method achieves its highest AUC value at a certain pixel resolution and integration scale. A pixel resolution of 200 μm and an integration scale of 32×32 pixels have led HAR to its highest AUC value. In turn, a pixel resolution of 800 μm and an integration scale of 75×75 pixels have led LBP to its best AUC value. The integration scale and the pixel resolution interact with each other in a certain way. In the case of LBP, LDN, and HOG, the texture features of each method are represented in a histogram. This histogram includes the repetition of the patterns detected by each method at a certain pixel resolution and integration scale. LBP features calculated at pixel resolution 200 μm are different from those calculated at pixel resolution 400 μm. LDN and HOG also produce different patterns at different pixel resolutions. The local patterns of LBP, LDN, and HOG are usually calculated within a certain integration scale. Different integration scales will yield different histograms for the local patterns. For instance, the histograms of LBP that are calculated with the integration scales 75×75 and 100×100 are different.

ANOVA results show that the mean AUC values of the pixel resolutions are significantly different in the case of LBP with the LSVM. In addition, the mean AUC values of the integration scales are significantly different with LBP, HOG, and HAR and the NLSVM. The performance differences with respect to the pixel resolutions and the integration scales are only significantly different with the LBP and the NLSVM ($p = 0.0095$). These results indicate that the choice of the pixel resolution and the integration scale has a direct implication on the performance of a texture-based CAD system, because our choice substantially affects the performance of the utilized texture method.

Image preprocessing also affects the performance of the texture analysis methods. HOG and GF achieve the highest AUC values with CLAHE, while LDN and HAR perform better with NP. Indeed, CLAHE, MF, and SH change the intensities of the mammograms in different ways. As a result, each texture analysis method will produce a different AUC value with each preprocessing technique. In general, the preprocessing approach that makes the small-scale structures in the ROIs more visible would give the texture methods more discriminative power. For instance, CLAHE leads GF to its best AUC value (0.75). There is also a coherent relation between the principle of operation of some texture methods and the utilized preprocessing. For instance, the binary patterns of the LDN are calculated based on the edge responses of each pixel in the image. MF removes the outliers before calculating the edge responses. Thus, the edge responses will be properly calculated, and the discriminative power of LDN will improve.

Prior to mass classification, the calculated texture features should be normalized to prevent attributes with higher numeric ranges from dominating those with lower numeric ranges. As shown in our experiments, each texture method produces its highest AUC value with a certain normalization method. This is because each normalization method produces numerical values with different distributions. Consequently, the arrangement of the texture features in the feature space with a certain normalization method is different than with other normalization methods. Thus, the normalization technique changes the final values of the features computed by each texture method. As shown in Table 5, LBP and LDN achieve the highest AUC values with zs normalization, HOG with mn, HAR with nh, and GF with ℓ^1.

In the classification stage, we utilize two widely used classifiers in the field of mammogram analysis: the LSVM and the NLSVM. The first one tries to linearly separate the texture features in the feature space, while the second one uses a kernel function (RBF) to separate the features. As shown in Table 4, the LSVM has led LBP, LDN, HOG, and HAR to the highest AUC values. Conversely, GF achieves the best AUC value with the NLSVM, indicating that GF features are not linearly separable.

Table 4 shows a summary of the levels of pixel resolution, integration scale, preprocessing, and normalization methods that have led each texture method to its best AUC value considering the experiments in Sections 3.1, 3.2, and 3.3. HAR and GF achieve the best AUC values at pixel resolution 200 μm, while LDN and HOG give their best results at pixel resolution 600 μm. No method achieves its best AUC value with the integration scales 25×25 and 50×50 pixels.

The greedy, SFS, and ExS approaches are used to find the best combination among the levels of all factors. Although the greedy approach is the least complex approach, it yielded poor AUC values. In contrast, the ExS achieved good results, but its computational complexity is the highest. The SFS approach provides a trade-off between the accuracy and the computational complexity. It is not as complex as the ExS approach and it does not produce poor AUC values as the greedy approach. In the case of LBP, LDN, HOG, and HAR, Table 7 shows that the SFS approach produces approximately the same results as those obtained with the ExS approach. The GF achieved better AUC values with the SFS approach because it used the NLSVM, whereas using it with the ExS approach presents some additional challenges in the calculation of the optimal values of its internal parameters (γ and C).

Rangayyan et al. extracted 111 ROIs from mammograms, which were obtained from three different sources: mammographic image analysis society (MIAS), the teaching library of the Foothills Hospital in Calgary, and a screening test (the Alberta program for the early detection of breast cancer) [10]. Although using mammograms from different sources may be helpful to assess the robustness of the studied texture methods, the three mammogram sets used by Rangayyan et al. were digitized at different pixel resolutions. Thus, the characteristics of the textures extracted from the 111 ROIs may be different. This changes the characteristics of the extracted features, so the effect of pixel resolution on the performance of the texture methods may have not been properly studied. In contrast, in the current study, the ROIs were extracted from a single source (the mini-MIAS database). Rangayyan et al. extracted ROIs with different sizes (each ROI included a mass) and they did not mention the effect of the integration scale on the performance of the texture methods. Conversely, the current study has considered six integration scales. With pixel resolution 800 μm, integration scale 75×75, no preprocessing, zs normalization method, and the LSVM, the LBP achieves the best AUC value (0.78) compared to other texture methods, exceeding the best AUC value (0.75) achieved by Rangayyan et al. [10]. This is encouraging, so our future work will focus on improving the capabilities of an LBP-based approach by complementing it with the analysis of the fractal dimensions in multiple integration scales at different pixel resolutions.

As mentioned above, the work of [10] has some similarities to our analysis; however it obtained an AUC value less than the one of our study; in addition, the authors of [45] have studied the effect of ROI size and location on texture methods when classifying the low-risk women and the BRCA1/BRCA2 gene-mutation carriers. In turn, our study focuses on analyzing the impact of pixel resolution, integration scale, preprocessing, and feature normalization on texture methods when classifying breast tumors into benign or malignant.

In the current work we studied the impact of the abovementioned factors on the performance of texture

methods, achieving the best AUC value with the LBP (0.78). However, some methods in the literature achieved better benign/malignant breast cancer classification results, such as the ones of [33–35]. For instance, the authors of [35] achieved an AUC of 0.92 because they used ROIs of different dataset (DDSM) and extracted the GLCM features from subwindows or regions (they added spatial information). We expect that the classification results of our study will be improved when utilizing the region-based approach of [35] with each texture method. One of our future research lines is to integrate the region-based approach of [35] with our analysis.

5. Conclusion

Texture analysis methods, when applied to *benign/malignant* mass classification in mammograms, are sensitive to the changes of pixel resolution, integration scale, preprocessing, and feature normalization. The best combination of the aforementioned factors should be identified to achieve the best discriminative power of each texture analysis method. We expect that the assessment performed in this study will help researchers to accomplish this task. Due to its computational cost advantage, sequential forward selection would be a suitable approach to determine a reasonable (possibly the best) factor configuration.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] M. Malvezzi, P. Bertuccio, F. Levi, C. La Vecchia, and E. Negri, "European cancer mortality predictions for the year 2014," *Annals of Oncology*, vol. 25, no. 8, pp. 1650–1656, 2014.

[2] F. J. Gilbert, S. M. Astley, M. G. C. Gillan et al., "Single reading with computer-aided detection for screening mammography," *The New England Journal of Medicine*, vol. 359, no. 16, pp. 1675–1684, 2008.

[3] M. Abdel-Nasser, A. Moreno, and D. Puig, "Towards cost reduction of breast cancer diagnosis using mammography texture analysis," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 28, no. 1-2, pp. 385–402, 2016.

[4] R. Bellotti, F. De Carlo, S. Tangaro et al., "A completely automated CAD system for mass detection in a large mammographic database," *Medical Physics*, vol. 33, no. 8, pp. 3066–3075, 2006.

[5] J. Melendez, C. I. Sánchez, B. Van Ginneken, and N. Karssemeijer, "Improving mass candidate detection in mammograms via feature maxima propagation and local feature selection," *Medical Physics*, vol. 41, no. 8, Article ID 081904, 2014.

[6] A. Oliver, X. Lladó, J. Freixenet, and J. Martí, "False positive reduction in mammographic mass detection using local binary patterns," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2007*, pp. 286–293, Springer, Berlin, Germany, 2007.

[7] V. Pomponiu, H. Hariharan, B. Zheng, and D. Gur, "Improving breast mass detection using histogram of oriented gradients," in *Medical Imaging: Computer-Aided Diagnosis*, vol. 9035 of *Proceedings of SPIE*, pp. 1–6, International Society for Optics and Photonics, San Diego, Calif, USA, March 2014.

[8] Y. Zheng, "Breast cancer detection with Gabor features from digital Mammograms," *Algorithms*, vol. 3, no. 1, pp. 44–62, 2010.

[9] H.-P. Chan, B. Sahiner, N. Patrick et al., "Computerized classification of malignant and benign microcalcifications on mammograms: texture analysis using an artificial neural network," *Physics in Medicine and Biology*, vol. 42, no. 3, pp. 549–567, 1997.

[10] R. M. Rangayyan, T. M. Nguyen, F. J. Ayres, and A. K. Nandi, "Effect of pixel resolution on texture features of breast masses in mammograms," *Journal of Digital Imaging*, vol. 23, no. 5, pp. 547–553, 2010.

[11] H. Soltanian-Zadeh, F. Rafiee-Rad, and D. S. Pourabdollah-Nejad, "Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms," *Pattern Recognition*, vol. 37, no. 10, pp. 1973–1986, 2004.

[12] A. Oliver, J. Freixenet, R. Martí et al., "A novel breast tissue density classification methodology," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 55–65, 2008.

[13] A. Oliver, J. Freixenet, J. Martí et al., "A review of automatic mass detection and segmentation in mammographic images," *Medical Image Analysis*, vol. 14, no. 2, pp. 87–110, 2010.

[14] M. P. Sampat, A. C. Bovik, G. J. Whitman, and M. K. Markey, "A model-based framework for the detection of spiculated masses on mammography," *Medical Physics*, vol. 35, no. 5, pp. 2110–2123, 2008.

[15] E. D. Pisano, S. Zong, B. M. Hemminger et al., "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital Imaging*, vol. 11, no. 4, pp. 193–200, 1998.

[16] S. Anand, R. S. S. Kumari, S. Jeeva, and T. Thivya, "Directionlet transform based sharpening and enhancement of mammographic X-ray images," *Biomedical Signal Processing and Control*, vol. 8, no. 4, pp. 391–399, 2013.

[17] T. S. Subashini, V. Ramalingam, and S. Palanivel, "Automated assessment of breast tissue density in digital mammograms," *Computer Vision and Image Understanding*, vol. 114, no. 1, pp. 33–43, 2010.

[18] J. Suckling, J. Parker, D. Dance et al., "The mammographic image analysis society digital mammogram database," in *Proceedings of the 2nd International Workshop on Digital Mammography*, pp. 375–378, York, UK, July 1994.

[19] N. Karssemeijer, "Automated classification of parenchymal patterns in mammograms," *Physics in Medicine and Biology*, vol. 43, no. 2, pp. 365–378, 1998.

[20] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[21] A. R. Rivera, J. R. Castillo, and O. Chae, "Local directional number pattern for face analysis: face and expression recognition," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1740–1752, 2013.

[22] R. A. Kirsch, "Computer determination of the constituent structure of biological images," *Computers and Biomedical Research*, vol. 4, no. 3, pp. 315–328, 1971.

[23] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 886–893, IEEE, San Diego, Calif, USA, June 2005.

[24] R. M. Haralick, I. Dinstein, and K. Shanmugam, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[25] W. Gómez, W. C. A. Pereira, and A. F. C. Infantosi, "Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound," *IEEE Transactions on Medical Imaging*, vol. 31, no. 10, pp. 1889–1899, 2012.

[26] R. P. Ramos, M. Z. do Nascimento, and D. C. Pereira, "Texture extraction: an evaluation of ridgelet, wavelet and co-occurrence based methods applied to mammograms," *Expert Systems with Applications*, vol. 39, no. 12, pp. 11036–11047, 2012.

[27] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *Journal of Neurophysiology*, vol. 58, no. 6, pp. 1233–1258, 1987.

[28] D. T. Puff, E. D. Pisano, K. E. Muller et al., "A method for determination of optimal image enhancement for the detection of mammographic abnormalities," *Journal of Digital Imaging*, vol. 7, no. 4, pp. 161–171, 1994.

[29] I. Christoyianni, A. Koutras, E. Dermatas, and G. Kokkinakis, "Computer aided diagnosis of breast cancer in digitized mammograms," *Computerized Medical Imaging and Graphics*, vol. 26, no. 5, pp. 309–319, 2002.

[30] P. Agrawal, M. Vatsa, and R. Singh, "Saliency based mass detection from screening mammograms," *Signal Processing*, vol. 99, pp. 29–47, 2014.

[31] L. de Oliveira Martins, A. C. Silva, A. C. De Paiva, and M. Gattass, "Detection of breast masses in mammogram images using growing neural gas algorithm and Ripley's K function," *Journal of Signal Processing Systems*, vol. 55, no. 1–3, pp. 77–90, 2009.

[32] D. C. Moura and M. A. G. López, "An evaluation of image descriptors combined with clinical data for breast cancer diagnosis," *International Journal of Computer Assisted Radiology and Surgery*, vol. 8, no. 4, pp. 561–574, 2013.

[33] L. Nanni, S. Brahnam, and A. Lumini, "A very high performing system to discriminate tissues in mammograms as benign and malignant," *Expert Systems with Applications*, vol. 39, no. 2, pp. 1968–1971, 2012.

[34] L. Nanni, S. Brahnam, S. Ghidoni, E. Menegatti, and T. Barrier, "Different approaches for extracting information from the co-occurrence matrix," *PLoS ONE*, vol. 8, no. 12, Article ID e83554, 2013.

[35] L. Nanni, S. Brahnam, S. Ghidoni, and E. Menegatti, "Region-based approaches and descriptors extracted from the co-occurrence matrix," *International Journal of Latest Research in Science and Technology*, vol. 3, pp. 192–200, 2014.

[36] S. Ergin and O. Kilic, "A new feature extraction framework based on wavelets for breast cancer diagnosis," *Computers in Biology and Medicine*, vol. 51, pp. 171–182, 2014.

[37] S. Dhahbi, W. Barhoumi, and E. Zagrouba, "Breast cancer diagnosis in digitized mammograms using curvelet moments," *Computers in Biology and Medicine*, vol. 64, pp. 79–90, 2015.

[38] S. Aksoy and R. M. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 563–582, 2001.

[39] P. Juszczak, D. Tax, and R. Duin, "Feature scaling in support vector data description," in *Proceedings of the 8th Annual Conference of the Advanced School for Computing and Imaging (ASCI '02)*, pp. 95–102, 2002.

[40] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[41] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, article 27, 2011.

[42] G. B. Junior, A. C. Cardoso de Paiva, A. Corrêa Silva, and A. C. M. de Oliveira, "Classification of breast tissues using Moran's index and Geary's coefficient as texture signatures and SVM," *Computers in Biology and Medicine*, vol. 39, no. 12, pp. 1063–1072, 2009.

[43] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[44] P. Armitage, G. Berry, and J. N. Matthews, *Statistical Methods in Medical Research*, John Wiley & Sons, New York, NY, USA, 2002.

[45] H. Li, M. L. Giger, Z. Huo et al., "Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location," *Medical Physics*, vol. 31, no. 3, pp. 549–555, 2004.