Computer Methods and Programs in Biomedicine 160 (2018) 75–83

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

False-positive reduction in computer-aided mass detection using mammographic texture analysis and classification

Sami Dhahbi[a,b,*], Walid Barhoumi[a,c], Jaroslaw Kurek[d], Bartosz Swiderski[d], Michal Kruk[d], Ezzeddine Zagrouba[a]

[a] Université de Tunis El Manar, Institut Supérieur d'Informatique, Research Team on Intelligent Systems in imaging and Artificial Vision (SIIVA), Laboratoire de recherche en Informatique, Modélisation et Traitement de l'Information et de la Connaissance (LIMTIC), 2Rue Abou Raihane Bayrouni, Ariana 2080, Tunisia

[b] Université de Monastir, Faculté de Sciences de Monastir, Avenue Environnement Monastir 5019, Tunisia

[c] Université de Carthage, Ecole Nationale d'Ingénieurs de Carthage, 45 Rue des Entrepreneurs, 2035 Charguia II, Tunis-Carthage, Tunisia

[d] The Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences, 166 Nowoursynowska Street, Warsaw 02-787, Poland

ARTICLE INFO

Article history:
Received 17 February 2017
Revised 13 March 2018
Accepted 28 March 2018

Keywords:
Mammography
False-positive reduction
Breast cancer diagnosis
Hilbert's image representation

ABSTRACT

Background and objective: The aim of computer-aided-detection (CAD) systems for mammograms is to assist radiologists by marking region of interest (ROIs) depicting abnormalities. However, the confusing appearance of some normal tissues that visually look like masses results in a large proportion of marked ROIs with normal tissues. This paper copes with this problem and proposes a framework to reduce false positive masses detected by CAD.

Methods: To avoid the error induced by the segmentation step, we proposed a segmentation-free framework with particular attention to improve feature extraction and classification steps. We investigated for the first time in mammogram analysis, Hilbert's image representation, Kolmogorov–Smirnov distance and maximum subregion descriptors. Then, a feature selection step is performed to select the most discriminative features. Moreover, we considered several classifiers such as Random Forest, Support Vector Machine and Decision Tree to distinguish between normal tissues and masses. Our experiments were carried out on a large dataset of 10168 ROIs (8254 normal tissues and 1914 masses) constructed from the Digital Database for Screening Mammography (DDSM). To simulate practical scenario, our normal regions are false positives asserted by a CAD system from healthy cases.

Results: The combination of all the descriptors yields better results than each feature set used alone, and the difference is statistically significant. Besides, the feature selection steps yields a statistically significant increase in the accuracy values for the three classifiers. Finally, the random forest achieves the highest accuracy (81.09%), outperforming the SVM classifier (80.01%)) and decision tree (79.12%), but the difference is not statistically significant.

Conclusions: The accuracy of discrimination between normal and abnormal ROIs in mammograms obtained with the proposed gray level texture features sets are encouraging and comparable to these obtained with multiresolution features. Combination of several features as well as feature selection steps improve the results. To improve false positives reduction in CAD systems for breast cancer diagnosis, these features could be combined with multiresolution features.

1. Introduction

Breast cancer is the most diagnosed female cancer and the second leading cause of cancer deaths among women worldwide. In 2012, this disease affected 1.67 million new cases (25% of all female cancers) and killed 522,000 women (14.7% of cancer-related deaths) [8]. Breast cancer is arising from cells of the breast that develops locally in the breast and metastasizes to the lymph nodes and internal organs. Even though some studies pointed out several risk factors of having breast cancer such as genetics and tobacco, the cause of cell metastasize is not clearly known. This makes early prevention of breast cancer not possible. Thus, early detection is considered as the cornerstone in breast cancer treatment. A breast cancer detected at early stage is easy to handle, whereas

* Corresponding author.
E-mail addresses: sami.dhahbi@laposte.net (S. Dhahbi), walid.barhoumi@enicarthage.rnu.tn (W. Barhoumi), jaroslaw_kurek@sggw.pl (J. Kurek), bartosz_swiderski@sggw.pl (B. Swiderski), Michal_kruk@sggw.pl (M. Kruk), ezzeddine.zagrouba@fsm.rnu.tn (E. Zagrouba).

https://doi.org/10.1016/j.cmpb.2018.03.026
0169-2607/

late detection decreases treatment options and increases mortality rates. Unfortunately, traditional detection methods such as palpable breast examination or clinical women request (after pain) usually detect cancers at non-recoverable stages. Screening programs are applied to cope with this issue: women with higher risk (such as old women) but with no clinical symptoms are examined. The best cost-effective tool for breast cancer detection is currently screening mammography. It can detect breast abnormalities at an early stage when they are not detectable by a woman or doctor, what increases very significantly the chance of cure [16]. For instance, several developed countries established systematic nationwide screening mammography programs for early breast cancer detection: each woman over 50 years undergoes a mammography examination every two (or three) years. These programs helped in reducing the mortality rate of breast cancer [13,27].

However, mammography interpretation is a difficult and error prone task, even for skilled radiologists, mainly due to the subtle signs of breast abnormalities and the overlapping dense fibroglandular tissue. Misinterpretation of mammograms leads to false positives and/or false negatives. False positives -taking normal tissue for abnormalities- result in high recall rates with unnecessary treatments (such as invasive biopsies). False negatives -missing a cancer abnormality- complicate treatment options and threaten patient life. Due to the huge amount of screening mammograms to be analyzed and the limited number of radiologists double reading by a second radiologist is not feasible in screening programs. A cost-effective alternative to human double reading is the use of Computer Aided Detection systems (CAD), which may replace the second reader and alert the radiologist to suspicious regions. These systems have improved the detection rates of breast abnormalities such as micro-calcification and masses. Given that the detection of masses is much more challenging because some normal tissues may look visually similar to masses and can be taken for abnormalities, the aim of this work is to reduce mass false positives produced by CAD systems. Indeed, the main drawback of CAD systems for masses is false positives, which may increase unnecessary biopsy rates. Besides, given the low rate of cancer occurrence in screening programs, radiologists may over-look CAD outputs that contain suspicious regions. Hence, a posterior step of reduction of false positive rates in CAD systems is crucial to reduce unnecessary follow-up treatments and to ensure radiologistâs acceptance. Indeed, two issues are typically addressed in computer-aided detection of mammographic masses. The first one is the detection step, which aims to localize candidate ROIs that may depict abnormalities. The second is false positive reduction step, which aims to detect and eliminate normal ROIs. In this paper, we assume that we have already ROIs from the detection step and, thus, we address only the false positive reduction step.

Several studies have focused on reducing false positives for masses. Hence, given suspicious regions detected by a CAD system, false positive reduction (FPR) methods aim to classify these regions as normal or abnormal. A first class of methods aimed to include information from the current mammogram or other views. For instance, Vállez et al. [30] included a breast density classification step prior to lesion detection so that to improve the detection results. Li et al. [22] used a bilateral similarity analysis to combine information from right and left breast views. Tan et al. [29] proposed a score fusion to combine detection results of medio-lateral oblique (MLO) view and cranio-caudal (CC) view. A second class of methods solved the pattern recognition problem of distinguishing between normal tissues and masses in three steps: lesion segmentation, feature extraction, and binary classification. For instance, Junior et al. [17] used a spatial approach of diversity indexes to describe patterns detected in previously segmented regions and a SVM classifier to classify regions into masses and non-masses. Liu and Zeng [24] used adaptive region growing and narrow band based active contour for mammographic masses segmentation, gray level co-occurrence matrix (GLCM) and complete local binary patterns (CLBP) for the description of the segmented lesions, and SVM for classification. de Sampaio et al. [5] performed masses segmentation based on micro-genetic algorithm. Then, Local binary patterns and SVM were used to classify the suspicious regions. However, the segmentation of mammographic masses is a very difficult task and error-prone. Therefore, some authors proposed to omit the segmentation step and to compute the feature vector directly from the ROI. For instance, Dhahbi et al. [6] proposed a multiscale texture analysis method for false positive reduction based on curvelet moments. Curvelet transform was first performed on region of interest (ROI). Moment statistics (mean, variance, kurtosis and skewness) were then computed from each curvelet band and were input into a k-nearest-neighbor (kNN) classifier. A feature selection step was also included to select the most relevant curvelet moments. Zyout et al. [32] reduced false positive of mass detection based on multiscale textural feature extraction, particle swarm optimization and support vector machines (SVM) classification. Hussain [14] introduces a method for false positive reduction based on Weber law descriptor (WLD) and SVM classification. Scale-invariant feature transform (SIFT) features have been used to describe mammographic ROIs for scalable mammogram retrieval [15,23]. 2D wavelet transform modulus maxima method (WTMM), that was originally introduced to perform multifractal analysis of rough surfaces using the continuous wavelet transform [1], has been applied to quantify the roughness fluctuations in mammographic images and to segment these images into dense tissue regions, fat tissue regions and micro-calcifications [19]. Batchelder et al. [2] used the WTMM method to detect microcalcification in human breast tissue in mammograms and to characterize the fractal geometry of benign and malignant microcalcification clusters, the former being Euclidean whereas the latter being fractal.

Our main motivation in this work is to improve the performance of mammographic masses detection through false-positive reduction. To avoid the error induced by the segmentation step and to take into account the texture around the lesion [9–11,25], we proposed a segmentation-free framework with particular attention to improve feature extraction and classification steps. The flowchart of the proposed framework is given in Fig. 1.

Hence, we propose five different feature extraction methods to describe mammographic regions. The first method extracts 23 features based on Hilbert's image representation. The second method uses fractal texture analysis to extract 36 features. The third set is composed of 38 Kolmogorov–Smirnov distances. The fourth set consists of 15 features corresponding to statistics from the gray level co-occurrence matrix (GLCM). The last method is based on K-S statistic and Minkowsky approach to compute 10 features corresponding to maximum sub-region descriptors. To the best of our knowledge, Hilbert's image representation, Kolmogorov–Smirnov distances and maximum sub-region descriptors are introduced for the first time for mammographic masses analysis and false positive reduction. Once the 122 features are computed, we performed a sequential feature selection method to choose the most discriminative subset. For the classification step, even though the SVM classifier is the most common classifier for mammographic masses classification, other classifiers could yield better results. Therefore, we investigated three different classification methods: SVM, decision tree and random forest. The main contributions of the paper are as follows:

1. Effective representation of mammographic masses through Hilbert's image representation, Kolmogorov–Smirnov distances and maximum sub-region descriptors.

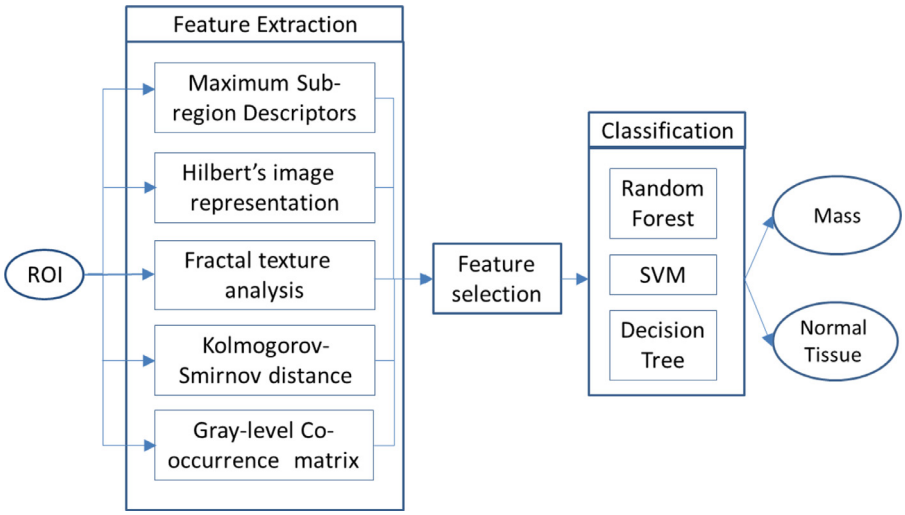

Fig. 1. Flowchart of the proposed framework.

2. Combination of the proposed three feature sets with two other commonly used sets: fractal texture features and gray level co-occurrence matrix statistics.
3. Selection of the most discriminative features for masses description.
4. Investigation of three different classification methods.
5. Experimental evaluation of the proposed framework in a large and challenging database Fig. 3.

The rest of this paper is organized as follows. Sections 2 describes the used mammograms in the experiments. Section 3 presents the feature extraction methods we used to generate diagnostic features. Section 4 recalls the feature selection method. Section 5 is devoted to numerical experiments. Conclusions are stated in Section 6.

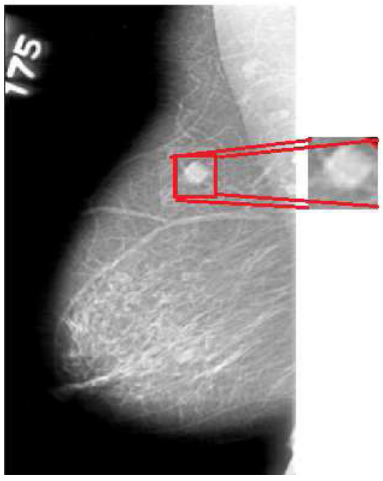

Fig. 2. Illustration of region of interest (ROI) extraction.

2. Materials and methods

2.1. Materials

Our experimental dataset is constructed from the Digital Database for Screening Mammography (DDSM) [12], which is the largest public mammogram dataset. DDSM is an old database that includes 2604 cases made from digitized films, each case is composed of four mammograms corresponding to Cranial-Caudal (CC) and Medio-Lateral-Oblique (MLO) views for right and left breasts. The dataset includes the ground truth of each mammogram, mainly its diagnostic result (normal, benign or malignant) and the location of existing lesions.

Our experiments are carried out on a large dataset of 10,168 ROIs (8254 normal tissues and 1914 masses) constructed in [15] from the DDSM. For this, Jiang et al. [15] used a CAD program [31] to detect masses. False positives asserted by this CAD system from healthy cases [31] are our normal regions. For ROIs depicting masses, a manual cropping was performed to extract rectangular areas centered on the coordinates of lesions provided in the ground truth (Fig. 2). Hence, compared to common databases in which normal cases are randomly selected, our dataset is more realistic and more challenging [15].

In the experiments, all the ROIs were resized to 256×256 pixels. Fig. 2 illustrates examples of ROIs used in the experiments.

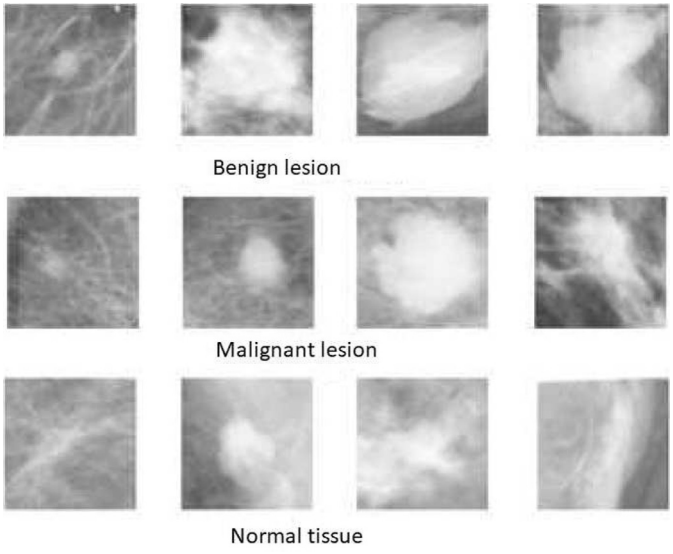

Fig. 3. Sample of ROIs from the DDSM database.

2.2. Generation of diagnostic features

Our whole database of masses consists of totally 10,168 cropped images (ROIs). This set has been divided into normal tissues and abnormal tissues, based on accurate diagnosis performed manually by doctors. Our approach now is to generate features that have the capability to distinguish between normal tissues and masses.

During the manual analysis of the images of normal and abnormal tissues, it is easy to notice that the main differences between normal tissues and masses are connected with the shape, size, granularities, texture and intensity of the images associated with different states of tissues. In our approach to this problem, we will base on the features referred to characteristics of chaotic image, changes in the features distribution, co-occurrence Haralick's matrix, fractal dimension and regular statistical features.

On the basis of all extracted ROIâs, we generated 122 potential diagnostic features in total for binary classification. Features have been extracted based on every gray-scale image from the whole database. The dataset of potential diagnostic features consists of the following groups:

- features generated based on Hilbert's image representation (35 features) .
- features generated based on coaxial rings image representation and Kolmogorov–Smirnov distance approach (14 features).
- features generated based on Maximum regions descriptors (6 features).
- features generated based on forest fire modeling (2 features).
- features generated based on the gray-level co-occurrence matrix - GLCM (4 features).
- features generated based on box-counting fractal dimension (9 features).
- features generated based on segmentation-based fractal texture analysis (36 features).
- features generated based on 4-level wavelet packet decomposition (16 features).

2.2.1. Hilbert's image representation

A Hilbert curve or Hilbert space-filling curve is a continuous fractal space-filling curve, first described by the German mathematician David Hilbert in 1891. It was one variant of filling space with curves discovered by Giuseppe Peano in 1890. The Hilbert's curve (Fig. 4) appears to have useful characteristics.

Both the true Hilbert curve and its discrete approximations are useful because they provide a mapping between 1D and 2D space that fairly well preserves locality [15,26].

For many extracted features (especially statistical features), where one dimensional matrix (vector) is required as an algorithm input, we should find the best representation of every grayscale image. One of the best 1-dimensional image representation is the application of the Hilbert' curve. One-dimensional matrix is the result of covering discrete Hilbert's transformation mask for the image. Hence, this one dimensional matrix can be treated as a vector and can be a base for any feature extraction algorithm, where one dimensional matrix is strictly required.

Let's assume that we have to fill the space with curves and the space has a square size $n \times m$, where n is the power of 2 (e.g. 4, 8, 16, 32). Hilbert's curves of order 0 to 5 are depicted on Fig. 4. Based on the above assumption, we can generate the discrete Hilbert's curves using the following recursive formula:

$$\begin{cases} x_0 = 0 \\ y_0 = 0 \\ x_n = \frac{1}{2}[(y_{n-1} - 0.5)(x_{n-1} - 0.5)(x_{n-1} + 0.5)(-y_{n-1} + 0.5)] \\ y_n = \frac{1}{2}[(x_{n-1} - 0.5)(y_{n-1} + 0.5)(y_{n-1} + 0.5)(-x_{n-1} - 0.5)] \end{cases} \quad (1)$$

The advantages of applying the discrete Hilbert's curve for one-dimensional image representation are as follows:

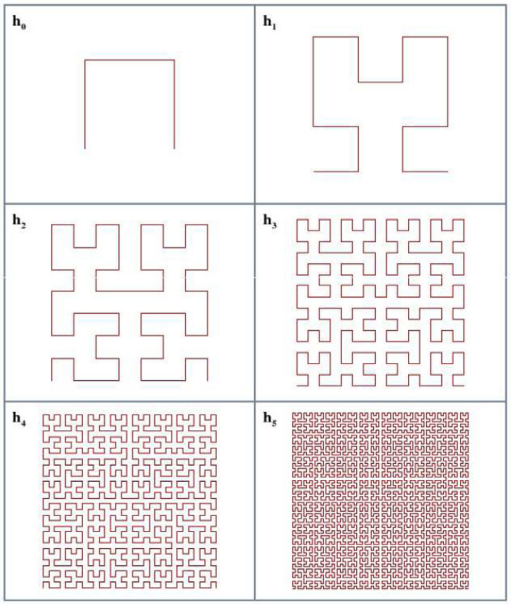

Fig. 4. Example of discrete Hilbert's curve for order 0 to 5.

- mapping between 1D and 2D space fairly well preserves locality
- when we follow every point of discrete Hilbert's curve, we follow through the image neighbor pixels
- if (x, y) are the coordinates of a point within the unit square, and d is the distance along the curve, when it reaches that point, then the points that have nearby d values will also have nearby (x, y) values.

In this article, we have constructed the Hilbert's curve of order 1024. Then the discrete Hilbert's curve has been mapped to the image. The last stage is the creation of a vector image representation, which consists of all points of the discrete Hilbert's curve. The Matlab code generating the discrete Hilbert's curve is presented below:

```
function [x,y] = hilbert(n)
if n<=0
x = 0;
y = 0;
else
[xo,yo] = hilbert(n-1);
x = .5*[-.5+yo -.5+xo .5+xo  .5-yo];
y = .5*[-.5+xo  .5+yo .5+yo -.5-xo];
end
```

2.2.2. Potential 35 features generated based on Hilbert's image representation

After converting matrix image representation to vector image representation in the form of the discrete Hilbert's curve, the authors have applied many statistic measures where a vector form is required as an input. Some of these statistic measures applied by the authors are presented in Table 1. In Total, the authors defined 27 statistic features on the basis of the discrete Hilbert's curve image representation.

2.2.3. Kolmogorov–Smirnov distance

For some new texture segmentation features generated from the images, the authors have applied Kolmogorov Smirnov (KS) statistical distance. Some propositions of applying this statistical measure to the medical image characterization have been already presented in (Demidenko, 2004, Pauwels and Frederix, 2000).

The KS test determines if the samples are drawn from the same underlying continuous population characterized by the cumulative

Table 1
Some statistical features based on Hilbert's curve image representation.

Name	Equation
Mean	$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
Standard deviation	$\sigma = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(x_i-\mu)^2}$
Skewness	$S = \frac{1}{n}\sum_{i=1}^{n}(\frac{x_i-\mu}{\sigma})^3$
Kurtosis	$K = \frac{1}{n}\sum_{i=1}^{n}(\frac{x_i-\mu}{\sigma})^4 - 3$
RMS	$\sqrt{\frac{\sum_{i=1}^{n}\lvert x_i\rvert}{n}}$
Crest factor	$C = \frac{x_{peak}}{x_{RMS}}$

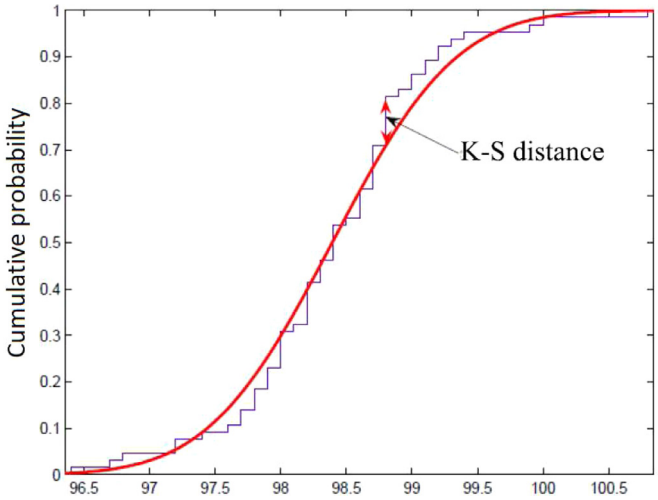

Fig. 5. Example of Kolmogorov–Smirnov test statistic.

distributions $F(x_i)$ and $F(x_j)$ (Fig. 5). The distance between these two populations is defined by the KS test in the following formula:

$$d_{KS} = max|F(x_i) - F(x_j)| \quad (2)$$

2.2.4. Representation of the neighborhood pixel as coaxial rings

The aim of this approach is to find features which provide us with the information on the evolution of the 3-dimensional (R,G,B) pixelâs distribution surrounding this point [21]. These features should indicate what is a pace of changes of neighborhood pixel colors distribution. In our experiment, only the grayscale image has been taken into consideration, hence, we should measure the pace of changes of pixel intensity distribution. To execute the calculation in a reasonable time, only 4 coaxial rings have been constructed, each ring is composed of 56 pixels (Fig. 5). To compare mentioned distribution, we performed the Kolmogorov–Smirnov test. The representation of the neighborhood pixel in the form of coaxial rings has been depicted on Fig. 6.

Instead of constructing coaxial rings for the whole image, we divided the image into subimages of equal size and we constructed coaxial rings for each subimage (Fig. 7). For this, analyzed gray scale image was scaled to 289 × 289 pixels and subdivided into 289 subimages of size 17 × 17 pixels. A set of 4 coaxial rings was then constructed for each subimage. In this way, each image is represented by 289 sets of coaxial rings.

For every set of 4 coaxial rings corresponding to a subimage, we compute the *KS* distances between proper rings as described in Table 2.

Once the *KS* distances were computed for all subimages of a given image, we calculated the average *KS* statistic of coaxial rings of the same level. Then, we performed linear regression to the relationship of the average and median KS distance d_{KS} as the function

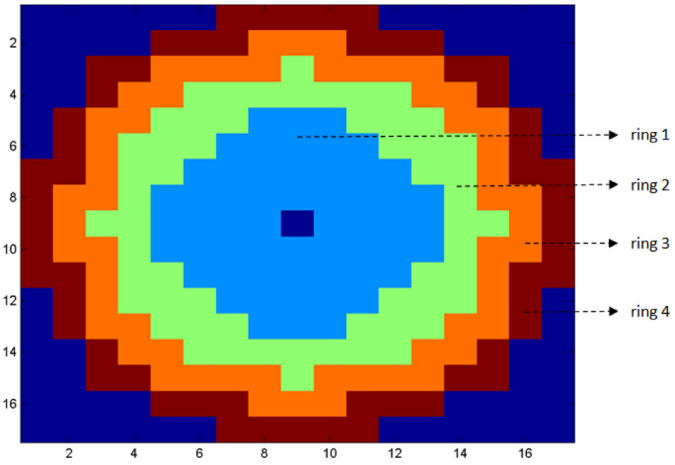

Fig. 6. Representation of the neighborhood pixel as coaxial rings.

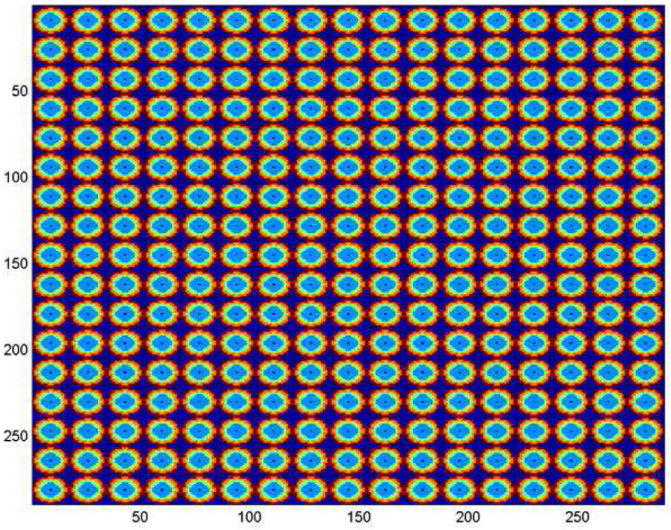

Fig. 7. Subdivision of image into 17 × 17 subimages. Each subimage is represented by a set of 4 coaxial rings.

Table 2
Calculation of Kolmogorov–Smirnov distance between coaxial rings.

No	$Ring_i$	$Ring_j$	Skipped level ($\lvert i-j\rvert$)	KS
1	1	2	1	KS1
2	1	3	2	KS2
3	1	4	3	KS3
4	2	3	1	KS4
5	2	4	2	KS5
6	3	4	1	KS6

of the level l (Eq. (3)).

$$d_{KS} = \alpha_0 + \alpha_1 l + \varepsilon \quad (3)$$

where l denotes skipped level, and α_0 and α_1 represent the regression coefficients.

Finally, based on coaxial rings image representation and Kolmogorov–Smirnov distance, 14 features have been calculated. These features have been presented in Table 3.

2.2.5. Maximum regions descriptors

The main idea is to observe disaggregating the image to smaller consistent subgroups on the basis of the thresholding. We have searched for the threshold level from which most subgroups are

Table 3
Definition of 14 features based on coaxial rings image representation and Kolmogorov–Smirnov distance approach.

No	Definition	Description
1	$d_{KSmean}12$	mean of KS statistics between ring No. 1 and ring No. 2
2	$d_{KSmean}13$	mean of KS statistics between ring No. 1 and ring No. 3
3	$d_{KSmean}14$	mean of KS statistics between ring No. 1 and ring No. 4
4	$\frac{d_{KSmean}13}{d_{KSmean}12}$	$d_{KSmean}13$ over $d_{KSmean}12$ ratio
5	$\frac{d_{KSmean}14}{d_{KSmean}12}$	$d_{KSmean}14$ over $d_{KSmean}12$ ratio
6	α_{0mean}	intercept coefficient of approximation line d_{KSmean}
7	α_{1mean}	slope coefficient of approximation line d_{KSmean}
8	$d_{KSmedian}12$	median of KS statistics between ring No. 1 and ring No. 2
9	$d_{KSmedian}13$	median of KS statistics between ring No. 1 and ring No. 3
10	$d_{KSmedian}14$	median of KS statistics between ring No. 1 and ring No. 4
11	$\frac{d_{KSmedian}13}{d_{KSmedian}12}$	$d_{KSmedian}13$ over $d_{KSmedian}12$ ratio
12	$\frac{d_{KSmedian}14}{d_{KSmedian}12}$	r $d_{KSmedian}14$ over $d_{KSmedian}12$ ratio
13	$\alpha_{0median}$	intercept coefficient of approximation line $d_{KSmedian}$
14	$\alpha_{1median}$	slope coefficient of approximation line $d_{KSmedian}$

Fig. 8. Input image for maximum regions descriptors approach.

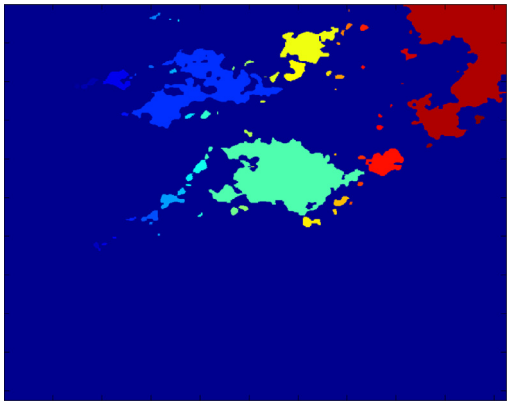

Fig. 9. Output image for maximum regions descriptors approach.

derived. To obtain the result for dozen thousands images in a reasonable period of time, percentiles of pixels intensity have been applied. For each image, we search for the quantile q and its intensity threshold value th_q, which splits the image into the largest number of compact groups of pixels. The searching is performed using the quantiles of pixel's intensity, i.e., 0.01, 0.02, ... , 0.99. The application of Maximum Regions Descriptors application is depicted on Figs. 8 and 9.

After finding the threshold associated with disaggregating the image to most subgroups, we can obtain values of our 3 features.

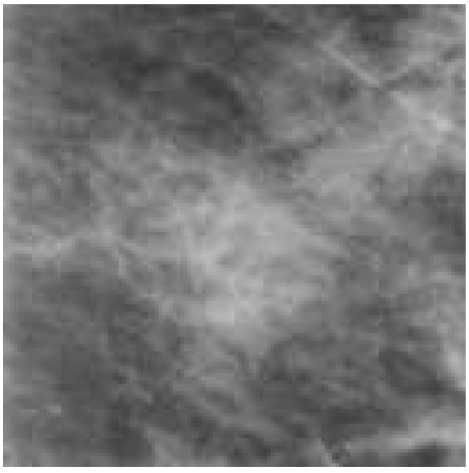

Fig. 10. Image input of forest fire modeling approach.

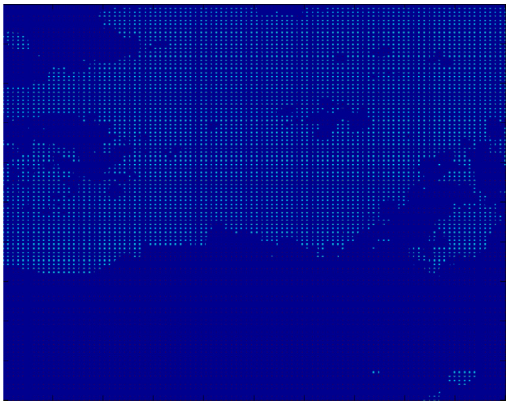

Fig. 11. Setting fire to the forest.

These features can be calculated for order operators $<$ and $>$, hence we have a total of 6 features (Table 4).

2.2.6. Percolation coefficients

The goal in this approach is the use of forest fire modeling to create a potential diagnostic feature based on an image. The idea is based on 9 thresholds (for each decile) of a an image. For every result of thresholding we start setting fire to the forest and then measure the duration of the fire (number of iterations). The fire can be spread only on thresholded area. The authors have applied 8 neighborhood pixels.

The results of stages of 3 of the described process applied to the image on Fig. 9 are depicted on Figs. 10–13.

The output of this process applied to this example is showed in Table 5.

Finally, we computed a potential feature q_w as follows:

$$q_w = \frac{\sum_{i=1}^{9} q_i d_i}{\sum_{i=1}^{9} d_i} \tag{4}$$

The above equation can be applied to two order operators ($\geq$ and $\leq$), resulting thus in 2 potential features.

2.2.7. Potential 4 features from the gray-level co-occurrence matrix and other statistics

Four features have been generated using gray-level co-occurrence matrix (GLCM) from a given image. GLCM calculates how often a pixel with the gray-level value i occurs adjacent to a pixel with the value j. Based on the GLCM approach, 4 potential features have been generated and presented in Table 6:

Table 4
Definition of 6 features based on maximum regions descriptors approach.

No	Definition	Description
1	$p \geq$	associated value of percentile (percentile to intensity: 0.01 â0 ; 0.99â255) for sign $\geq$
2	$q \geq$	associated percentile (order of quantile) for sign $\geq$
3	$a \geq$	area of largest subgroup for sign $\geq$
4	$p \leq$	associated value of percentile (percentile to intensity: 0.01 â0 ; 0.99â255) for sign $\leq$
5	$q \leq$	associated percentile (order of quantile) for sign $\leq$
6	$a \leq$	area of largest subgroup for sign $\leq$

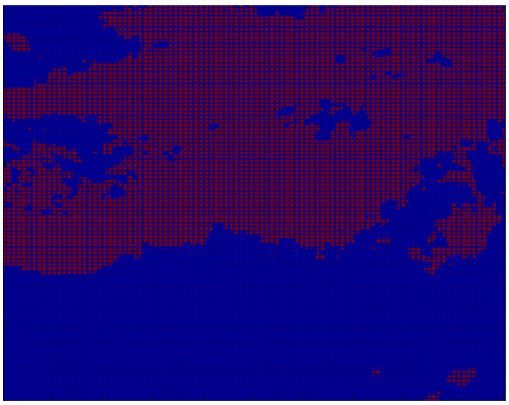

Fig. 12. Image after a few iterations.

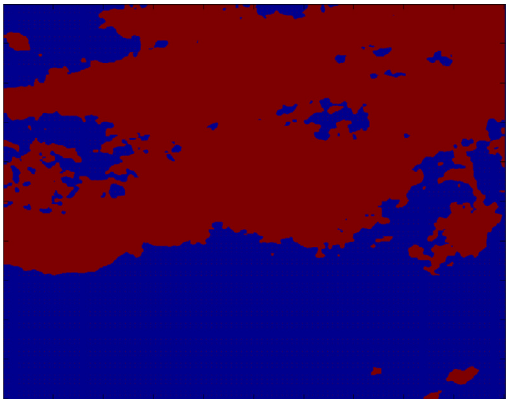

Fig. 13. Result of forest fire modeling after 40 iterations.

Table 5
Result of forest fire modeling.

q_i	d_i
0.1	31
0.2	40
0.3	31
0.4	30
0.5	40
0.6	24
0.7	29
0.8	27
0.9	22

Table 6
Definition of 4 features from histogram of co-occurring gray-scale (GLCM).

Name	Description
Contrast	measure of the intensity contrast between a pixel and its neighbor over the whole image
Correlation	measure of how correlated a pixel is to its neighbor over the whole image
Energy	sum of squared elements in the GLCM
Homogeneity	measure of the closeness of the distribution of elements in the GLCM to the GLCM diagonal

2.2.8. Potential 9 features based on box-counting fractal dimension

The box-counting fractal dimension is a measure characterizing the fractal complexity. It is the particular case of the Mandelbrot fractal dimension, based on the notion of self-similarity of the structure at different scales. It measures how the length of the complex curve changes when the measurement is performed with an increased accuracy [28]. To characterize any curve, this curve is covered with the set of regular squared areas of size ε. Thus, we have to calculate the number of squared areas containing any part of the given curve. Then number of found areas is denoted as N(ε) (Formula 5). In our approach, we have used 9 boxes.

$$d = \lim_{\varepsilon \to 0} \frac{log(N(\varepsilon))}{log(\frac{1}{\varepsilon})} \tag{5}$$

2.2.9. 36 Features on the basis of segmentation-based fractal texture analysis

This approach allows us to generate 36 texture features on the basis of the SFTA algorithm (Segmentation-based Fractal Texture Analysis) and returns 16 vectors extracted from the input gray scale image. Applying 6th fractal order, we can obtain 36 texture features. The SFTA algorithm decomposes images into various thresholded images using several sets of lower and upper threshold values. The implementation has been based on Costa approach [3]. Thresholded images are used to extract the fractal dimension. Since images with more jagged edges and prominent color differences tend to have higher fractal dimension, images with more uniform texture properties will have a lower fractal dimension.

2.2.10. 16 Features based on 4-level wavelet packet decomposition

In this approach, the authors have applied Daubechies wavelets (db10) family [4]. After wavelet packet decomposition, the portion of energy for every terminal node has been calculated based on the following equation (Formula 6):

$$E = \Sigma_{k=1}^{N} |S_{jk}|^2 \tag{6}$$

where N is the number of wavelet sub-bands, k is a counter and S_{jk} are the appropriate discrete wavelet transformation coefficients.

3. Features selection

To choose the most appropriate diagnostic features amongst the 122 potentially available ones, best separating the two classes of tissues (normal and abnormal), we have applied the method of sequential feature selection to create a subset of the most discriminative features for breast cancer diagnosis.

This approach begins with an empty set and iteratively adds one feature among the remaining available features until there is no improvement in prediction [26]. At each stage, every feature among the remaining available ones is added to the current set to construct candidate feature subsets. Each candidate feature set was evaluated using an SVM classifier and 10-fold cross-validation. The feature that yields the best classification rate for the new enlarged feature set is selected. After applying sequential features selection, we obtained 48 diagnostic features from the 122 available ones.

Table 7
Accuracy (mean ± Standard deviation) of binary classification (normal vs. abnormal tissues) of mammography images using SVM, Decision tree and Random forest.

Feature set	SVM	Decision tree	Random forest
GLCM	63.32% (± 2.43)	63.19% (± 2.13)	64.66% (± 3.40)
Fractal analysis	66.61% (± 2.79)	64.55% (± 2.23)	67.36% (± 2.17)
Hilbert's image representation	57.18% (± 2.15)	58.18% (± 2.71)	61.88% (± 3.23)
Kolmogorov-Smirnov	55.87% (± 2.31)	55.66% (± 1.90)	58.22% (± 2.39)
Maximum sub-regions	60.26% (± 2.13)	62.92% (± 2.71)	61.52% (± 2.63)
All descriptors (without FS[a])	71.63% (± 0.96)	70.11% (± 2.53)	71.42% (± 3.14)
All descriptors (with FS[a])	80.01% (± 5.05)	79.12% (± 5.51)	81.09% 1 (± 4.51)

[a] FS: feature Selection.

4. Experimental study

In the experiments, we evaluated the performance of each proposed feature set for false-positive mammographic mass reduction. Furthermore, to highlight the importance of combining feature sets and feature selection steps, feature sets composed of all proposed descriptors (with and without feature selection) were considered. We have applied three types of classifiers: SVM, decision tree and random forest. For numerical experiments, the equal subset of normal and abnormal tissues trials have been chosen in a randomized approach. The 10-fold stratified cross validation has been performed to compute the classification accuracy of each classifier.

4.1. Results

Table 7 summarizes the classification results achieved by the different proposed feature sets using SVM, decision tree and random forest classifiers. Hence, accuracy is calculated as the ratio of normals classified as normals and total of normals.

We can see that feature sets obtained through the combination of all the descriptors (with and without feature selection) yield better results than each feature set used alone, regardless the classification method that is used, and the difference is statistically significant. This proves the usefulness of the combination step. Besides, the feature selection step increases the accuracy values for the three classifiers, and the improvement is statistically significant. Finally, the random forest achieves the highest accuracy (81.09%), outperforming decision tree (80.01%) and the SVM classifier (79.12%) that is commonly used for mammographic masses classification, but the difference is not statistically significant.

For comparison purposes, we implemented state-of-the-art curvelet-based mammogram characterization. Indeed, we have already used multiresolution texture analysis for mammogram characterization, and the proposed curvelet moments set outperforms state-of-the-art curvelet-based mammogram analysis methods [6]. To further improve mammogram characterization, we investigate in this paper gray level mammogram features. Bearing in mind the superiority of multiresolution texture features over gray level texture features, we are not looking for a gray level mammogram description that outperforms multiresolution texture features, but rather for the combination of the two approaches. Our findings show that the proposed feature set yields better results than several curvelet-based methods. Indeed, our highest accuracy result (Random Forest-81%) is slightly better (but not statistically significant) than biggest curvelet coefficients (BCC) (=80.65% std 0.65%), Statistical Curvelet Coefficients (SCC) (=80.01% std 0.58%) and co-occurence matrix of curvelet coefficients (=80.69% std 0.52%). Our method is only outperformed by Curvelet Level Moments (CLM) (86.46% std 0.91%), and the difference is statistically significant.

Table 8
Accuracy (mean ± Standard deviation) comparison of the proposed method against state-of-the-art methods for binary classification (normal vs. abnormal tissues) of mammography images.

Method	Accuracy (mean ± Std)
Biggest curvelet coefficients	80.65% ± 0.65
Statistical curvelet coefficients	80.01% ± 0.58
Co-occurrence matrix of curvelet coefficients	80.69% ± 0.52
Curvelet moments	86.46% ± 0.91
Proposed method	81% ± 4.5

4.2. Discussion

In this paper, we have investigated several gray level feature extraction methods for false positive reduction of mammogram masses and tested three classification methods. The poor results obtained with each set alone is not only due to the limitation of each feature set, but also to the challenging problem in hand. Combination of the proposed feature sets statistically improves the results, which shows the importance of combining several feature sets. Furthermore, the feature selection step reduce the size of the feature set and also statistically improves the accuracy. This shows the importance of feature selection step. For all the compared sets, the differences between the results gathered from the three classifiers are not statistically significant. This shows that, in our case, the feature extraction step affects more the results than the classification step. The superiority of curvelet moments show that multiresolution feature may be better for mass characterization than gray level features. Motivated by the encouraging results obtained by the proposed gray level texture features as well as curvelet moments, our future work will include combination of such features to further improve the characterization of mammogram ROIs Table 8.

Motivated by the great success of deep learning methods, several researchers have investigated it for mammogram analysis [7,20]. For instance, Kooi et al. have shown that a deep learning model in the form of a Convolutional Neural Network(CNN) trained on a large dataset of mammographic lesions outperforms a state-of-the art system in Computer Aided Detection (CAD) [20]. A comparison of the CNN to three experienced readers have shown that the network is not far from the radiologists performance, but still substantially below the mean of the readers. As pointed in [20], even though a large performance increase is still possible, However, further improving the network requires more training data. In this context, a possible future direction is to employ our feature set with the CNN, which may be more effective than adding thousands of extra samples to the training.

5. Conclusion

In this paper, we have investigated several gray level texture analysis methods for the characterization of mammogram ROIs. The proposed methods include gray level co-occurrence matrix, fractal analysis, Hilbert's image representation, Kolmogorov–Smirnov distance and maximum sub-region descriptors. By extracting features directly from the entire ROIs, we not only escape the challenging problem of mammographic mass segmentation, but we also take into account the texture surrounding the lesion, which is useful for breast cancer diagnosis [18]. A feature selection step was

also included to choose the optimal feature set. Several classifiers (Random Forest, Support Vector Machine and Decision Tree) were considered to distinguish normal tissues from masses. Empirical evaluation in a large database composed of challenging suspicious regions extracted from the DDSM database prove the efficiency of the suggested method for false-positive reduction in mammographic mass detection. Our future work will be devoted to the combination of our feature set with multi-resolution curvelet moments, which provide better accuracy than our proposed gray level feature set. Furthermore, we plan to consider deep learning for comparison in our future work. Another future direction is to test the proposed framework in the case of current digital full breast mammographic images.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

References

[1] A. Arneodo, N. Decoster, P. Kestener, S. Roux, A wavelet-based method for multifractal image analysis: from theoretical concepts to experimental applications, Adv. Imaging Electron Phys. 126 (2003) 1–92.
[2] K.A. Batchelder, A.B. Tanenbaum, S. Albert, L. Guimond, P. Kestener, A. Arneodo, A. Khalil, Wavelet-based 3d reconstruction of microcalcification clusters from two mammographic views: new evidence that fractal tumors are malignant and euclidean tumors are benign, PLoS ONE 9 (9) (2014) e107580.
[3] A.F. Costa, G. Humpire-Mamani, A. Traina, An efficient algorithm for fractal analysis of textures, in: Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on, IEEE, 2012, pp. 39–46.
[4] I. Daubechies, Ten Lectures on Wavelets, SIAM, 1992.
[5] W.B. de Sampaio, A.C. Silva, A.C. de Paiva, M. Gattass, Detection of masses in mammograms with adaption to breast density using genetic algorithm, phylogenetic trees, lbp and svm, Expert Syst. Appl. 42 (22) (2015) 8911–8928.
[6] S. Dhahbi, W. Barhoumi, E. Zagrouba, Breast cancer diagnosis in digitized mammograms using curvelet moments, Comput. Biol. Med. 64 (2015) 79–90.
[7] N. Dhungel, C. Gustavo, P.B. Andrew, A deep learning approach for the analysis of masses in mammograms with minimal user intervention, Med. Image Anal. 37 (2017) 114–128.
[8] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D. Forman, F. Bray, Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012, Int. J. Cancer 136 (5) (2015).
[9] E. Gerasimova, B. Audit, S.G. Roux, A. Khalil, O. Gileva, F. Argoul, O. Naimark, A. Arneodo, Wavelet-based multifractal analysis of dynamic infrared thermograms to assist in early breast cancer diagnosis, Front. Physiol. 5:176 (2014).
[10] E. Gerasimova-Chechkina, B. Toner, Z. Marin, B. Audit, S. Roux, F. Argoul, A. Khalil, O. Gileva, O. Naimark, A. Arneodo, Combining multifractal analyses of digital mammograms and infrared thermograms to assist in early breast cancer diagnosis, in: AIP Conference Proceedings, 1760, AIP Publishing, 2016, p. 020018.
[11] E. Gerasimova-Chechkina, B. Toner, Z. Marin, B. Audit, S.G. Roux, F. Argoul, A. Khalil, O. Gileva, O. Naimark, A. Arneodo, Comparative multifractal analysis of dynamic infrared thermograms and x-ray mammograms enlightens changes in the environment of malignant tumors, Front. Physiol. 7:336 (2016).
[12] M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer, R. Moore, K. Chang, S. Munishkumaran, Current status of the digital database for screening mammography, in: Digital mammography, Springer, 1998, pp. 457–460.
[13] S. Hofvind, G. Ursin, S. Tretli, S. Sebuødegård, B. Møller, Breast cancer mortality in participants of the norwegian breast cancer screening program, Cancer 119 (17) (2013) 3106–3112.
[14] M. Hussain, False-positive reduction in mammography using multiscale spatial weber law descriptor and support vector machines, Neural Comput. Appl. 25 (1) (2014) 83–93.
[15] M. Jiang, S. Zhang, H. Li, D.N. Metaxas, Computer-aided diagnosis of mammographic masses using scalable image retrieval, IEEE Trans. Biomed. Eng. 62 (2) (2015) 783–792.
[16] A. Jotwani, J. Gralow, Early detection of breast cancer, Mol.Diagn.Ther. 13 (6) (2009) 349–357.
[17] G.B. Junior, S.V. da Rocha, M. Gattass, A.C. Silva, A.C. de Paiva, A mass classification using spatial diversity approaches in mammography images for false positive reduction, Expert Syst. Appl. 40 (18) (2013) 7534–7543.
[18] A.N. Karahaliou, I.S. Boniatis, S.G. Skiadopoulos, F.N. Sakellaropoulos, N.S. Arikidis, E.A. Likaki, G.S. Panayiotakis, L.I. Costaridou, Breast cancer diagnosis: analyzing texture of tissue surrounding microcalcifications, IEEE Trans. Inf. Technol. Biomed. 12 (6) (2008) 731–738.
[19] P. Kestener, J.M. Lina, P. Saint-Jean, A. Arneodo, Wavelet-based multifractal formalism to assist in diagnosis in digitized mammograms, Image Anal. Stereol. 20 (3) (2011) 169–174.
[20] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Mérida, C.I. Sánchez, R. Mann, A. den Heeten, N. Karssemeijer, Large scale deep learning for computer aided detection of mammographic lesions, Med. Image Anal. 35 (2017) 303–312.
[21] M. Kruk, B. Świderski, S. Osowski, J. Kurek, M. Słowińska, I. Walecka, Melanoma recognition using extended set of descriptors and classifiers, Eurasip J. Image Video Process. 2015 (1) (2015) 43.
[22] Y. Li, H. Chen, Y. Yang, L. Cheng, L. Cao, A bilateral analysis scheme for false positive reduction in mammogram mass detection, Comput. Biol. Med. 57 (2015) 84–95.
[23] J. Liu, S. Zhang, W. Liu, X. Zhang, D.N. Metaxas, Scalable mammogram retrieval using anchor graph hashing, in: Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on, IEEE, 2014, pp. 898–901.
[24] X. Liu, Z. Zeng, A new automatic mass detection method for breast cancer with false positive reduction, Neurocomputing 152 (2015) 388–402.
[25] Z. Marin, K.A. Batchelder, B.C. Toner, L. Guimond, E. Gerasimova-Chechkina, A.R. Harrow, A. Arneodo, A. Khalil, Mammographic evidence of microenvironment changes in tumorous breasts, Med. Phys. 44 (4) (2017) 1324–1336.
[26] I. Mathworks, Matlab User Manual, Mathworks, Natick, 2015.
[27] H.D. Nelson, K. Tyne, A. Naik, C. Bougatsos, B.K. Chan, L. Humphrey, Screening for breast cancer: an update for the us preventive services task force, Ann. Intern. Med. 151 (10) (2009) 727–737.
[28] M. Schroeder, Fractals, chaos, power laws: minutes from an infinite paradise, Cour.Corp., 2009.
[29] M. Tan, J. Pu, B. Zheng, Reduction of false-positive recalls using a computerized mammographic image feature analysis scheme, Phys. Med. Biol. 59 (15) (2014) 4357.
[30] N. Vállez, G. Bueno, O. Déniz, J. Dorado, J.A. Seoane, A. Pazos, C. Pastor, Breast density classification to reduce false positives in cade systems, Comput. Methods Programs Biomed. 113 (2) (2014) 569–584.
[31] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, 1, IEEE, 2001. I–I
[32] I. Zyout, J. Czajkowska, M. Grzegorzek, Multi-scale textural feature extraction and particle swarm optimization based model selection for false positive reduction in mammography, Comput. Med. Imaging Gr. 46 (2015) 95–107.