BMC Bioinformatics

A machine learning approach on multiscale texture analysis for breast microcalcification diagnosis

Annarita Fanizzi[1], Teresa M. A. Basile[2,3], Liliana Losurdo[1*], Roberto Bellotti[2,3], Ubaldo Bottigli[4], Rosalba Dentamaro[1], Vittorio Didonna[1], Alfonso Fausto[5], Raffaella Massafra[1], Marco Moschetta[6], Ondina Popescu[1], Pasquale Tamborra[1], Sabina Tangaro[3] and Daniele La Forgia[1]

Abstract

Background: Screening programs use mammography as primary diagnostic tool for detecting breast cancer at an early stage. The diagnosis of some lesions, such as microcalcifications, is still difficult today for radiologists. In this paper, we proposed an automatic binary model for discriminating tissue in digital mammograms, as support tool for the radiologists. In particular, we compared the contribution of different methods on the feature selection process in terms of the learning performances and selected features.

Results: For each ROI, we extracted textural features on Haar wavelet decompositions and also interest points and corners detected by using Speeded Up Robust Feature (SURF) and Minimum Eigenvalue Algorithm (MinEigenAlg). Then a Random Forest binary classifier is trained on a subset of a sub-set features selected by two different kinds of feature selection techniques, such as filter and embedded methods. We tested the proposed model on 260 ROIs extracted from digital mammograms of the BCDR public database. The best prediction performance for the normal/abnormal and benign/malignant problems reaches a median AUC value of 98.16% and 92.08%, and an accuracy of 97.31% and 88.46%, respectively. The experimental result was comparable with related work performance.

Conclusions: The best performing result obtained with embedded method is more parsimonious than the filter one. The SURF and MinEigen algorithms provide a strong informative content useful for the characterization of microcalcification clusters.

Keywords: Computer-aided diagnosis, Microcalcifications, Digital mammograms, Haar wavelet transform, SURF, Minimum eigenvalue algorithm, Random forest, Feature selection

Background

Breast cancer is the first cause of death among women and, although it is difficult to prevent, an early diagnosis of breast lesions increases the chances of survival and reduce the mortality rate [1]. Currently, screening programs use mammography [2, 3] as primary diagnostic tool for detecting breast cancer at an early stage. However the identification of some lesions remains still difficult

for radiologists. In particular, the 55% of breast diseases with tumor lesions are accompanied by the presence of microcalcifications (MCs), that are tiny spots of calcium deposits localized or broadly diffused on the breast areas, especially when they appear extremely minute (sometimes they do not exceed 0.1 mm) and grouped in clusters.

The diagnosis of microcalcifications is usually based on radiologists expertise resulting in some cases in inaccurate lesion detection [4–6] or in performing unnecessary breast biopsies on benign calcification clusters. This limit

*Correspondence: lilianalosurdo@gmail.com
[1]I.R.C.C.S. Istituto Tumori "Giovanni Paolo II", viale O. Flacco 65, Bari, Italy
Full list of author information is available at the end of the article

Fanizzi *et al. BMC Bioinformatics* 2020, **21**(Suppl 2):91

Page 2 of 11

becomes more evident in women with dense breast tissue that can hide lesions causing cancer to be detected at later stages [7, 8]. To overcome such limits, a solution is represented by the double blind reading of the mammograms by two radiologists [9] with a consequent higher workload and cost. A more interesting solution could be represented by using intelligent techniques to automatize the process of identification, normal vs abnormal tissue, and diagnosis, benign vs malignant, of clustered microcalcifications.

Several works have presented computerized methods to detect abnormalities in mammograms, playing a key role in the early detection of breast cancer thus helping to reduce the mortality rate due to breast pathologies in a cost-effective manner [5]. Such methods are known as Computer-Aided Detection/Diagnosis (CAD) systems and may offer to radiologists a reliable support in the evaluation of mammographic images [4, 10–12].

Many methods have been proposed to achieve a robust mammography-based CAD system for microcalcification diagnosis [13–18], in some cases well performing in dealing with specific abnormalities. Nevertheless, the automatic and accurate classification of microcalcification clusters, especially in differentiating the benign from the malignant ones, remains still complicated due to their nature. For this reason, the focus of this study is on the identification of a general model able to discriminate and, at the same time, characterize breast tissue and lesions with the aim of facing the fundamental challenge of improving the accuracy of breast lesion identification in order to decrease unnecessary biopsies and later surgeries. Accordingly, in the proposed model, an important role should be played by the features extraction and selection process, used to describe and characterize the regions of interest (ROIs), as well as by the classifier employed in the last phase of the CAD scheme that should be responsible of the decision regarding the origin, benign or malignant, of the region.

For what concerns the extraction of a representative set of features from the ROIs, different models have been proposed in literature ranging from those describing shapes of the clusters and classical statistical measurements [10, 19] to the ones exploiting morphological features [5, 16]. In some works, the texture analysis model uses a set of local statistical properties of pixel intensity. The textural features are obtained by the decomposition of the image into different frequency sub-bands by a wavelet transform [13, 14, 20] or by considering the spatial relationship between pixels with different gray-levels using the gray level co-occurrence-matrix [21–24]. Differently from textural descriptions of the breast lesions, some works concentrate on the potential correlation between the topology of clustered MCs and their pathological nature [17, 25]. Overall, all the works reported in literature use a broad variety of machine learning techniques such as k-Nearest Neighbours (kNN) [17, 23], Artificial Neural Networks (ANN) [14], and Support Vector Machines (SVM) [13, 14, 21] to build a classifier model able to discriminate the ROIs containing microcalcifications as benign or malignant using the extracted feature sets.

The development and integration of such a tool able to classify breast regions seem to be the natural prosecution of work presented in [26] where a CAD system working on full-field digital mammograms for the detection of clustered microcalcifications is reported. Indeed, as a succeeding step one expects the CAD system to perform a classification of the identified regions firstly in normal or abnormal tissue, as to reduce the erroneously detected regions that represent the false positive instances, and successively, on the abnormal classified regions, in benign or malignant lesions so to reduce recalls for unnecessary and stressful biopsies or ultrasound scans.

In this work, we propose the exploitation of texture analysis methods combined with machine learning techniques in order to characterize breast regions into normal/abnormal tissue and successively into benign/malignant lesions. The process is a multi-phase approach made up of a feature extraction step, performed by texture analysis methods, a features selection step, carried out with filter and embedded methods, and a breast region classification, performed by means of machine learning techniques, in order to categorize clusters of MCs in digital mammograms. Specifically, for each ROI a set of well-defined textural features, such standard statistical features, on a multiscale decomposition of the image based on the Haar wavelet transform [27, 28] are extracted. Moreover, interest points and corners are detected by using Speeded Up Robust Feature (SURF) [29] and Minimum Eigenvalue Algorithm (MinEigenAlg) [30], respectively. Successively, two different kinds of feature selection techniques, such as filter and embedded methods [31, 32] are exploited. Specifically, the filter methods include algorithms that evaluate the capacity of the individual features to predict the expected result. Embedded methods allow optimization between the interaction of the selected features and the classification algorithm used. Finally, a training test by means of a state of art classifier, such as Random Forest [33], is performed to classify the clustered microcalcifications. The proposed approach was tested on full-field digital mammograms extracted from the public database BCDR (Breast Cancer Digital Repository − https://bcdr.ceta-ciemat.es) [34]. The model performance was tested in cross-validation and evaluated in terms of accuracy, sensitivity and specificity, obtaining results in agreement with the literature.

Fanizzi *et al. BMC Bioinformatics* 2020, **21**(Suppl 2):91

Page 3 of 11

Materials and methods

Dataset selection

The image dataset consisted of a set of digital mammograms randomly selected from the Breast Cancer Digital Repository [34] (BCDR). Currently, the BCDR contains cases of 1734 patients with mammography and ultrasound images, clinical history, lesion segmentation and selected pre-computed image-based descriptors. Patient cases are classified according to BIRADS categories [35] and annotated by specialized radiologists covering all the possibilities of diagnosis. In the database all available medio-lateral oblique (MLO) and cranial caudal (CC) views of the left and right breast are included. The BCDR is subdivided in two different repositories: (1) a Film Mammography-based Repository (BCDR-FM) and (2) a Full Field Digital Mammography-based Repository (BCDR-DM). In particular, BCDR-DM includes 724 patient cases with digital mammograms. The MLO and CC images are gray-level mammograms with a resolution of 3328 (width) by 4084 (height) or 2560 (width) by 3328 (height) pixels, depending on the compression plate used in the acquisition (according to the breast size of the patient).

For this study, digital mammograms from BCDR-DM both in MLO and CC views were considered. Since the BCDR images reported the segmentation of the main lesions only, each extracted image was evaluated in double blind by two radiologists of our Institute dedicated to senological diagnostics, which have manually identified and classified ROIs containing the microcalcification clusters. Then, after a comparison between these independent readings, only ROIs for which both radiologists agreed were taken into account. As result, the dataset exploited in this study consists of 130 ROIs with clustered MCs, where 75 benign and 55 malignant, and 130 ROIs without any pathology.

Textural feature extraction

In this paper, we propose a fully automated model for the characterization of regions containing clustered microcalcifications in digital mammograms mainly based on a texture analysis approach. Since a fundamental property of the image texture is the scale at which the image is observed and analyzed, in this work a wavelet transform based on multiscale texture analysis approach, and specifically the Haar wavelet transform, was exploited. The Haar wavelet [27, 28] is a sequence of rescaled "square-shaped" functions which together form a wavelet family or basis. Wavelet approach is similar to Fourier analysis in that it allows a target function over an interval to be represented in terms of an orthonormal basis. This basis is composed by scaled and translated basis functions and denoted as $\phi(x,y)$ and $\psi(x,y)$. Conceptually, the scaling function is the low frequency component of the scaling function in 2 dimensions, and therefore there is one 2D

scaling function. The translated function has three different wavelet components, namely horizontal, vertical and diagonal. However, the wavelet function is related to the order at which apply the low- and high-filters and, since the wavelet function is separable, i.e.

$$f(x,y) = f_1(x)f_2(y)$$

these functions can be written as follows:

$$
\begin{aligned}
\phi(x,y) &= \phi(x)\phi(y) \rightarrow LL = \text{low} - \text{low}, \\
\psi^H(x,y) &= \psi(x)\phi(y) \rightarrow HL = \text{high} - \text{low}, \\
\psi^V(x,y) &= \phi(x)\psi(y) \rightarrow LH = \text{low} - \text{high}, \\
\psi^D(x,y) &= \psi(x)\psi(y) \rightarrow HH = \text{high} - \text{high},
\end{aligned}
\tag{1}
$$

where the corresponding filter order is denoted [27].

The Haar sequence is recognized as the first known wavelet basis and extensively used as a teaching example. In the 2D Haar wavelet decomposition on the image, the original image is first low-pass filtered and downscaled, yielding an approximation coefficients sub-image (LL_1 in Fig. 1a, top left), and then high-pass filtered, yielding the three detail coefficients sub-images (Fig. 1a, top right: horizontal HL_1, bottom left: vertical LH_1, and bottom right: diagonal HH_1), according to the general form of 2D wavelet transform (Eq. 1). To compute the successive level of decomposition, the process is iterated on LL_1, i.e. the approximation coefficient sub-image (Fig. 1b, top left). Specifically, in this study, we performed the 2D Haar Transform at two levels of decomposition.

However, to perform texture analysis, a number of attributes or descriptors that differentiate the textures have to be identified. Of course, such descriptors are assumed to be uniform within the regions with the same texture. Many works in literature report the texture analysis process based on first- or second-order statistics computed on the image histogram. The use of such texture descriptors rely on the assumption that texture can be defined by local statistical properties of pixel gray levels. For this reason, in our study, for each of the eight sub-images obtained in the Haar decomposition (LL, HL, LH and HH for levels 1 and 2) the following features are computed: mean, variance, skewness, kurtosis, entropy, relative smoothness; thus resulting, for each ROI, in a set of 48 Statistical Features (SF set).

Interest point/corner detection

As pointed out, the microcalcifications are characterized as to be tiny spots of calcium deposits localized or broadly diffused alone the breast areas or in some cases extremely minute and grouped in clusters. For this particular characterization of such lesions, in our model we enriched the information coming from the texture analysis with the information about points and corners of interest that can

Fig. 1 Image Haar decomposition. **a** One- and **b** two- level of decomposition
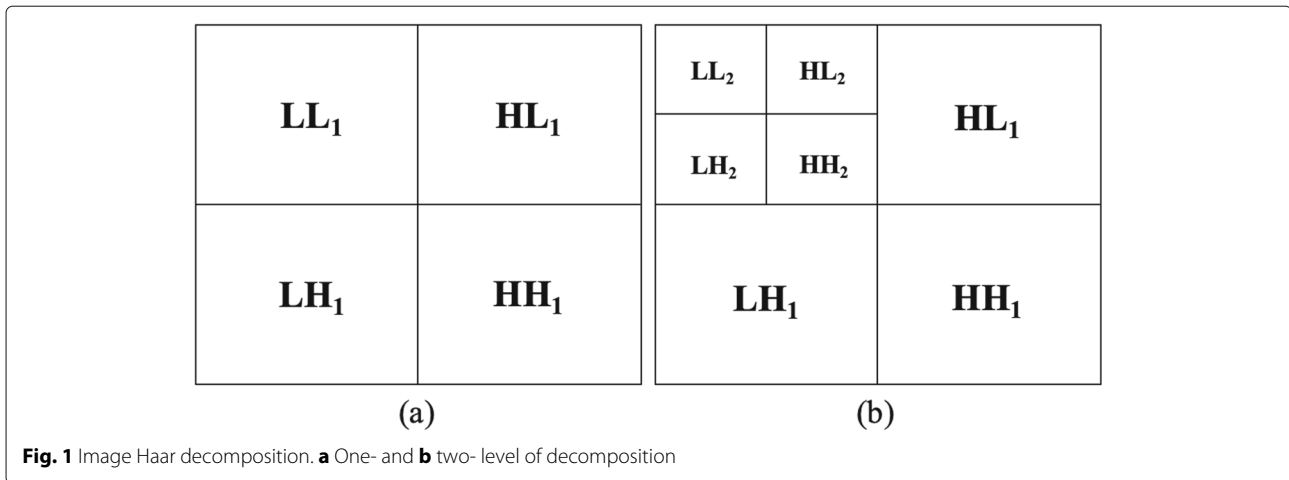
be identified in the ROIs. Specifically, they were obtained by using SURF and MinEigenAlg.

SURF method [29] is an interest local point detector and descriptor that relies on integral images for image convolutions. It consists of three main parts: interest point detection, local neighborhood description and matching. For the detection of interest points, first it uses square-shaped filters as an approximation of Gaussian smoothing, and then evaluates the sum of the original image within a rectangle using the integral image and requiring evaluations at the four corners of that rectangle. Then, SURF algorithm employs an integer approximation of the determinant of Hessian matrix of an image I, defined at the point $p(x, y)$ and scale σ as follows:

$$\mathcal{H}(p, \sigma) = \begin{pmatrix} L_{xx}(p, \sigma) & L_{xy}(p, \sigma) \\ L_{yx}(p, \sigma) & L_{yy}(p, \sigma) \end{pmatrix}, \qquad (2)$$

where $L_{xx}(p, \sigma)$ is the convolution of the second-order derivative of Gaussian with the image $I(x, y)$ at the point x, and similarly for L_{xy}, L_{yx} and L_{yy}; the scale σ represents the layers obtained by filtering the image with gradually bigger mask (9x9, 15x15, 27x27, etc.). In this way, a pyramidal scale space is built: rather than serial downsampling (Fig. 2a), each successive level of the pyramid is built by upscaling the image in parallel (Fig. 2b) [29]. In order to find points of interest in the image and over scales, local change around the point is measured and detected points are highlighted. Finally, the maxima of the Hessian matrix determinant are interpolated in scale and image space. The descriptor of local neighborhood is made by means a description of the intensity distribution of the pixels within the neighborhood of the point of interest in order to provide a solid description of an image feature. The SURF descriptor fixes a reproducible orientation by using information from a circular region around the points of interest, and adds the Haar wavelet responses. Then, the interest region is split into smaller sub-regions,

and for each of them, the Haar wavelet responses are extracted. Finally, the responses are weighted with a Gaussian in order to offer more robustness for deformations, noise and translation. For the last part of the algorithm, matching pairs can be found by comparing the descriptors obtained from different images.
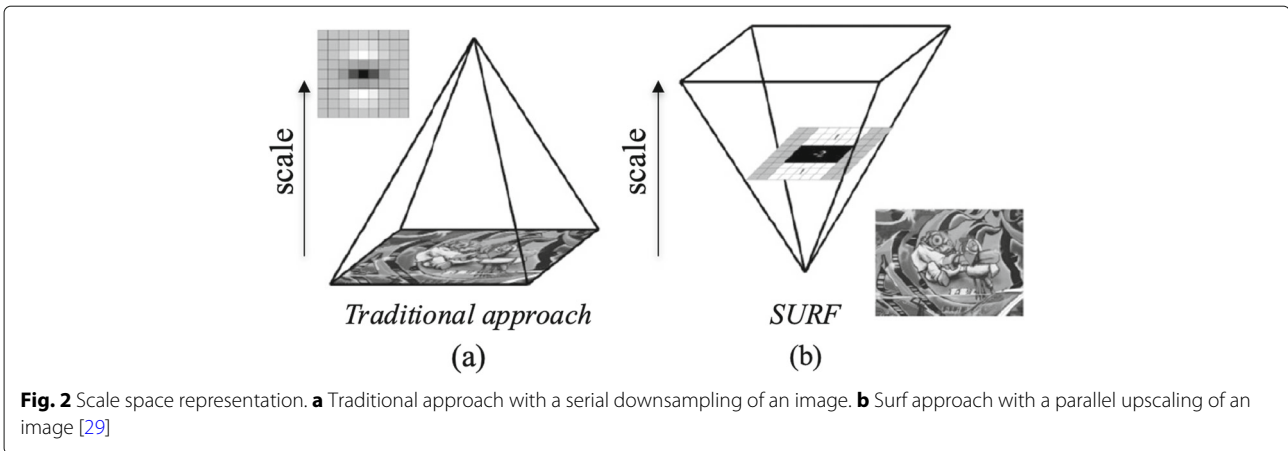
MinEigenAlg uses the Shi-Tomasi detector to identify the interest corners of an image. It is based on the Harris corner detector [30] with a modification in the score calculation. A corner can be defined as a point where two dominant and different edge directions meet in a local neighborhood of the point, differently from an edge with no change along the edge direction and from a flat region where no change are in all directions. The corner selection criterion of the Harris corner detector is that a score is calculated for each pixel with respect to all the directions (horizontal, vertical and on the two diagonals) by using the two eigenvalues (λ_1 and λ_2) of a symmetric matrix, known as Harris matrix. According to this corner detector, the Harris matrix provides two "large" eigenvalues for an interest corner. Then, a function taking into account the determinant and trace of the Harris matrix gives the following score:

$$R_H = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2, \qquad (3)$$

where k is an empirical constant ($k = 0.04 - 0.06$). The Shi-Tomasi corner detector is different from the Harris corner detector in the score computation:

$$R_{ST} = min(\lambda_1, \lambda_2), \qquad (4)$$

where the score depends only on two eigenvalues and not on k constant. Considering this minimum value between two eigenvalues as score, when it is greater than a minimum value, the point can be marked as a corner (green hatched area in Fig. 3).

Fanizzi *et al. BMC Bioinformatics* 2020, **21**(Suppl 2):91

Page 5 of 11

Fig. 2 Scale space representation. **a** Traditional approach with a serial downsampling of an image. **b** Surf approach with a parallel upscaling of an image [29]

In this preliminary approach, only the number of interest points (IP) and corners (IC) has been taken into account by applying the two algorithms above described.
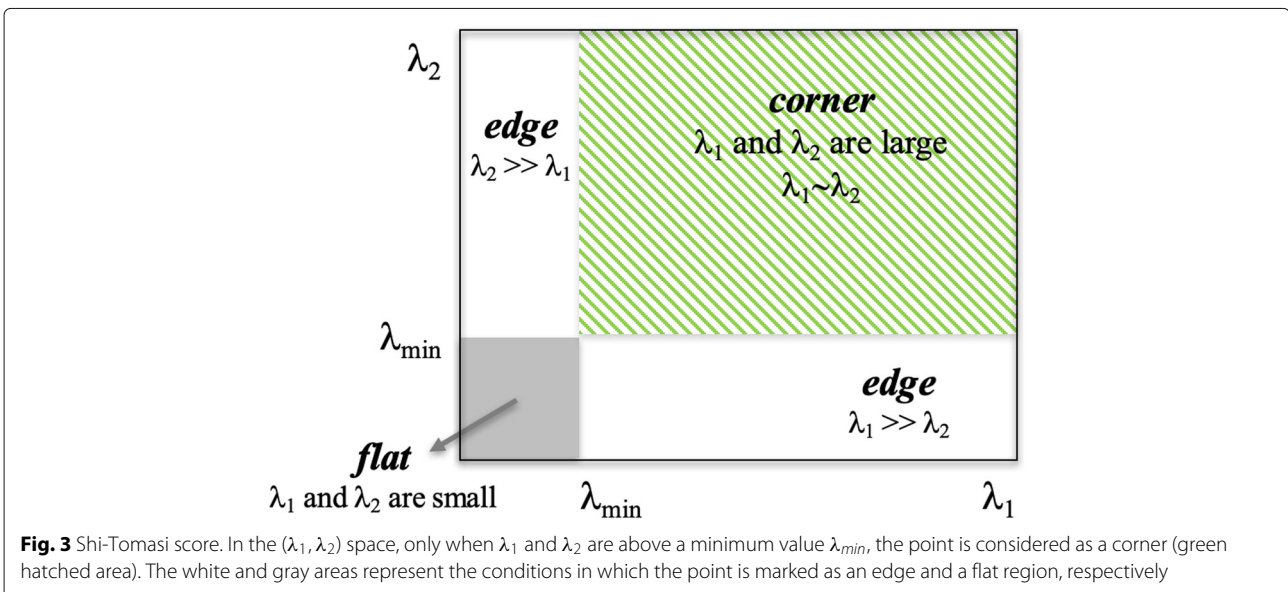
An example of the feature set extraction is shown in Fig. 4: the SF, IP and IC sets are obtained starting from an original ROI including microcalcifications by applying Haar wavelet transform, SURF and MinEigen algorithms, respectively. The extraction of these feature sets was performed by using the corresponding default functions implemented in the MATLAB R2017a (Mathworks, Inc., Natick, MA, USA) software.
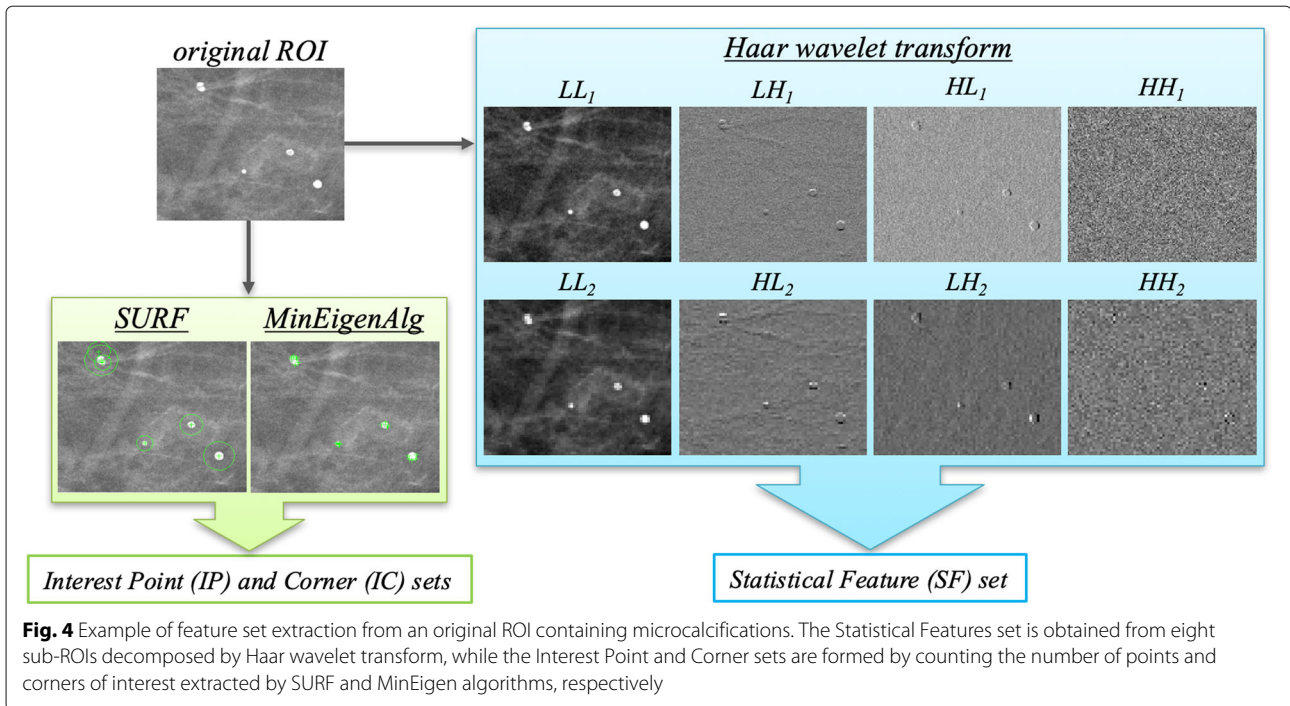
Classification model

The general structure of the classification model proposed is showed in Fig. 5. The method is developed in three phases: (i) for each ROI a set of features are extracted by using the methods above described; (ii) a features subset is selected on training set; (iii) finally, a Random Forest (RF)

binary classifier [33] is trained to discriminate ROIs using the selected features sub-set.

In this work, we evaluate two different kinds of feature selection techniques, such as filter and embedded methods [31, 32]. The filter methods include algorithms that evaluate the capacity of the individual features to predict the expected result. Usually, univariate parametric and non-parametric statistical tests are used to evaluate the significance of the different distributional form of features in sub-samples (classes). Since the feature selection procedure is independent from the machine learning algorithm used, it is possible that the selected features in the first phase will result in a subset that may not work very well downstream of the learning algorithm. Although the learning pipeline is faster, the contribution to discrimination problem generated by the combination of features is lost. In our work, in order to filter out features that have little chance to be useful in data analysis, we used the

Fig. 3 Shi-Tomasi score. In the (λ_1, λ_2) space, only when λ_1 and λ_2 are above a minimum value λ_{min}, the point is considered as a corner (green hatched area). The white and gray areas represent the conditions in which the point is marked as an edge and a flat region, respectively

Fanizzi *et al. BMC Bioinformatics* 2020, **21**(Suppl 2):91

Page 6 of 11

Fig. 4 Example of feature set extraction from an original ROI containing microcalcifications. The Statistical Features set is obtained from eight sub-ROIs decomposed by Haar wavelet transform, while the Interest Point and Corner sets are formed by counting the number of points and corners of interest extracted by SURF and MinEigen algorithms, respectively
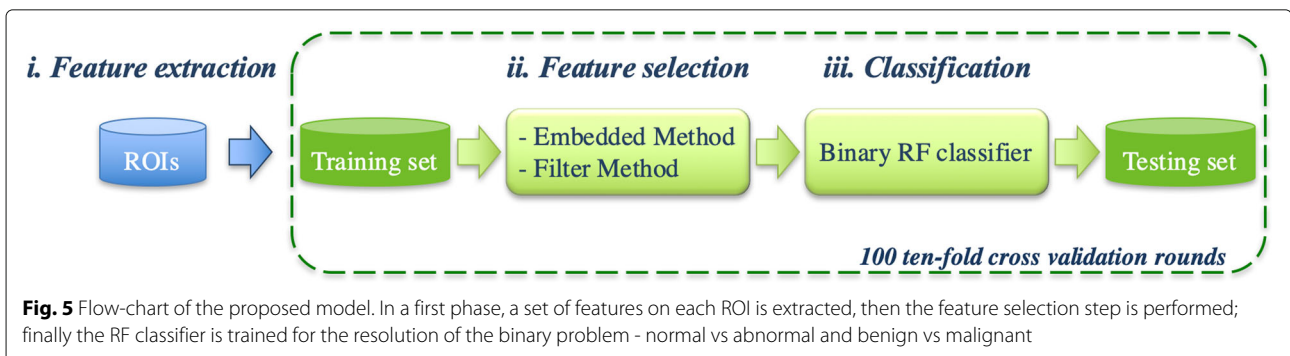
non-parametric Wilcoxon-Mann-Whitney test [36] to verify whether the medians of distributions of the two classes of the binary problem are equal.

Embedded methods allow optimization between the interaction of the selected features and the classification algorithm used. In fact, the selection criterion is grafted into the chosen machine learning algorithm: an analysis of feature importance with respect to its expected result is intrinsically elaborated, therefore these methods are essentially fulfilling the goal, i.e. optimizing the classifier performance. However, they are computationally more expensive than the repeated learning steps and cross-validation. In this work, the feature relevance problem-driven is calculated by the same RF [33] that is often used for feature selection task. Indeed, the tree-based strategy used by random forests naturally ranks by how well they improve the purity of the node: nodes with the greatest decrease in impurity happen at the start of the trees, while nodes with the least decrease in impurity occur at the end of trees. Thus, by pruning trees below a particular node, we can create a sub-set of the most important features.

At each step of cross-validation, a feature ranking is calculated with respect to their predictive power assessed with the two different approaches (filter and embedded method) on training set. Then, a binary classification model is trained by selecting iteratively an increasing number of features sorted by their discriminating power.

By using the filter method, the features are sorted in descending order by the p-value non-parametric test calculated on the training dataset; on the contrary, with the embedded method the features are sorted in increasing order by the relevance calculated on the training set. The subset selected at each step of the cross-validation is in turn used to train a RF classifier: a standard configuration

Fig. 5 Flow-chart of the proposed model. In a first phase, a set of features on each ROI is extracted, then the feature selection step is performed; finally the RF classifier is trained for the resolution of the binary problem - normal vs abnormal and benign vs malignant

Fanizzi *et al. BMC Bioinformatics* 2020, **21**(Suppl 2):91

Page 7 of 11

was adopted with 100 trees and 20 features (as described in [33]) randomly selected at each split.

The proposed model was evaluated on two binary discrimination problems, i.e. normal vs abnormal tissue and benign vs malignant lesions. The performance of the prediction models was evaluated in terms of Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve, accuracy, sensitivity and specificity on 100 ten-fold cross-validation rounds.

Results

In a previous work [37], we proposed a CAD for characterizing and discriminating ROIs that did not provide a real feature selection process. It was trained on statistical features calculated on the multiscale decomposition of the image based on the Haar wavelet transform, and on interest points and corners detected by using two known algorithms, SURF and MinEigenAlg. In particular, a state-of-the-art machine learning classifier, such as a Random Forest classifier, was trained to solve a binary discrimination problem. The performance of the proposed model was evaluated in cross validation on 260 ROIs (130 normal, 75 benign and 55 malignant ROIs); the experimental outcomes showed that the developed model was high performing both for the normal vs abnormal classification problem, with a median AUC value of 98.46% and an accuracy of 95.83%, and for benign vs malignant one, with a median AUC value of 94.19% and an accuracy of 88.19%.

Starting from these encouraging results, here we would examine the contribution of different methods on the feature selection process, and consequently on the learning performance, as well as analyze the discrimination power of some features according to their assigned relevance on a larger and different set of ROIs.

Specifically, as previously reported, two different approaches for the feature selection task were evaluated, namely filter and embedded methods and the performance of the classification model in cross-validation was measured. In particular, as described in the "Methods" section, at each step of cross-validation, a features ranking is calculated with the two approaches on training set and a binary RF classifier is trained by selecting iteratively an increasing number of features.

Figure 6 shows the performance results for solving the normal/abnormal (a) and benign/malignant (b) problems on ROIs extracted from BCDR dataset. The mean accuracy of classifier models (%) has been calculated on 100 rounds of 10-fold cross validation. The experiment results show that the number of features needed to optimize the performance for discriminating normal and abnormal ROIs are 2 in case of the embedded method is used, whereas they become 6 with the exploitation of the filter method. In particular, the embedded method always selects only the interest points and corners with SURF

and MinEigen algorithms, respectively; instead, among the first 6 most significant features selected with the filter method there is also the kurtosis measurement calculated on different Haar decompositions (Table 1).

The discrimination problem of the benign and malignant ROIs requires more features to solve the classification problem; specifically, 10 features with the embedded method and 26 with the filter method. Among the first 10 most significant features selected with the first method, there are yet interest points and corners provided by SURF and MinEigen algorithms, respectively, but also relative smoothness, variance and entropy measurements calculated on different Haar decompositions. With reference to the features selected by the filter method, skewness and kurtosis measurements result also significant.

Table 2 shows the best classification performances calculated on 100 rounds of 10-fold cross-validation. The binary models trained to discriminate normal/ abnormal ROIs by using the two different feature selection approaches are highly performing; however, the performance of the model trained on features selected by the embedded method is slightly higher (p-value Wilcoxon-Mann Whitney test $\leq$ 0.01) of those obtained with the filter approach, except in the identification of malignant ROIs (sensitivity); in particular, with only 2 features, it reaches a median AUC value of 98.16%, an accuracy of 97.31%, a sensitivity of 94.62%, and a specificity of 100%.

Indeed, the models trained to discriminate benign/malignant ROIs with two different feature selection methods are yet highly performing and significantly comparable (p-value Wilcoxon-Mann Whitney test $\leq$ 0.01), however the embedded method selected fewer features: with 10 features selected by the embedded feature selection method, the classification performances reaches a median AUC value of 92.08%, an accuracy of 88.46%, a sensitivity of 89.09%, and a specificity of 88.00%.

Discussion

In this work we have developed a binary classification model of ROIs containing microcalcification clusters. Firstly, for each ROI obtained from images of BCDR database, we have extracted textural features on a multiscale decomposition based on the Haar wavelet transform, and also detected interest points and corners by using two known algorithms, SURF and MinEigenAlg. In particular, we have evaluated the classification performance of a RF classifier for increasing values of features selected on training set of the cross-validation selected by two different approaches, that are embedded and filter methods.

The experimental results on the dataset considered have highlighted, regardless of the method of feature selection used, the normal/abnormal problem can be effectively solved with a number of features decidedly contained (no more than 6), achieving high performance comparable
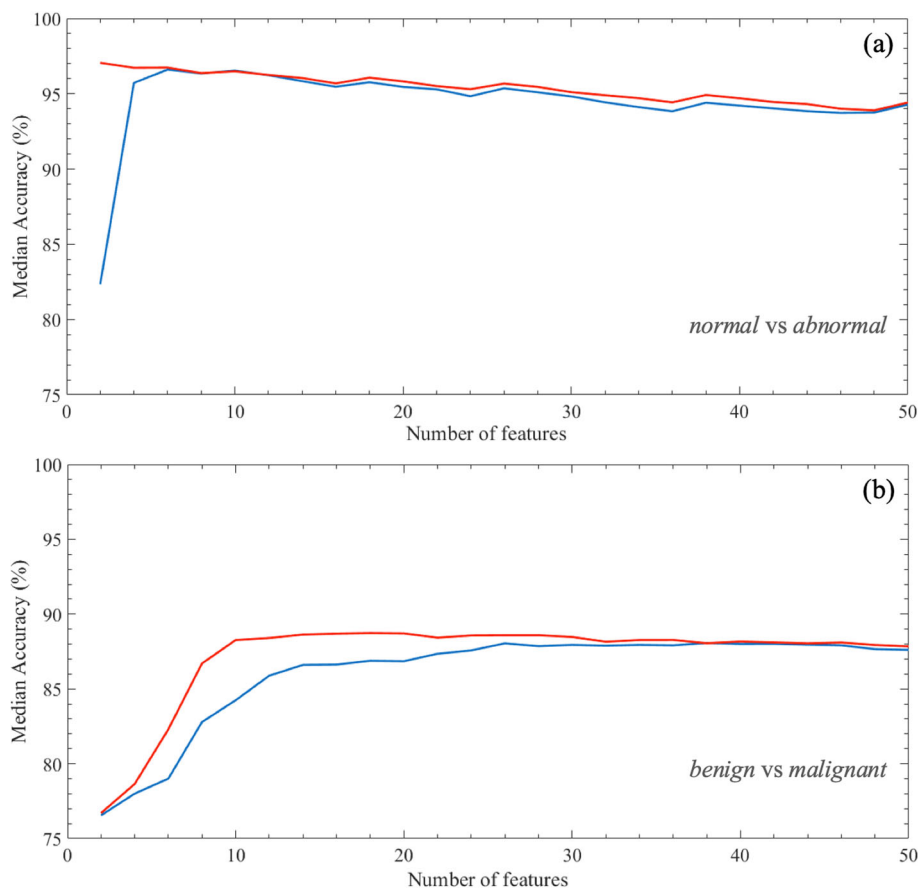
Fig. 6 Median accuracy (%) with the growth of the number of features. The median value is calculated on 100 rounds of 10-fold cross validation for increasing values of the number of features used to train the proposed model to classify ROIs into **a** normal/abnormal **b** and benign/malignant. Two different feature selection approaches are used, that are embedded (red line) and filter (blue line) methods

to the state-of-the-art. Specifically, the best performance is obtained with the embedded feature selection method using only two features, that are interest points and corners provided by SURF and MinEigenAlg, respectively.

For what concerns the much more complex problem of benign/malignant classification, which represents the main focus of the works proposed in the literature about the characterization of microcalcification clusters, the best results of the proposed method (median AUC value of 92.08%, accuracy of 88.46%) were obtained using 10 features selected by the embedded method: the joint contribution of these has allowed the achievement of comparable performance with respect to the best result obtained by independently selecting the 26 features with the highest discriminated power.

Experimental results showed that interest points and corners, relative smoothness, variance and entropy measurements calculated on different Haar decompositions seem to have significant information content for discriminating benign/malignant ROIs.

Results of this study are quite promising. Table 3 shows the performance of state-of-the-art models for the classification into benign and malignant microcalcifications. For this comparison, works on classification of microcalcification clusters mainly using textural features [38] or a combination of these with other types of features [21], but also topological ones [17, 25] and statistical features [19] were taken into account. Moreover, different machine learning approaches and databases have been used. Therefore, the comparison is purely qualitative. However, the classification performances obtained by our approach are more performing with respect to these works and do not require a manual segmentation of the lesion by radiologist but only the identification of a suspicious area. Moreover, the proposed approach is able to characterize the lesions by exploiting a reduced set of features.

In the next stage of our studies, we will evaluate the proposed model on different databases and also evaluate other features to improve the classification

Fanizzi *et al. BMC Bioinformatics* 2020, **21**(Suppl 2):91

Page 9 of 11

Table 1 Significant features on BCDR database

Normal/Abnormal				Benign/Malignant			
Embedded M. (k ≤ 2)	freq (%)	Filter M. (k ≤ 6)	freq (%)	Embedded M. (k ≤ 10)	freq (%)	Filter M. (k ≤ 26)	freq (%)
# Interest Points	100	# Interest Points	100	Variance_LL2	100	Variance_LL1	100
# Interest Corners	100	Kurtosis_HL2	99.80	# Interest Corners	100	Skewness_LL1	100
		# Interest Corners	99.10	Variance_LL1	99.90	Entropy_LL1	100
		Kurtosis_HL1	97.80	RelSmoothness_LL2	99.90	RelSmoothness_LL1	100
		Kurtosis_LH1	76.40	RelSmoothness_LL1	99.60	Entropy_HL1	100
		Kurtosis_LH2	61.90	# Interest Points	91.30	Entropy_HH1	100
		Variance_LH2	24.80	Variance_HH1	77.70	Kurtosis_HH1	100
		RelSmoothness_LH2	21.90	RelSmoothness_HH1	77.40	Variance_LL2	100
				Entropy_HH1	58.90	Skewness_LL2	100
				Entropy_HL1	44.80	Entropy_LL2	100
				Mean_HH1	41.20	RelSmoothness_LL2	100
						Kurtosis_LH2	100
						Kurtosis_HL2	100
						Kurtosis_HH2	100
						# Interest Points	100
						# Interest Corners	100
						Entropy_LH1	99.20
						Entropy_LH2	98.60
						Entropy_HH2	97.80
						Kurtosis_HL1	97.10
						RelSmoothness_HH1	96.10
						Variance_HH1	88.80
						Skewness_HL2	76.30
						Mean_LL1	59.00

The features whose occurrence in the first k positions of the rankings defined by the filter and embedded methods is significantly different from the case (p-value null model test ≤ 0.05) are reported. k is the number of features that maximizes the accuracy of normal vs abnormal and benign vs malignant classification problems

Table 2 Best classification performance on BCDR database

		Normal/Abnormal	Benign/Malignant
	AUC	98.16 (97.87 − 98.48) ∗∗	92.08 (91.61 − 92.58)
Embedded Method	Accuracy	97.31 (96.92 − 97.31) ∗∗	88.46 (87.69 − 89.23)
	Sensitivity	94.62 (93.85 − 94.62)	89.09 (87.27 − 90.91)
	Specificity	100 (100 − 100) ∗∗	88.00 (86.67 − 89.33)
	AUC	98.67 (98.57 − 98.76)	92.13 (91.66 − 92.78)
Filter Method	Accuracy	96.92 (96.54 − 96.92)	87.69 (86.92 − 89.23)
	Sensitivity	93.85 (93.85 − 94.62)	89.09 (87.27 − 90.91)
	Specificity	99.23 (99.23 − 100)	87.33 (85.33 − 89.33)

The classification performance calculated in correspondence with the best result highlighted in the 100 rounds of 10-fold cross-validation for increasing the number of selected features, are summarized. We tested the significance of the diversity of performance measures obtained with the two different feature selection techniques on the same classification problem. Statistical significance is measured with the Wilcoxon-Mann-Whitney test: ∗∗ p-value < 0.01 (Bonferroni correction)

Fanizzi *et al. BMC Bioinformatics* 2020, **21**(Suppl 2):91

Page 10 of 11

Table 3 Benign vs Malignant microcalcifications: accuracy (Acc) and Area Under the Curve (AUC) performances

Method	No. ROIs	Feature type	Classifier	Acc (%)	AUC (%)
Chen et al. (2015) [17]	300	topological features	kNN	85	91
Ren et al. (2012) [19]	295	statistical features	kNN	82	86
Khehra et al. (2013) [21]	380	statistical, shape and textural features	LS-SVM	89	89
Strange et al. (2014) [25]	300	mereotopological features	Barcodes	80	82
Hu et al. (2017) [38]	150	textural features	ELM	-	92
Proposed approach	**260**	**textural features and #interest points/corners**	**RF**	**88**	**92**

performances of benign and malignant microcalcification clusters.

Conclusion

The diagnosis of microcalcifications is usually based on radiologists expertise sometimes resulting in an inaccurate lesion detection with unnecessary biopsies and subsequent surgery.

Several methods have been developed for the task of microcalcification diagnosis, in some cases well performing. Nevertheless, the automatic and accurate classification of microcalcification clusters, especially in differentiating the benign from the malignant ones, remains still complicated due to their nature. In this paper, we propose the use of texture analysis methods combined with machine learning techniques in order to select the optimal subset for characterizing breast regions. In particular, we trained a binary RF classifier on an increasing number of features sorted by their statistical significance in the set of data; these features were calculated using two different feature selection approaches, such as embedded and filter methods.

Both feature selection techniques revealed highly performing, nevertheless the best accuracy result obtained with embedded method is more parsimonious than the filter one: it needs only 2 features to discriminate ROIs into normal/abnormal and 10 into benign/malignant.

The measurements provided by SURF and MinEigen algorithms seems to provide a strong informative content useful for the characterization of microcalcification clusters. In this work we have limited ourselves to considering only the number of interest points and corners and not also the features associated with them. In order to improve the performances, specially in the classification of benign and malignant microcalcifications, it would be interesting to deepen the analysis of these features. Future work will concern also the simultaneous exploitation of the information coming from the CC and MLO views of the same breast lesion [39] and the evaluation of more complex features selection methods by wrapped approach.

Abbreviations
ANN: Artificial neural networks; AUC: Area under the curve; BCDR: Breast cancer digital repository; CAD: Computer-aided detection/diagnosis; CC: CranioCaudal; DM: Digital mammography; FM: Film mammography; HH: High-high; HL: High-low; IC: Interest corner; IP: Interest point; kNN: k-nearest neighbours; LH: Low-high; LL: Low-low; MC: MicroCalcification; MinEigen(Alg): Minimum eigenvalue (algorithm); MLO: MedioLateral oblique; RF: Random forest; ROC: Receiver operating characteristics; ROI: Region of interest; SF: Statistical feature; SURF: Speed up robust feature; SVM: Support vector machine

Authors' contributions
Conceptualization: AnF, LL, TMAB, ST and DLF; Data curation: DLF, AF, MM, AnF and LL; Formal analysis: AnF, LL, TMAB and ST; Funding acquisition: DLF; Methodology: AnF, LL, and TMAB; Software: AnF and LL; Supervision: ST and DLF; Validation: AnF; Writing - original draft: AnF, LL, TMAB, ST and DLF; Writing - review editing: RB, UB, RD, VD, AF, RM, MM, OP and PT. All authors read and approved the final manuscript.

Author details
[1] I.R.C.C.S. Istituto Tumori "Giovanni Paolo II", viale O. Flacco 65, Bari, Italy. [2] Dip. Interateneo di Fisica "M. Merlin", Università degli Studi di Bari "A. Moro", via G. Amendola 173, Bari, Italy. [3] INFN - Istituto Nazionale di Fisica Nucleare, sezione di Bari, via G. Amendola 173, Bari, Italy. [4] Dip. di Scienze Fisiche, della Terra e dell'Ambiente, Università degli Studi di Siena, strada Laterina 2, Siena, Italy. [5] Dip. di Diagnostica delle Immagini, Ospedale Universitario di Siena, viale Bracci 16, Siena, Italy. [6] Dip. Interdisciplinare di Medicina, Università degli Studi di Bari "A. Moro", piazza G. Cesare 11, Bari, Italy.

Fanizzi *et al. BMC Bioinformatics* 2020, **21**(Suppl 2):91

Page 11 of 11

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
2. Howell A. The emerging breast cancer epidemic: early diagnosis and treatment. Breast Cancer Res. 2010;12(4):10.
3. Cronin KA, Lake AJ, Scott S, Sherman RL, Noone A-M, Howlader N, Henley SJ, Anderson RN, Firth AU, Ma J, et al. Annual report to the nation on the status of cancer, part i: National cancer statistics. Cancer. 2018;124(13):2785–800.
4. Doi K. Computer-aided diagnosis in medical imaging: achievements and challenges. In: World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany. Springer; 2009. p. 96. https://doi.org/10.1007/978-3-642-03904-1_26.
5. Jalalian A, Mashohor SB, Mahmud HR, Saripan MIB, Ramli ARB, Karasfi B. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. Clin Imaging. 2013;37(3):420–6.
6. Iacomi M, Cascio D, Fauci F, Raso G. Mammographic images segmentation based on chaotic map clustering algorithm. BMC Med Imaging. 2014;14(1):12.
7. Boyd NF, Martin LJ, Bronskill M, Yaffe MJ, Duric N, Minkin S. Breast tissue composition and susceptibility to breast cancer. J Natl Cancer Inst. 2010;102(16):1224–37.
8. Checka CM, Chun JE, Schnabel FR, Lee J, Toth H. The relationship of mammographic density and age: implications for breast cancer screening. Am J Roentgenol. 2012;198(3):292–5.
9. Azavedo E, Zackrisson S, Mejàre I, Arnlind MH. Is single reading with computer-aided detection (cad) as good as double reading in mammography screening? a systematic review. BMC Med Imaging. 2012;12(1):22.
10. Elter M, Horsch A. Cadx of mammographic masses and clustered microcalcifications: a review. Med Phys. 2009;36(6Part1):2052–68.
11. Shiraishi J, Li Q, Appelbaum D, Doi K. Computer-aided diagnosis and artificial intelligence in clinical imaging. In: Seminars in Nuclear Medicine. Elsevier; 2011. p. 449–62. https://doi.org/10.1053/j.semnuclmed.2011.06.004.
12. Guzmán-Cabrera R, Guzmán-Sepúlveda J, Torres-Cisneros M, May-Arrioja D, Ruiz-Pinales J, Ibarra-Manzano O, Aviña-Cervantes G, Parada AG. Digital image processing technique for breast cancer detection. Int J Thermophys. 2013;34(8-9):1519–31.
13. Jian W, Sun X, Luo S. Computer-aided diagnosis of breast microcalcifications based on dual-tree complex wavelet transform. Biomed Eng Online. 2012;11(1):96.
14. Phadke AC, Rege PP. Detection and classification of microcalcifications using discrete wavelet transform. Int J Emerg Trends Technol Comput Sci. 2013;2(4):130–4.
15. Khehra BS, Pharwaha APS. Classification of clustered microcalcifications using mlffbp-ann and svm. Egypt Informa J. 2016;17(1):11–20.
16. Boulehmi H, Mahersia H, Hamrouni K. A new cad system for breast microcalcifications diagnosis. Int J Adv Comput Sci Appl. 2016;7(4):133–43.
17. Chen Z, Strange H, Oliver A, Denton ER, Boggis C, Zwiggelaar R. Topological modeling and classification of mammographic microcalcification clusters. IEEE Trans Biomed Eng. 2015;62(4):1203–14.
18. Cascio D, Fauci F, Iacomi M, Raso G, Magro R, Castrogiovanni D, Filosto G, Ienzi R, Vasile MS. Computer-aided diagnosis in digital mammography: comparison of two commercial systems. Imaging Med. 2014;6(1):13.
19. Ren J. Ann vs. svm: Which one performs better in classification of mccs in mammogram imaging. Knowl Based Syst. 2012;26:144–53.
20. Pawar MM, Talbar SN. Genetic fuzzy system (gfs) based wavelet co-occurrence feature selection in mammogram classification for breast cancer diagnosis. Perspect Sci. 2016;8:247–50.
21. Khehra BS, Pharwaha APS. Least-squares support vector machine for characterization of clusters of microcalcifications. World Acad Sci Eng Technol Int J Comput Inf Sci Eng. 2013;7(12):932–41.
22. Beura S, Majhi B, Dash R. Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. Neurocomputing. 2015;154:1–14.
23. Mohamed H, Mabrouk MS, Sharawy A. Computer aided detection system for micro calcifications in digital mammograms. Comput Methods Prog Biomed. 2014;116(3):226–35.
24. Malar E, Kandaswamy A, Chakravarthy D, Dharan AG. A novel approach for detection and classification of mammographic microcalcifications using wavelet analysis and extreme learning machine. Comput Biol Med. 2012;42(9):898–905.
25. Strange H, Chen Z, Denton ER, Zwiggelaar R. Modelling mammographic microcalcification clusters using persistent mereotopology. Pattern Recogn Lett. 2014;47:157–63.
26. Fanizzi A, Basile T, Losurdo L, Amoroso N, Bellotti R, Bottigli U, Dentamaro R, Didonna V, Fausto A, Massafra R, et al. Hough transform for clustered microcalcifications detection in full-field digital mammograms. In: Applications of Digital Image Processing XL, vol. 10396. International Society for Optics and Photonics; 2017. p. 1039616. https://doi.org/10.1117/12.2273814.
27. Gonzalez RC, Woods RE. Digital Image Processing, 3rd end. Upper Saddle River: Prentice-Hall; 2007.
28. Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell. 1989;11(7):674–93.
29. Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-up robust features (surf). Comput Vis Image Underst. 2008;110(3):346–59.
30. Shi J, et al. Good features to track. In: Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference On. IEEE; 1994. p. 593–600. https://doi.org/10.1109/cvpr.1994.323794.
31. Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. bioinformatics. 2007;23(19):2507–17.
32. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003;3(Mar):1157–82.
33. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
34. Ramos-Pollán R, Guevara-López MA, Suárez-Ortega C, Díaz-Herrero G, Franco-Valiente JM, Rubio-del-Solar M, González-de-Posada N, Vaz MAP, Loureiro J, Ramos I. Discovering mammography-based machine learning classifiers for breast cancer diagnosis. J Med Syst. 2012;36(4):2259–69.
35. D'Orsi C, Sickles E, Mendelson E, Morris E. 2013 ACR BI-RADS Atlas: Breast Imaging Reporting and Data System. Reston: American College of Radiology; 2014.
36. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat. 194750–60. https://doi.org/10.1214/aoms/1177730491.
37. Losurdo L, Fanizzi A, Basile TM, Bellotti R, Bottigli U, Dentamaro R, Didonna V, Fausto A, Massafra R, Monaco A, et al. A combined approach of multiscale texture analysis and interest point/corner detectors for microcalcifications diagnosis. In: International Conference on Bioinformatics and Biomedical Engineering. Springer; 2018. p. 302–13. https://doi.org/10.1007/978-3-319-78723-7_26.
38. Hu K, Yang W, Gao X. Microcalcification diagnosis in digital mammography using extreme learning machine based on hidden markov tree model of dual-tree complex wavelet transform. Expert Syst Appl. 2017;86:135–44.
39. Batchelder KA, Tanenbaum AB, Albert S, Guimond L, Kestener P, Arneodo A, Khalil A. 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. 2014;9(9):107580.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.