**World Scientific**
www.worldscientific.com

# TESTING THE PERFORMANCES OF DIFFERENT IMAGE REPRESENTATIONS FOR MASS CLASSIFICATION IN DIGITAL MAMMOGRAMS

E. ANGELINI, R. CAMPANINI, E. IAMPIERI, N. LANCONELLI and M. MASOTTI*

*Department of Physics, University of Bologna, and INFN Bologna*
*Viale Berti–Pichat 6/2, 40127 Bologna, Italy*
*\*matteo.masotti@bo.infn.it*

M. ROFFILLI

*Department of Computer Science, University of Bologna*
*Mura Anteo Zamboni 7, 40127, Bologna, Italy*

The classification of tumoral masses and normal breast tissue is targeted. A mass detection algorithm which does not refer explicitly to shape, border, size, contrast or texture of mammographic suspicious regions is evaluated. In the present approach, classification features are embodied by the image representation used to encode suspicious regions. Classification is performed by means of a support vector machine (SVM) classifier. To investigate whether improvements can be achieved with respect to a previously proposed overcomplete wavelet image representation, a pixel and a discrete wavelet image representations are developed and tested. Evaluation is performed by extracting 6000 suspicious regions from the digital database for screening mammography (DDSM) collected by the University of South Florida (USF). More specifically, 1000 regions representing biopsy-proven tumoral masses (either benign or malignant) and 5000 regions representing normal breast tissue are extracted. Results demonstrate very high performance levels. The area $A_z$ under the receiver operating characteristic (ROC) curve reaches values of $0.973 \pm 0.002$, $0.948 \pm 0.004$ and $0.956 \pm 0.003$ for the pixel, discrete wavelet and overcomplete wavelet image representations, respectively. In particular, the improvement in the $A_z$ value with the pixel image representation is statistically significant compared to that obtained with the discrete wavelet and overcomplete wavelet image representations (two-tailed $p$-value $< 0.0001$). Additionally, 90% true positive fraction (TPF) values are achieved with false positive fraction (FPF) values of 6%, 11% and 7%, respectively.

*Keywords*: Computer-aided detection; mammography; support vector machine; image processing; wavelets.

PACS Nos.: 87.57.Nk, 87.57.Ra, 87.59.Ek.

## 1. Introduction

Breast cancer remains one of the leading causes of death among women from all over the world.[1] An early detection of this disease is fundamental in order to increase the probability for the patient to survive. Screening mammography, namely the periodical radiographic examination of asymptomatic women's breast, is undoubtedly the most effective tool to this purpose. However, it is known that between 10%–30% of breast cancers are missed by radiologists during the interpretation of radiographic images.[2,3] This is mostly due to subtle nature of mammographic lesions, poor image quality, eye fatigue or oversight. In order to increase the detection rate of breast cancer, computer-aided detection (CAD) systems have been recently adopted.[4−6] Those automatic systems identify suspicious regions in radiographic images by means of advanced computer vision and artificial intelligence techniques. The basic idea behind those systems is to provide the radiologist with a second opinion, but leaving the final decision strictly to him. With this approach, recent studies demonstrated that CAD systems have a very positive impact on early detection of breast cancer.[7,8]

Masses are lesions commonly associated with the presence of breast cancer. They are thickenings of breast tissue which appear in radiographic images as lesions with size ranging from 3 mm to 30 mm. In order to detect them, the true majority of mass detection algorithms follows a two-step approach. First, suspicious regions are detected on radiographic images by means of traditional image processing techniques, i.e., segmentation, filtering, thresholding, etc. Second, the suspicious regions detected by the first step are classified as mass or non-mass (normal tissue) by means of discriminant techniques, i.e., neural networks, linear discriminant analysis, template matching, etc. The final aim is twofold:

(a) To achieve a high true positive fraction (TPF) value, namely a high fraction of masses correctly classified as masses.
(b) To achieve a low false positive fraction (FPF) value, namely a low fraction of non-masses incorrectly classified as masses.

The overwhelming majority of the mass detection algorithms so far developed addresses the second step by extracting features from the suspicious regions identified during the first step. Typical features refer explicitly to shape, border, size, contrast and texture of those regions. Thus far, for example, texture features in combination with linear discriminant analysis have been used by Wei *et al.*,[9] thus achieving a TPF value of 90% and an FPF value of approximately 35%. On the same dataset, texture features in combination with convolution neural networks yielded a TPF value of 90% and an FPF value of 31%, as described by Sahiner *et al.*[10] Features based on geometry, intensity, gray-levels and gradients of suspicious regions have been investigated in combination with neural networks as well. Template matching techniques in combination with different kinds of extracted features have been discussed by Chang *et al.*, Baydush *et al.* and Tourassi *et al.*[11−13]

For those works, a TPF value of 90% and FPF values of approximately 49%, 14% and 35% have been achieved, respectively.

However, considering the great variability of masses in shape, border, size, contrast and texture, it is often very difficult to get a common set of features effective for every kind of mass. Motivated by this consideration, our group has proposed a novel approach to mass detection[14] in which classification features are simply embodied by the specific image representation used to encode each suspicious region. The coefficients resulting from this encoding are directly classified by means of a support vector machine (SVM) classifier,[15] an artificial intelligence technique based on V. Vapnik's statistical learning theory. The main motivation for using SVM rather than traditional classifiers relies in its ability in handling very high dimensional feature spaces.[16] It should be noted that the number of classification features is generally quite limited in classic approaches. Conversely, for the approach discussed above, the number of classification features is represented by the amount of coefficients resulting from each suspicious region, thus much higher. This ability has been already successfully experimented on CAD issues and biological datasets.[17,18]

The main purpose of this study is thus to investigate whether similar or better performances can be achieved with respect to those achieved by the image representation originally developed, namely an overcomplete wavelet image representation.[19] To this aim, pixel and discrete wavelet[20] image representations will be developed, evaluated and compared to the overcomplete wavelet image representation.

## 2. Materials and Methods

### 2.1. *Image database*

The mass detection approach is evaluated using suspicious regions extracted from the Digital Database for Screening Mammography (DDSM) collected by the University of South Florida (USF).[21] The DDSM database is comprised of digitized radiographic images with associated ground truth which gives location, outline and subtlety ratings for each mass. Radiographic images are digitized with Lumisys scanner at 50 $\mu$m or Howtek scanner at 43.5 $\mu$m, both with 12-bit gray-level resolution. From the DDSM cases, a total of 1000 diagnosed masses are extracted using the provided ground truth annotations. In particular, a square crop centered on the location of each annotated mass is selected. The size is chosen so that the ratio between the crop area and the area of the annotated mass is nearly 1.3. All the crops containing a mass are thus characterized by having about 70% of their area occupied by the annotated mass and the remaining 30% by background. Note that, since SVM deals exclusively with dimensionally homogeneous vectors, all the crops containing a mass have to be resized to an arbitrarily prefixed size of $64 \times 64$ pixels. To this purpose, bilinear resizing has been chosen. On the contrary, for the non-mass class, a total of 5000 square crops are extracted randomly from the DDSM normal cases and selected as $64 \times 64$ pixel regions representing normal tissue. The whole
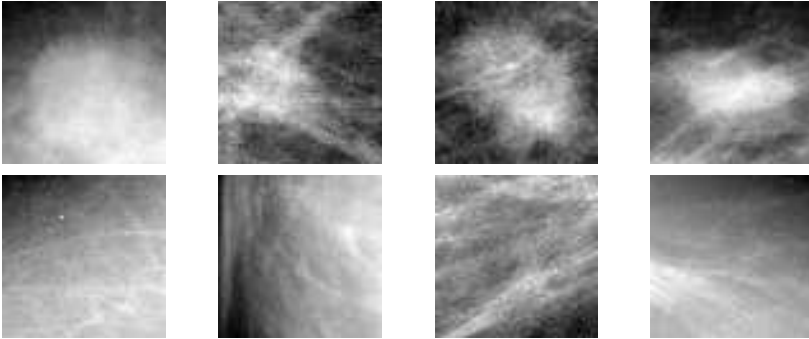
Fig. 1.   The two classes. Mass class (top) versus non-mass class (bottom).

dataset is thus composed of 6000 regions with pixel size $64 \times 64$, namely a total of 1000 crops representing benign and malignant masses plus a total of 5000 crops representing normal tissue, see Fig. 1.

## 2.2.  *Image representations*

### 2.2.1.  *Pixel image representation*

The simplest way to encode an image is by just concatenating all its gray-level intensity values, thus yielding a long vector with as many entries as the number of pixels in the image. This encoding is usually referred to as the pixel image representation of the image under analysis. In the context of image classification, this approach forces the classifier to discriminate the images under exam by simply using the information derived from their gray-level intensity values. For the specific classification problem presented herein, the pixel image representation forces SVM to learn the typical gray-level intensity content of regions representing tumoral masses and that of regions representing normal tissue.

When using the pixel image representation, masses and normal tissue are submitted to the classifier as they appear in radiographic images. Masses appear generally as nucleated objects with quite defined edges, whereas non-masses appear as less defined and very heterogeneous objects. An exception to nucleated masses is represented by spiculated masses, namely objects having a star-shaped boundary and margins with sharp fingers pointing away from the center of the mass. Although these two families of masses are quite different, common characteristics which differentiate them from non-masses are the tendency to have a fairly sharp boundary and to appear brighter than surrounding tissue. Given these considerations, a typical approach would extract few specific features encoding those discriminant characteristics, precisely the sharpness of the boundary and the contrast with the surrounding tissue. Conversely, the approach based on pixel image representation simply submits the gray-level intensity content of each suspicious region

to SVM. No a priori information is thus introduced, since the identification of the most discriminant features (i.e., pixels) is completely demanded to the classifier.

### 2.2.2. *Wavelet image representation*

In the image processing community, the wavelet transform is a well-known technique allowing for multi-resolution and multi-orientation image analysis.[20,22,23] Very broadly, this transform encodes the difference in gray-level intensity content between local regions along different resolutions and orientations of images. A strong response from the wavelet transform of an image indicates the presence of an intensity difference at a specific location, typically an edge or a boundary. On the contrary, a weak response indicates a uniform area. The main motivation for evaluating the application of the wavelet transform herein is precisely that it offers a suitable image representation for highlighting shape and detailed structures of masses at different resolutions and orientations.

The classical wavelet transform, also known as discrete wavelet transform,[20] is an orthogonal transform. By using a cascade of low-pass and high-pass filters together with subsampling operators, it transforms an $N \times N$ pixel image into $N \times N$ multi-resolution and multi-orientation wavelet coefficients. The application of each pair of filters and subsampling operators corresponds to a decomposition level, in other words to a specific resolution of the analysis. Figure 2 illustrates the discrete wavelet decomposition up to level three of a region containing a mass. In particular, for each level $j = 1, 2, 3$, the horizontal component $\mathbf{d}_j^{\mathsf{H}}$, the vertical component $\mathbf{d}_j^{\mathsf{V}}$ and the diagonal component $\mathbf{d}_j^{\mathsf{D}}$ are depicted. Haar wavelet filters have been used.

A redundant version of the discrete wavelet transform, commonly referred to as overcomplete wavelet transform,[19] is used in order to split the information of the image on a higher number of wavelet coefficients, thus achieving finer resolutions. By eliminating the subsampling operators from the first and even decomposition levels, the low-pass and band-pass images produced by the overcomplete wavelet decomposition consist of approximately $N - 1 \times N - 1$ wavelet coefficients at level 1, $N - 2 \times N - 2$ at level 2, $(N - 3)/2 \times (N - 3)/2$ at level 3, etc. In this way, the overcomplete wavelet transform provides a higher-resolution encoding of the image information. Figure 3 illustrates the overcomplete wavelet decomposition up to level three of a region containing a mass. In particular, for each level $j = 1, 2, 3$, the horizontal component $\mathbf{d}_j^{\mathsf{H}}$, the vertical component $\mathbf{d}_j^{\mathsf{V}}$ and the diagonal component $\mathbf{d}_j^{\mathsf{D}}$ are depicted. Haar wavelet filters have been used here as well.

### 2.3. *Classification*

### 2.3.1. *Support Vector Machine*

SVM[15] constructs a binary classifier from a set of $l$ training samples, consisting of labeled patterns $(\mathbf{x}_i, y_i) \in \mathbb{R}^N \times \{\pm 1\}$, $i = 1, \ldots, l$. Taking values $+1$ or $-1$,
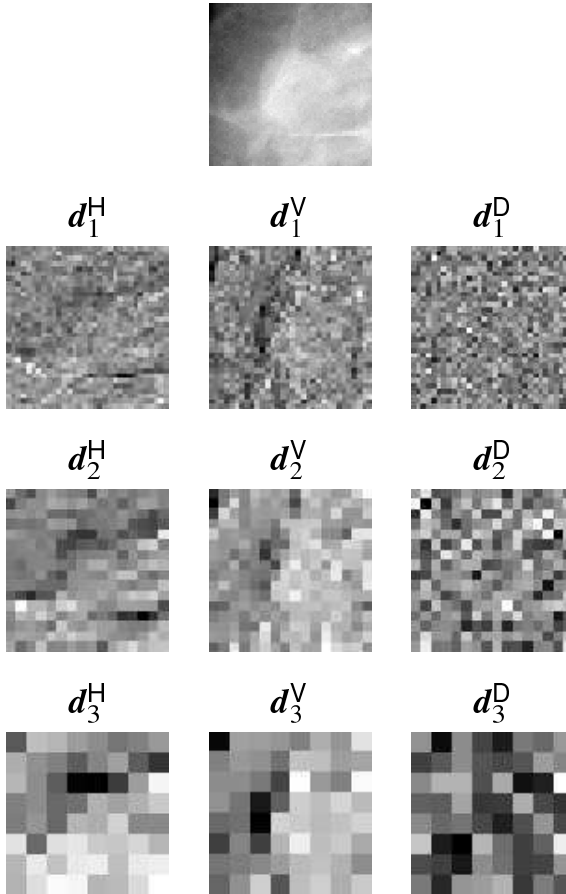
Fig. 2.   Discrete wavelet image representation. Detail components up to level three of a region containing a mass.

the label $y_i$ indicates the class of the pattern $\mathbf{x}_i$ considered. The classifier aims at estimating a decision function $f : \mathbb{R}^N \to \pm 1$ from a given class of functions, such that $f$ will correctly classify unseen test samples $(\mathbf{x}, y)$. A sample is assigned to the class $+1$ if $f(\mathbf{x}) \geq 0$ and to the class $-1$ otherwise. When dealing with the pixel image representation, the feature vector $\mathbf{x}$ contains the gray–level intensity values of each suspicious region. Similarly, when the wavelet image representation is adopted, it contains the wavelet coefficients resulting from the application of the wavelet transform to each of those regions.

SVM selects hyperplanes in order to separate the two classes. Among all the separating hyperplanes, it finds the maximal margin hyperplane (MMH), namely the one that causes the largest separation in the feature space between itself and the borderline training samples of the two classes. To this aim, the decision function
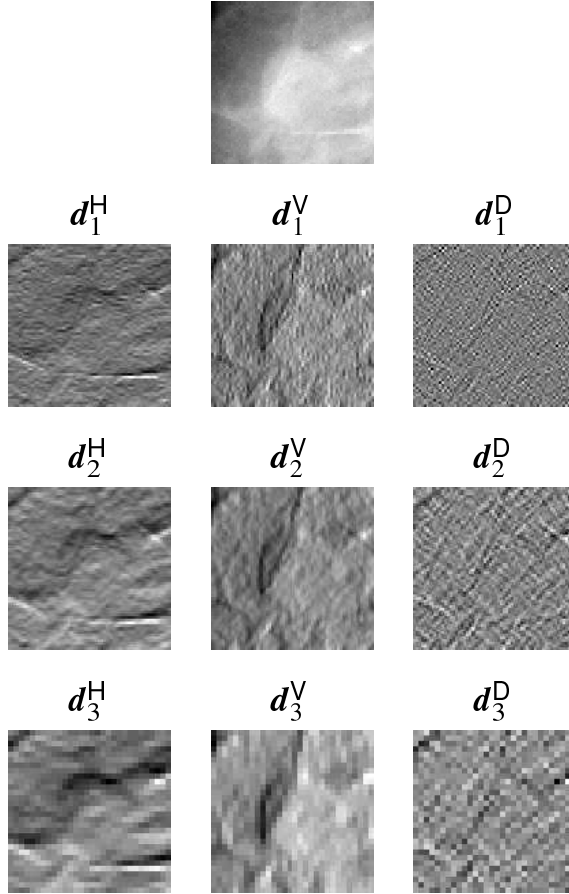
Fig. 3. Overcomplete wavelet image representation. Detail components up to level three of a region containing a mass.

is computed as:

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w} \cdot \mathbf{x} + b) = \text{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i (\mathbf{x} \cdot \mathbf{x}_i) + b\right),\tag{1}$$

where the coefficients $\alpha_i$ and $b$ are calculated by solving the following quadratic programming problem:

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,k=1}^{l} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j,\tag{2}$$

subject to:

$$0 \leq \alpha_i \leq C, \qquad i = 1, \dots, l,\tag{3}$$

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \,. \tag{4}$$

$C$ is a regularization parameter selected by the user. Once the training samples $(\mathbf{x}_i, y_i)$ are provided, the coefficients $\alpha_i$ and $b$ are calculated by solving the quadratic programming problem discussed above. The separating MMH is obtained by substituting in Eq. (1) the coefficients $\alpha_i$ and $b$ found. The classification of a previously unseen sample identified by the feature vector $\mathbf{x}$ is therefore achieved according to the values of $f(\mathbf{x})$ in Eq. (1). Note that, since the most part of the coefficients $\alpha_i$ are generally null, only a small fraction of training samples (called support vectors) contributes to the determination of the separating MMH. It turns out that the number of dot products $\mathbf{x}_i \cdot \mathbf{x}_j$ which must be actually computed is sensibly reduced. For this reason, SVM results to be particularly suited for classification problems with very high dimensional feature spaces.

When samples are not linearly separable in the feature space, a non–linear transformation $\phi(\mathbf{x})$ is used.[16] The rationale is to map feature vectors into a higher dimensional feature space where they are linearly separable. With this approach, classification problems which appear quite complex in the original feature space can be afforded by using simple decision functions in the mapped feature space, for instance linear hyperplanes. This point is crucial, since the use of low-complexity decision functions is fundamental in order to have learning machines with a good generalization ability. In order to implement this mapping, the dot products $\mathbf{x} \cdot \mathbf{x}_i$ of Eq. (1) are substituted by the values $\phi(\mathbf{x}) \cdot \phi(\mathbf{x}_i) \equiv K(\mathbf{x}, \mathbf{x}_i)$, commonly referred to as kernel functions. Admissible and typical kernels are:

$$\begin{cases} K(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y} & \text{Linear kernel} \\ K(\mathbf{x}, \mathbf{y}) = (\gamma \mathbf{x}^T \mathbf{y} + r)^d, \gamma > 0 & \text{Polynomial kernel} \end{cases}, \tag{5}$$

where $\gamma$, $r$ and $d$ are kernel parameters selected by the user.

### 2.3.2. *Performance evaluation*

Cross-validation is a common procedure used to train and test a classifier. Given a $n$-dimensional dataset $D$, first the entire dataset is divided into $f$ homogeneous sub-datasets $F_1, F_2, \ldots, F_f$, commonly referred to as folds. The classifier is then trained with the collection of the first $f - 1$ folds $(F_1, F_2, \ldots, F_{f-1})$ and tested on $F_f$, namely the fold left over. The procedure is permuted for each $F_i$, $i = 1, \ldots, f$. Compared to the holdout method, the major advantage of this technique is that larger portions of the dataset can be used for training the classifier. Furthermore, classification performances are estimated as the average error rate over the $f$ test folds $F_i$, thus preventing the problems arising from unfortunate splits of the dataset. As already mentioned, the dataset used in this work is composed of 1000 crops representing the mass class and 5000 crops representing the nonmass class. With the purpose of implementing a 10-folds cross-validation procedure, the dataset is divided into 10 folds, each one containing 100 mass crops and 500 nonmass crops.

For each permutation of the cross-validation procedure, SVM is thus trained with 900 mass crops and 4500 nonmass crops, then tested respectively on 100 and 500.

Classification performances are evaluated using receiver operating characteristic (ROC) curves, namely plots of the classifier's TPF values ($y$-axis, ranging between 0 and 1) versus its FPF values ($x$-axis, ranging between 0 and 1). Reasoning informally, one ROC curve results to be better than another if it lies closer to the upper-left corner. This would mean having higher average TPF values and lower average FPF values. The most common technique to perform ROC curve analysis consists in varying the free parameters of the classifier, thus altering the values of TPF and FPF on the same test set. In this study, ROC curves are calculated by shifting the MMH found by SVM. This is achieved by changing the threshold $b$ in Eq. (1). For each choice of $b$, TPF and FPF values are then computed and plotted. Furthermore, in order to summarize the classification performances, the area $A_z$ under the ROC curve is employed. Its value ranges between 0 and 1 and can be interpreted as the average TPF values over all possible values of FPF. It turns out that one ROC curve results to be better than another if its area $A_z$ is greater. The areas $A_z$s are estimated using the ROCKIT software by Metz *et al.*[24]

### 2.4. *Tests*

The image representations discussed above are evaluated stand alone and with the combined effect of some image processing techniques, namely histogram equalization and resizing.

Histogram equalization transforms the image gray-level intensity values so that the histogram of the equalized image matches an approximately flat histogram. The net effect is to obtain images having higher contrast and exhibiting a larger variety of gray-level intensity values. This actually results in having crops with enhanced edges and boundaries.

Image resizing involves the interpolation of image adjacent pixels to estimate an image with lower dimensions. In this work, bilinear interpolation is used, i.e., each pixel of the resized image results from a weighted average of the pixels in the nearest 2-by-2 neighborhood. The resulting crops are characterized by a lower spatial resolution which supplies an approximate idea about edges and boundaries, but provides an effective picture of the brightness distribution of the pixels.

A further technique, called scaling, is tested in combination with the aforementioned ones. It consists of scaling the 2-sigma interval of correspondent features in the range [0, 1]. Scaling coefficients are calculated for each feature during the training phase, then are used to scale correspondent features both in the train and test set. In this work, correspondent features are represented by correspondent pixels when evaluating the pixel image representation and by correspondent wavelet coefficients when evaluating the wavelet image representations. This technique is very common in the pattern classification community, since it is useful to avoid that features of greater value dominate those of smaller value. Furthermore, since

classification depends mainly on the inner products of feature vectors, the scaling technique is useful to avoid numerical difficulties.

### 2.4.1. *Tests with pixel image representation*

Aimed at evaluating in detail the pixel image representation, three main tests are performed. For the sake of conciseness, they will be referred to as **PixS**, **PixRS** and **PixHRS**.

First, the pixel image representation is evaluated in combination with scaling for different SVM kernels, namely linear and polynomial with degree 2 and 3. With this image representation, SVM is required to classify $64 \times 64$ pixel crops whose correspondent pixels are scaled in the range [0, 1]. Being characterized by scaling, this test will be referred to as **PixS**.

Second, the influence of image resizing is tested by applying bilinear resizing to the crops. The original $64 \times 64$ pixel crops are resized to $16 \times 16$ by means of bilinear resizing. Resized crops are then scaled between [0, 1] and finally classified by using the aforementioned linear and polynomial SVM kernels. Being characterized by resizing and scaling, this test will be referred to as **PixRS**.

Third, histogram equalization is evaluated. In particular, the original $64 \times 64$ pixel crops are all processed by means of histogram equalization. The crops obtained are resized to $16 \times 16$ by means of bilinear resizing, scaled between [0, 1] and finally classified by using the same SVM kernels. Being characterized by histogram equalization, resizing and scaling, this test will be referred to as **PixHRS**.

### 2.4.2. *Tests with discrete wavelet image representation*

As far as the discrete wavelet image representation is concerned, two main tests are performed. They will be referred to as **DwtS** and **DwtHS**.

First, the discrete wavelet image representation is evaluated in combination with scaling for a linear SVM kernel. With this image representation SVM is asked to classify 4032 wavelet coefficients obtained by applying the discrete Haar wavelet transform up to level 3 to the crops, extracting the detail components of levels 1, 2, 3 and scaling them in the interval [0, 1]. Being characterized by scaling, this test will be referred to as **DwtS**.

Second, the influence of histogram equalization is explored by equalizing crops before transforming them with the discrete Haar wavelet transform. In particular, the original $64 \times 64$ pixel crops are all processed by using histogram equalization. The crops obtained are transformed by means of the discrete Haar wavelet transform up to level 3. The detail components of levels 1, 2, 3 are extracted, scaled between [0, 1] and finally classified by using a linear SVM kernel, as for the first test. Being characterized by histogram equalization and scaling, this test will be referred to as **DwtHS**.

Third, the effects of a different choice for the SVM kernel are evaluated. Other than for a linear SVM kernel, the discrete wavelet image representations discussed

above are tested for polynomial kernels with degree 2 and 3. These tests are referred to as **DwtHS2** and **DwtHS3**, the number indicating the degree of the polynomial kernel used.

### 2.4.3. *Tests with overcomplete wavelet image representation*

Similarly to the discrete wavelet image representation, two main tests are performed for the overcomplete case. They will be referred to as **OwtS** and **OwtHS**.

First, the overcomplete wavelet image representation is evaluated in combination with scaling for different SVM kernels, namely linear and polynomial with degree 2 and 3. With this image representation the classification features handled by SVM are 2955 wavelet coefficients obtained by applying the overcomplete Haar wavelet transform to the crops, extracting the decomposition levels 4, 6 and scaling them in the interval [0, 1]. Being characterized by scaling, this test will be referred to as **OwtS**.

Second, the influence of histogram equalization is explored. In particular, the original $64 \times 64$ pixel crops are all processed by using histogram equalization. The obtained crops are transformed by means of the overcomplete Haar wavelet transform. The detail components of levels 4, 6 are extracted, scaled between [0, 1] and finally classified by using linear and polynomial SVM kernels. Being characterized by histogram equalization and scaling, this test will be referred to as **OwtHS**.

### 3. Results and Discussion

As previously mentioned, in order to optimize SVM to the different image representations, several polynomial kernels (i.e., parameter $d$ in Eq. (5)) are evaluated, from degree 1 up to 3. As regards the other parameters (i.e., parameters $\gamma$ and $r$ in Eq. (5) and parameter $C$ in Eq. (3)) classification performances are found to be almost completely unaffected by any change on them. For this reason, they are arbitrarily set equal to unit.

### 3.1. *Results with pixel image representation*

For instance, crops resizing has not a tangible effect on classification performances. This means that the classification results achieved by the scaled pixel image representation (**PixS**) and its correspondent bilinear resized version (**PixRS**) are almost the same. This is an important point, since it demonstrates that similar results can be achieved by using $16 \times 16 = 256$ features instead of $64 \times 64 = 4096$, thus sensibly reducing computational times. Due to the similarity between the performances achieved by **PixS** and **PixRS**, only the ROC curve correspondent to the faster image representation is thus plotted in Fig. 4, namely the latter. A possible explanation for this result must be searched in that masses are generally characterized by a brightness distribution of their gray-level intensity values which is peaked in the center of the region under consideration, whereas non-masses appear as much
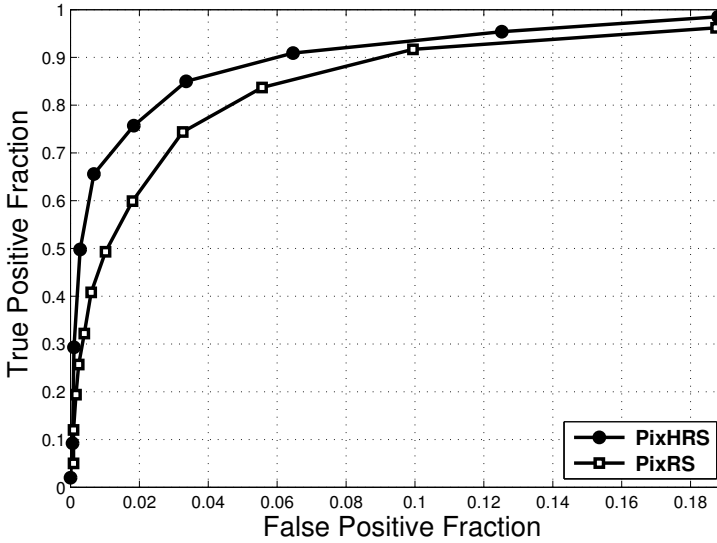
Fig. 4.    ROC analysis of the pixel image representation. 256 classification features. Linear SVM kernel.

more uniform objects. In particular, this heavily discriminant characteristic is quite insensitive to image resizing, since the peaked brightness distribution of masses and that much more uniform of non-masses are not sensibly affected by lowering the spatial resolution.

A further interesting result is that histogram equalization has a very positive effect on classification performances. The original crops processed by means of histogram equalization, bilinear resizing and scaling, in fact, are those achieving the best classification results. As evident from Fig. 4, the ROC curve correspondent to this image representation (**PixHRS**) is significantly better than that correspondent to **PixRS**, in particular for FPF values comprised between 1% and 5%. A practical explanation is that, by enhancing the image contrast, histogram equalization leads to a more evident separation between the central bright pixels of suspicious regions containing a mass and the uniform surrounding tissue. In this sense, it goes precisely in the direction of stressing the aforementioned heavily discriminant characteristic which differentiate masses from non-masses. Note, by the way, that as for the above discussed case the number of features here is equal to $16 \times 16 = 256$.

As regards SVM, the kernel which performs globally better is the linear. This is reasonable. Working with the pixel image representation, correlations among correspondent pixels are much more reliable as features than correlations among distant pixels.[25] In the case of linear kernel, in particular, the correlations considered are precisely those among correspondent pixels. The inner products computed are, in fact, $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})$, where $\mathbf{x}$ and $\mathbf{y}$ are two vectors containing the pixels of two crops.
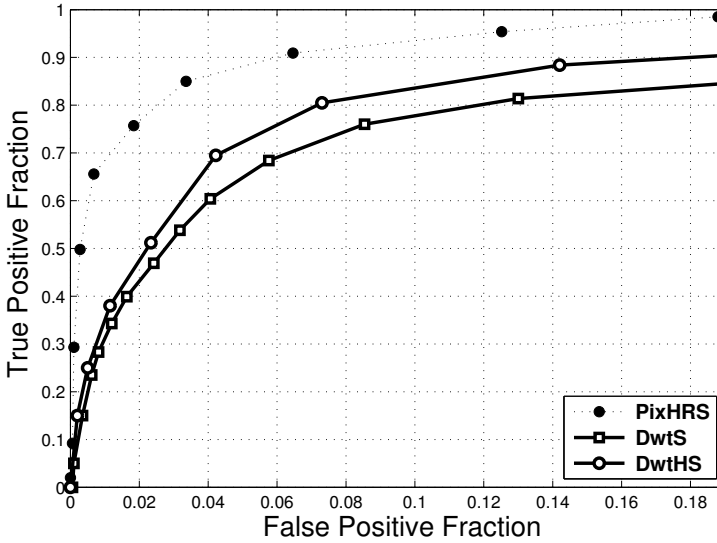
Fig. 5. ROC analysis of the discrete wavelet image representation. 4032 classification features. Linear SVM kernel. The ROC curve of **PixHRS** is depicted for comparison purposes.

### 3.2. *Results with discrete wavelet image representation*

Again, histogram equalization has a positive influence, although smaller, on classification performances. As for the pixel image representation, the original crops processed by means of histogram equalization, transformed by the discrete Haar wavelet transform and finally scaled are those achieving the best classification results among the discrete wavelet image representations. This is evident in Fig. 5, where the ROC curve correspondent to this image representation (**DwtHS**) is slightly better than that correspondent to **DwtS**. The main reason for the weaker impact of histogram equalization on classification performances is that the discrete Haar wavelet transform enhances itself edges and boundaries of images. Looking at Fig. 2, for example, it is evident that discrete wavelet coefficients have a strong response in presence of gray-level intensity differences, for instance edges and boundaries. This makes the improvements of histogram equalization inferior with respect to those reported for the pixel image representation. Note, by the way, that the number of features is 4032 for both **DwtS** and **DwtHS**.

A further interesting aspect is that, contrarily to the pixel image representation, the SVM kernels which performs globally better are the polynomial with degree 2 or 3, see Fig. 6. Dealing with a wavelet image representation, feature vectors are represented by a concatenation of detail components. Each pixel of the original crop is thus represented many times in the feature vectors, namely by one wavelet coefficient in each one of the three detail components. This is repeated for each decomposition level. It is thus evident that, in this case, correlations among distant features are important as well as correlations among correspondent features. For
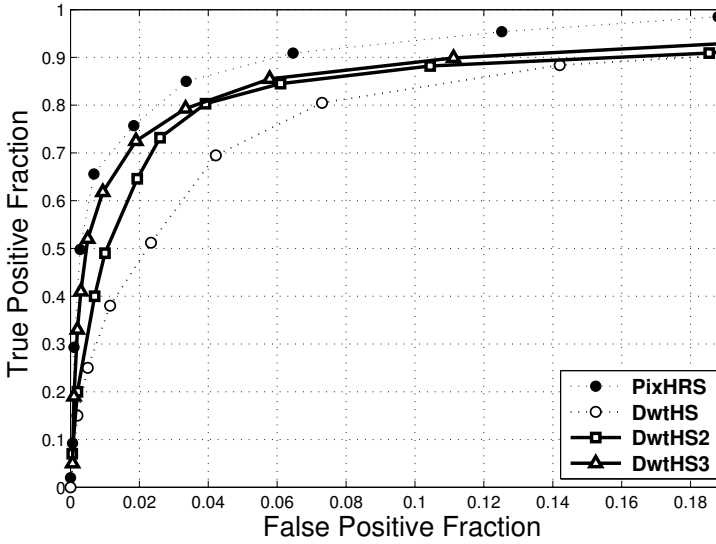
Fig. 6.   ROC analysis of the discrete wavelet image representation. 4032 classification features. Polynomial SVM kernel with degree 2 and 3. The ROC curves of **PixHRS** and **DwtHS** are depicted for comparison purposes.

this reason, SVM kernels which address specifically this task, such as the polynomial kernels with degree 2 or 3 (i.e., $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^2$ or $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^3$), achieve better classification performances.

### 3.3.  *Results with overcomplete wavelet image representation*

Owing to the same reasons discussed for the discrete wavelet image representation, experiments confirm the importance of choosing polynomial SVM kernels with degree higher than one. By evaluating different SVM kernels, it turns out that the overcomplete wavelet image representation is best classified by means of polynomial kernels with degree 2. For this reason, the ROC curves plotted in Fig. 7 correspond to tests performed by using this polynomial kernel, respectively **OwtS2** and **OwtHS2**, the number indicating the degree of the polynomial kernel used.

The influence of histogram equalization on the classification performances, results contradict somehow what is obtained for the discrete wavelet image representation. As evident from Fig. 7, the ROC curve which corresponds to the crops processed by means of histogram equalization, transformed by the overcomplete Haar wavelet transform and finally scaled (**OwtHS2**) is slightly worse than that corresponding to the crops simply transformed by the overcomplete Haar wavelet transform and finally scaled (**OwtS2**). The reason is probably that the combined effect of histogram equalization together with a redundant wavelet analysis enhances too much crops, thus encoding in the wavelet coefficients unnecessary and unimportant image details, for instance noise. As depicted in Fig. 3, in fact, the
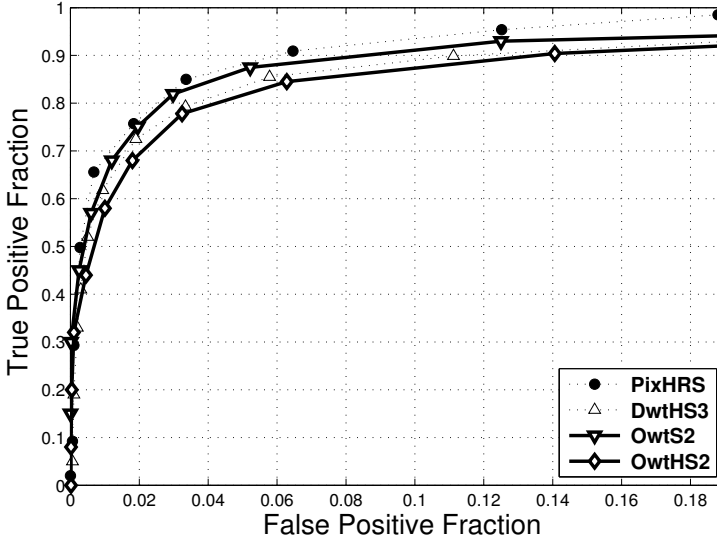
Fig. 7.   ROC analysis of the overcomplete wavelet image representation. 2955 classification features. Polynomial SVM kernel with degree 2. The ROC curves of **PixHRS** and **DwtHS3** are depicted for comparison purposes.

Table 1.   Classification results comparison. $A_z$ and FPF values achieved by the best performing image representations.

| Image representation | # Features | $A_z$ value | FPF for 90% TPF |
|---|---|---|---|
| **PixHRS** | 256 | $0.973 \pm 0.002$ | 6% |
| **DwtHS3** | 4032 | $0.948 \pm 0.004$ | 11% |
| **OwtS2** | 2955 | $0.956 \pm 0.003$ | 7% |

overcomplete wavelet coefficients are much more sensitive to gray-level intensity differences than their discrete wavelet analogous. Note, by the way, that the number of features is 2955 for both **OwtS2** and **OwtHS2**.

### 3.4.  *Comparison of the results*

The tests discussed above show that **PixHRS** is the image representation leading to the best results among all the pixel image representations evaluated, **DwtHS3** among all the discrete wavelet image representations and **OwtS2** among all the overcomplete wavelet image representations. To compare quantitatively the best results obtained for the three image representations tested, Table 1 is presented. Here, other than the values of the area $A_z$ under the ROC curves, the FPF values achieved by **PixHRS**, **DwtHS3** and **OwtS2** are reported for 90% TPF values.

By comparison, the pixel image representation **PixHRS** achieves slightly better classification performances with respect to the others, namely $A_z = 0.973 \pm 0.002$ and 90% TPF values with FPF values of 6%. In particular, the improvement in the

$A_z$ value with **PixHRS** is statistically significant compared to that with **DwtHS3** and **OwtS2** (two-tailed $p$-value $< 0.0001$). The reason is that, by applying histogram equalization and resizing, this pixel image representation enhances and exploits an important characteristic which discriminates suspicious regions containing masses from those representing normal tissue. As already mentioned, this characteristic is the presence in the most part of suspicious regions containing masses of a bright center, compared to the more heterogeneous structures characterizing normal tissue. In this context, the application of histogram equalization enhances the contrast of the suspicious region under study, thus underlying the separation between the central bright pixels and the surrounding tissue. At the same time, by lowering the spatial resolution, image resizing supplies an effective encoding which preserves the peaked or uniform nature of the brightness distribution.

The overcomplete wavelet image representation **OwtS2** achieves slightly worse performances, namely $A_z = 0.956 \pm 0.003$ and 90% TPF values with FPF values of 7%. If the pixel image representation exploits the gray-level intensity content of the suspicious regions, the overcomplete wavelet image representation is exclusively concerned with the difference in gray-level intensity content along different resolutions and orientations of those regions. No direct information about the gray-level intensity values is used. To some extent, the discriminant characteristic exploited is exactly the same as for the pixel image representation, however, not the possible presence of a central bright nucleus is considered, but the possible presence of a fairly sharp boundary delimiting the mass nucleus. To this aim, the rich spatial resolution intrinsically assured by the overcomplete wavelet transform is crucial. Differently from the pixel image representation, where information about the peaked or uniform brightness distribution can be supplied by a relative low resolution encoding, here the mass boundary is best identified and enhanced at finer resolutions, see Fig. 3. A further confirmation for that is represented by the discrete wavelet image representation **DwtHS3**. Even though without achieving statistical significance, however its classification performances are slightly inferior to those described for **OwtS2**, namely $A_z = 0.948 \pm 0.004$ and 90% TPF values with FPF values of 11%. The main reason for that is the lower spatial resolution which characterize this image representation. This, in fact, makes more difficult to identify and enhance the possible boundary delimiting the mass nucleus, see Fig. 2.

Finally, a key aspect to consider when evaluating the three different image representations is the computational time required by each of them for the analysis of an entire radiographic image. Looking at Table 1, it is evident that the number of features characterizing each image representation is quite different, namely 256 for **PixHRS**, 4032 for **DwtHS3** and 2955 for **OwtS2**. This different number of features affects rather sensibly the computational time required by SVM for calculating the dot products in Eq. (1) and therefore assigning a crop to the mass or non-mass class. On a dual Intel Xeon 2.6 GHz PC, **PixHRS** takes approximately 5 seconds for the analysis of an entire radiographic image, whereas **DwtHS3** and **OwtS2** respectively 1.5 and 1 minutes. It turns out that, other than for its higher

classification performances, **PixHRS** provides an image representation which is also much more reliable for real-time CAD implementations.

## 4. Conclusion

In a recent work,[14] our group has proposed a mass detection algorithm able to predict the presence or absence of tumoral masses in mammographic suspicious regions. With this approach, classification features are embodied by the image representation used in order to encode the suspicious regions under study, whereas classification is performed by means of an SVM classifier.

To investigate whether different image representations can achieve better or similar results with respect to the originally developed overcomplete wavelet image representation, two new image representations are tested herein, namely a pixel and a discrete wavelet image representation. To this purpose, a dataset of 6000 suspicious regions is collected from the DDSM database, 1000 of these regions being biopsy-proven benign or malignant masses, whereas the remaining 5000 representing normal tissue.

As far as the the pixel image representation is concerned, suspicious regions are encoded in the simplest way, namely by just concatenating their gray-level intensity values. In particular, best results are achieved by first processing the original $64 \times 64$ pixel suspicious regions by means of histogram equalization, then resizing the resulting crops to $16 \times 16$ by means of bilinear resizing and finally scaling them in the interval $[0, 1]$. The resulting 256 gray-level intensity values are then classified by means of a linear SVM kernel. By using this image representation, referred to as **PixHRS**, 90% TPF values are achieved with FPF values of 6% ($A_z = 0.973 \pm 0.002$).

For the discrete wavelet image representation, instead, best results are obtained by first processing the original $64 \times 64$ pixel suspicious regions by means of histogram equalization, then transforming the resulting crops by using the discrete Haar wavelet transform, extracting the detail components of levels 1, 2, 3 and finally scaling them in the interval $[0, 1]$. The resulting 4032 discrete wavelet coefficients are then classified by means of a polynomial SVM kernel with degree 3. With this image representation, referred to as **DwtHS3**, 90% TPF values are achieved with FPF values of 11% ($A_z = 0.948 \pm 0.004$).

Finally, with the overcomplete image representation image representation, suspicious regions are encoded by applying the overcomplete Haar wavelet transform, extracting the detail components of levels 4, 6 and scaling them in the interval $[0, 1]$. The 2955 overcomplete wavelet coefficients which result from this approach are then classified by means of a polynomial SVM kernel with degree 2. The results achieved by this image representation, referred to as **OwtS2**, show 90% TPF values with FPF values of 7% ($A_z = 0.956 \pm 0.003$).

The results reported here imply very high classification performances. This is much more evident when compared to published results from other mass detection

algorithms. In particular, the improvement in the $A_z$ value achieved by **PixHRS** over that of **DwtHS3** and **OwtS2** is statistically relevant (two-tailed $p$-value $<$ 0.0001). Combined with the lower computational times required for the analysis of an entire radiographic image, these aspects make **PixHRS** preferable for real-time CAD applications. Despite of that, we believe that the combination of the three image representations discussed herein could result very useful in order to sensibly improve the global efficiency of CAD systems. To this purpose, our efforts are currently invested in studying feasible strategies aimed at merging the opinion of each image representation regarding the presence or absence of a tumoral mass in a mammographic suspicious region.

## References

1. A. Jemal, T. Murray, E. Ward, A. Samuels, R. C. Tiwari, A. Ghafoor, E. J. Feuer and M. J. Thun, *CA*: *A Cancer J. Clinicians* **55**, 10 (2005).
2. R. E. Bird, T. W. Wallace and B. C. Yankaskas, *Radiology* **184**, 613 (1992).
3. R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy and E. A. Sickles, *Radiology* **219**, 192 (2001).
4. M. L. Giger, K. Doi, H. MacMahon, R. M. Nishikawa, K. R. Hoffmann, C. J. Vyborny, R. A. Schmidt, H. Jia, K. Abe and X. Chen, *Radiographics* **13**, 647 (1993).
5. N. Ibrahim, H. Fujita, T. Hara and T. Endo, *Phys. Med. Biol.* **42**, 2577 (1997).
6. A. Bazzani, A. Bevilacqua, D. Bollini, R. Brancaccio, R. Campanini, N. Lanconelli and D. Romani, *Int. J. Mod. Phys. C* **11**, 901 (2000).
7. M. L. Giger, N. Karssemeijer and S. G. Armato, *IEEE Trans. Med. Imaging* **20**, 1205 (2001).
8. M. L. Giger, *Comput. Science Engin.* **2**, 39 (2000).
9. D. Wei, H. P. Chan, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler and M. M. Goodsitt, *Med. Phys.* **22**, 1501 (1995).
10. B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler and M. M. Goodsitt, *IEEE Trans. Med. Imaging* **15**, 598 (1996).
11. Y. H. Chang, L. A. Hardesty, C. M. Hakim, T. S. Chang, B. Zheng, W. F. Good and D. Gur, *Med. Phys.* **28**, 455 (2001).
12. A. H. Baydush, D. M. Catarious, C. K. Abbey and C. E. Floyd, *Med. Phys.* **30**, 1781 (2003).
13. G. D. Tourassi, R. Vargas-Voracek, D. M. Catarious and C. E. Floyd, *Med. Phys.* **30**, 2123 (2003).
14. R. Campanini, D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi and M. Roffilli, *Phys. Med. Biol.* **49**, 961 (2004).
15. V. Vapnik, *Statistical Learning Theory* (John Wiley, 1998).
16. K. R. Müller, S. Mika, G. Rätsch and K. Tsuda, *IEEE Trans. Neural Networks* **12**, 181 (2001).
17. A. Bazzani, A. Bevilacqua, D. Bollini, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi and D. Romani, *Phys. Med. Biol.* **46**, 1651 (2001).
18. C. Z. Cai, W. L. Wang and Y. Z. Chen, *Int. J. Mod. Phys. C* **14**, 575 (2003).
19. E. P. Simoncelli, W. T. Freeman, E. H. Adelson and D. J. Heeger, *IEEE Trans. Inform. Th.* **38**, 587 (1992).
20. S. Mallat, *IEEE Trans. Pattern Anal. Mach. Intelligence* **11**, 674 (1989).

21. M. Heath, K. W. Bowyer, D. Copans, R. Moore and P. Kegelmeyer, *The Digital Database for Screening Mammography Digital Mammography: IWDW2000 5th International Workshop on Digital Mammography* (Medical Physics Publishing, 2000).
22. J. R. Sanchez, *Int. J. Mod. Phys. C* **14**, 491 (2003).
23. W. H. Steeb, Y. Hardy and R. Stoop, *Int. J. Mod. Phys. C* **13**, 771 (2002).
24. C. E. Metz, B. A. Herman and J. H. Shen, *Stat. Med.* **17**, 1033 (1998).
25. B. Schölkopf, P. Y. Simard, A. J. Smola and V. N. Vapnik, *Advances in Neural Information Processings Systems* (MIT Press, Cambridge, MA, 1998).