Discriminating Mass from Normal Breast Tissue: A Novel Ranklet Image Representation for ROI Encoding

Matteo Masotti
Department of Physics, University of Bologna,
Viale Berti–Pichat 6/2, 40127, Bologna, Italy
matteo.masotti@bo.infn.it
http://www.bo.infn.it/∼masotti/

Abstract

A support vector machine (SVM) classifier is used to determine whether regions of interest (ROIs) found on breast radiographic images contain mass or normal tissue. Before being presented to SVM, ROIs are encoded by means of a specific image representation. The coefficients resulting from the encoding are then used as classification features. Pixel and wavelet image representations have already been discussed in one of our previous works. A novel orientation–selective, non–parametric and multi–resolution image representation is developed and evaluated herein, namely a ranklet image representation. From the digital database for screening mammography (DDSM) collected by the University of South Florida, a database of ROIs is generated. A total of 1000 ROIs containing diagnosed masses are extracted from the DDSM benign and malignant cases, whereas 5000 ROIs containing normal tissue are extracted from the DDSM normal cases. The area A_z under the receiver operating characteristic curve is adopted for performance evaluation. By achieving A_z values of 0.978 ± 0.003, experiments demonstrate better classification results with respect to those reached by the previous image representations. In particular, the improvement on the A_z value over that achieved by the wavelet image representations is statistically relevant with two–tailed p–value < 0.0001.

1 Introduction

Masses are lesions of the breast tissue associated very often with the presence of cancer. They appear in radiographic images as regions with elevated intensity and size ranging from 3 mm to 30 mm [7]. In order to help radiologists in their identification, computer–aided detection (CAD) systems have been recently introduced [2]. By using computer vision and artificial intelligence techniques, these systems detect suspicious regions independently from the human reader, thus providing the radiologist with a second opinion.

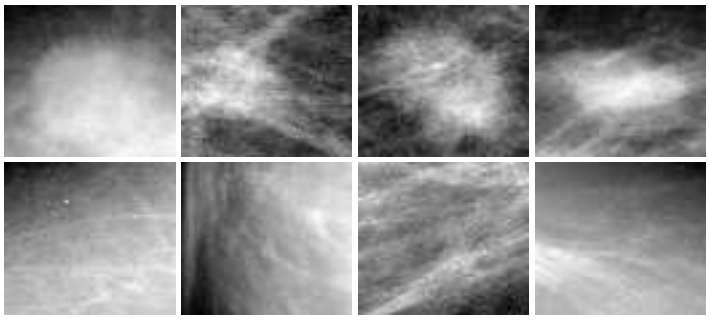


Figure 1: The two classes. Mass class (top) vs. non–mass class (bottom).

Our group is currently developing a CAD system for which mass detection is formulated as a supervised binary classification problem. At each location of the radiographic image, a region of interest (ROI) is first extracted, then encoded by means of a specific image representation and finally submitted to a previously trained support vector machine (SVM) classifier which guesses whether mass or normal tissue is present. Several image representations have already been developed and evaluated in one of our more recent works, namely a pixel, a discrete wavelet and an overcomplete wavelet image representation.

With the intention of improving classification performances, a novel image representation is developed and evaluated in this paper. Being based on an orientation–selective, non–parametric and multi–resolution transform known with the name of ranklet transform [8], in the following it will be referred to as ranklet image representation.

2 Data and Methods

The mass detection approach proposed herein consists of two steps. First, each ROI is submitted to the ranklet transform, thus being encoded by means of the ranklet image representation into a feature vector $\boldsymbol{x}$. Then, the ranklet coefficients stored in the feature vector $\boldsymbol{x}$ are presented to a previously trained SVM which decides whether mass or normal tissue is present.

2.1 ROI database

A database of ROIs is generated from the digital database for screening mammography (DDSM) collected by the University of South Florida [3], see Fig.1. A total of 1000 diagnosed masses are extracted from the DDSM benign and malignant cases using the provided ground truth annotations. A square crop centered on the location of each annotated mass is selected. The size is chosen so that 70% of its area is occupied by the annotated mass and the remaining 30% by background. For the non–mass class, a total of 5000 square crops are randomly extracted from the DDSM normal cases. Since SVM deals exclusively with dimensionally homogeneous vectors, all the crops are resized to an arbitrarily prefixed size of 64×64 pixels. To this purpose, bilinear resizing is chosen.

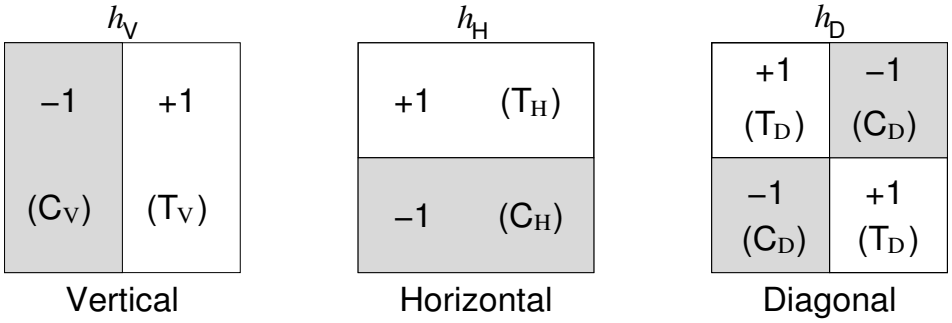


Figure 2: The three Haar wavelet supports h_{V}, h_{H} and h_{D}. From left to right, the vertical, horizontal and diagonal Haar wavelet supports.

2.2 The ranklet transform

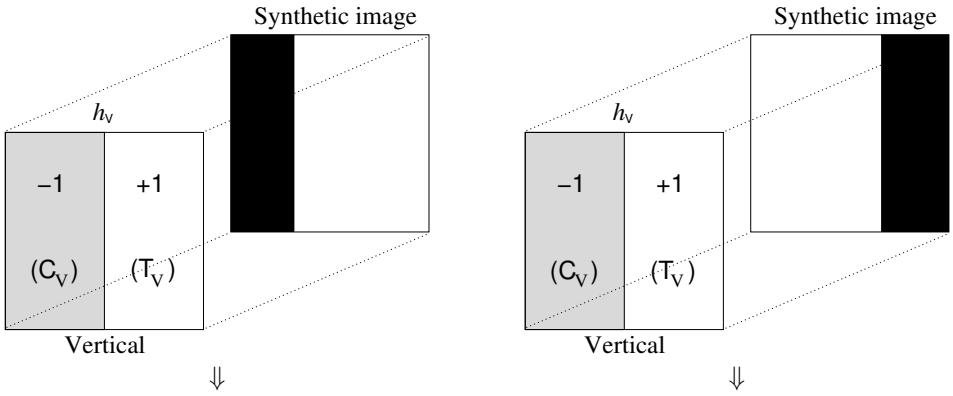
The ranklet transform is an orientation–selective, non–parametric and multi–resolution transform which has already been successfully exploited in targeting image classification tasks, specifically face recognition in image frames [9].

Suppose that an image is constituted by $p_1, \ldots, p_N$ pixels. The ranklet transform is defined by first splitting the N pixels into two subsets T and C of size $N/2$, thus assigning half of the pixels to the subset T and half to the subset C. The two subsets T and C are defined being inspired by the Haar wavelet supports depicted in Fig. 2. In particular, for the vertical Haar wavelet support, the two subsets T_{V} and C_{V} are defined; similarly for the horizontal and diagonal ones. The definition of the aforementioned Haar wavelet supports forms the basis for the orientation–selective property of the ranklet transform.

The second step consists in computing and normalizing in the range $[-1, +1]$ the number of pixel pairs (p_m, p_n), with $p_m \in \mathsf{T}$ and $p_n \in \mathsf{C}$, such that the intensity value of p_m is higher than the intensity value of p_n. This is done for each orientation, namely vertical, horizontal and diagonal. A direct calculation of this quantity would require approximately $O(N^2)$ operations. Nevertheless, it can be demonstrated that the same quantity can be calculated in approximately $O(NLogN)$ operations in the following way [8]:

$$R_j = \frac{\sum_{p \in \mathsf{T}_j} \pi(p) - \frac{N}{4}(\frac{N}{2} + 1)}{\frac{N^2}{8}} - 1, \quad j = \mathsf{V}, \mathsf{H}, \mathsf{D} \tag{1}$$

where $\sum_{p \in \mathsf{T}_j} \pi(p)$ is the sum of the pixels' ranks $\pi(p)$ in T_j. The geometric interpretation of the derived ranklet coefficient R_j is quite simple. Suppose that the image we are dealing with is characterized by a vertical edge, with the darker side on the left, where C_{V} is located, and the brighter side on the right, where T_{V} is located, see Fig.3. R_{V} will be close to $+1$ as many pixels in T_{V} will have higher intensity values than the pixels in C_{V}. Conversely, R_{V} will be close to -1 if the dark and bright side are reversed. Horizontal edges or other patterns with no global left–right variation of intensity will give a value close to 0. Analogous considerations can be drawn for the other ranklet coefficients, R_{H} and R_{D}. In this context, the use of the pixels' ranks, rather than their intensities, forms the basis for the non–parametric property of the ranklet transform.



$R_V = +0.59, (R_H = 0, R_D = 0)$ $R_V = -0.59, (R_H = 0, R_D = 0)$

Figure 3: Ranklet transform applied to some synthetic images. The application of the vertical Haar wavelet support (h_V) results in $R_V = \pm 0.59$, depending of the synthetic image analyzed. Conversely, due to symmetry reasons, the application of the horizontal and diagonal Haar wavelet supports (h_H, h_D) results in $R_H = 0$ and $R_D = 0$, regardless of the synthetic image analyzed.

Finally, the close correspondence between the Haar wavelet transform and the ranklet transform leads directly to the extension of the latter to its multi–resolution formulation. Similarly to what is done for the bi–dimensional Haar wavelet transform, the ranklet coefficients can be computed at different resolutions by simply stretching and shifting the Haar wavelet supports. The multi–resolution ranklet transform of an image is thus a set of triplets of vertical, horizontal and diagonal ranklet coefficients, each one corresponding to a specific stretch and shift of the Haar wavelet supports. Suppose, for example, that the horizontal and vertical shifts of the Haar wavelet supports along the horizontal and vertical dimensions of the image are of 1 pixel. The number nT of triplets $R_{V,H,D}$ at each resolution is thus computed as $nT = (I + 1 - S)^2$, where I and S represent the linear dimension of the image and that of the Haar wavelet support, respectively. Here, the possibility of computing ranklet coefficients at different resolutions forms the basis for the multi–resolution property of the ranklet transform.

2.3 SVM classifier

SVM constructs a binary classifier from a set of l training samples which consists of labeled patterns $(\boldsymbol{x}_i, y_i) \in \mathbb{R}^N \times \{\pm 1\}$, with $i = 1, \ldots, l$ [10]. Taking values $+1$ or -1, the label y_i indicates the class membership (mass or non–mass) of the correspondent feature vector $\boldsymbol{x}_i$ which contains a certain number of ranklet coefficients. In order to separate the two classes, SVM selects the maximal margin hyperplane, namely the hyperplane which causes the largest separation in the feature space between itself and the borderline training samples of the two classes (see Fig.4):

$$f(\boldsymbol{x}) = sgn(\boldsymbol{w} \cdot \boldsymbol{x} + b) = sgn\left(\sum_{i=1}^{l} y_i \alpha_i (\boldsymbol{x} \cdot \boldsymbol{x}_i) + b\right) \tag{2}$$

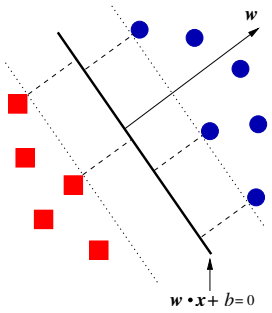


Figure 4: Maximal margin hyperplane separating two classes in a two–dimensional feature space.

The coefficients α_i and b are calculated by solving the quadratic programming problem:

$$\begin{cases} \max\limits_{\boldsymbol{\alpha}} & J = \sum\limits_{i=1}^{l} \alpha_i - \dfrac{1}{2} \sum\limits_{i,j=1}^{l} \alpha_i \alpha_j (\boldsymbol{x}_i \cdot \boldsymbol{x}_j) y_i y_j \\ \text{subject to} & \sum\limits_{i=1}^{l} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \ldots l \end{cases} \tag{3}$$

where C is a regularization parameter and J is the cost function to maximize. An unseen test sample $\boldsymbol{x}$ is assigned to the class $+1$ if $f(\boldsymbol{x}) \geq 0$ and to the class -1 otherwise. When samples are not linearly separable in the feature space, a non–linear transformation $\phi(\boldsymbol{x})$ is used [6]. 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. In order to implement this mapping, the dot products $\boldsymbol{x} \cdot \boldsymbol{x}_i$ are substituted by the values $\phi(\boldsymbol{x}) \cdot \phi(\boldsymbol{x}_i) \equiv K(\boldsymbol{x}, \boldsymbol{x}_i)$, commonly referred to as kernel functions. Admissible and typical kernels are the linear kernel $K(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{y}$, the polynomial kernel $K(\boldsymbol{x}, \boldsymbol{y}) = (\gamma \boldsymbol{x} \cdot \boldsymbol{y} + r)^d$, etc.

2.4 Tests performed

To determine an optimal setup for the ranklet image representation's and SVM classifier's parameters, two main tests are performed. The ranklet image representation is therein indicated as **RankletS**$_{[\text{RES1,RES2,}\ldots]}$. The pre–fix **Ranklet** stands for its being a ranklet image representation, whereas the post–fix **S** for its having correspondent features (i.e. ranklet coefficients) scaled in the interval $[-1, 1]$. An eventual integer number on the right side of the post–fix **S** indicates the degree of the polynomial SVM kernel used for classification. The subscript $_{[\text{RES1,RES2,}\ldots]}$ indicates the resolutions at which the multi–resolution ranklet transform is performed, namely the linear dimension of the Haar wavelet supports used. At this regard, the multi–resolution ranklet transform of a 64×64 crop results in a huge amount of ranklet coefficients, specifically when fine resolutions are considered. The 6000 crops of the image database are thus further resized from their original 64×64 pixel size. After initial experimentation, the new pixel size is chosen to be 16×16.

The first test performed is intended to understand the influence of different SVM kernels on classification performances. Original crops are resized from 64×64 pixel size to 16×16 by means of bilinear resizing. The multi–resolution ranklet transform is then applied at resolutions $[16, 8, 4, 2]$ to the resized crops, thus producing 1428 classification features for each ROI. Using as image representation the resulting ranklet coefficients, several SVM kernels are hence varied, namely linear, polynomial with degree 2 and 3.

The second test performed is intended to comprehend the effects of the multi–resolution property on classification performances. As for the previous test, original crops are resized from 64×64 pixel size to 16×16 by means of bilinear resizing. The multi–resolution ranklet transform is then applied to the resized crops by using several combinations of different resolutions. The number of resulting classification features hence varies according to the resolutions at which the analysis is performed. Owing to the results that will be discussed in Section 3.1, a polynomial SVM kernel with degree 3 is used.

2.5 Performance evaluation

In order to evaluate the performances of the classifier, a 10–folds cross–validation procedure is adopted [4]. The dataset is partitioned into 10 distinct and homogeneous folds, then SVM is trained with the collection of the first 9 folds (i.e., 900 mass crops and 4500 non–mass crops) and tested on the fold left out (i.e., 100 and 500). Training and test are repeated 10 times by changing the test fold in a round–robin manner. Compared to a single training and test, the major advantage of this technique is that classification performances are averaged over the 10 test folds, thus preventing the problems arising from unfortunate splits of the dataset. In particular, classification results are presented in terms of the average receiver operating characteristic (ROC) curve of the system [5]. ROC curves are plots showing on the y–axis the classifier's sensitivity (i.e., the fraction of masses correctly classified as masses), whereas on the x–axis its false positive fraction (i.e., the fraction of non–masses misclassified as masses). Reasoning informally, one ROC curve results better than another if it lies closer to the upper–left corner of the graph. This would mean having higher sensitivity and lower false positive fraction. 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 sensitivity over all possible values of false positive fraction. It turns out that one ROC curve results to be better than another if its area A_z is greater. The ROC curves and their associated areas A_z are estimated using the`ROCKIT` software by Metz *et al.* [1].

3 Results

3.1 SVM kernels

Classification performances are improved when the polynomial degree of the SVM kernel is increased. The linear SVM kernel (**RankletS**$_{[16,8,4,2]}$) achieves average A_z values of 0.946 ± 0.004, the polynomial SVM kernel with degree 2 (**RankletS2**$_{[16,8,4,2]}$) of 0.974 ± 0.003 and the polynomial SVM kernel with degree 3 (**RankletS3**$_{[16,8,4,2]}$) of 0.978 ± 0.003. In particular, the difference in the A_z values between the polynomial SVM kernels with degree 2 or 3 and the linear SVM kernel is statistically significant (two–tailed p–value < 0.0001).

Table 1: Summary of the average A_z values achieved by the different ranklet image representations evaluated.

Image representation	# Features	A_z
RankletS$_{[16,8,4,2]}$	1428	0.946 ± 0.004
RankletS2$_{[16,8,4,2]}$	1428	0.974 ± 0.003
RankletS3$_{[16,8,4,2]}$	1428	0.978 ± 0.003
RankletS3$_{[16,14,12,10,8,6,4,2]}$	2040	0.976 ± 0.003
RankletS3$_{[16,2]}$	678	0.977 ± 0.002
RankletS3$_{[16,14,12,10]}$	252	0.957 ± 0.004

3.2 Ranklet resolutions

Classification performances are quite similar when the whole range of low, intermediate and high resolutions is considered or a sampled version of them is taken into account. For example, **RankletS3**$_{[16,14,12,10,8,6,4,2]}$ and **RankletS3**$_{[16,8,4,2]}$ perform almost identically, the former achieving average A_z values of 0.976 ± 0.003. The difference between their areas does not reach statistical significance. Intermediate resolutions are not essential for classification purposes. The results obtained by ignoring intermediate resolutions (e.g. **RankletS3**$_{[16,2]}$, average A_z values of 0.977 ± 0.002), in fact, are not significantly different from those obtained by **RankletS3**$_{[16,8,4,2]}$. Conversely, high and low resolutions are quite important for classification purposes. The results achieved by ignoring high resolutions (e.g. **RankletS3**$_{[16,14,12,10]}$, average A_z values of 0.957 ± 0.004), in fact, perform significantly worse than those achieved by **RankletS3**$_{[16,8,4,2]}$ (two–tailed p–value < 0.0001). Similar results are found when low resolutions are ignored.

4 Discussion

The performances presented above show that **RankletS3**$_{[16,8,4,2]}$ provides the best classification results among all the ranklet image representations tested, namely an average A_z value of 0.978 ± 0.003. See Tab. 1 for a summary of the results reached by the different ranklet image representations evaluated.

4.1 SVM kernels

The achievement of better classification results with polynomial SVM kernels rather than linear ones is quite conceivable. When dealing with a multi–resolution encoding, such as the ranklet image representation, dot products among distant features are important as well as dot products among correspondent features. In this case, in fact, each pixel of the ROI is represented many times in the encoded feature vector, namely by a number of ranklet coefficients which is proportional to the number of multi–resolution levels at which the analysis is performed. Being computed as $K(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} \cdot \boldsymbol{y})^2$ or $K(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} \cdot \boldsymbol{y})^3$, polynomial SVM kernels with degree 2 or 3 are thus able to take this aspect into account, hence to provide better classification performances.

4.2 Ranklet resolutions

As far as ranklet resolutions are concerned, the greater importance of low and high resolutions with respect to intermediate ones relies in the ability of the former in encoding characteristics particularly useful to discriminate ROIs containing masses from ROIs containing normal tissue. Low ranklet resolutions, in fact, encode informations concerning the symmetry of the intensity distribution which characterizes the image when analyzed at coarse resolutions. Owing to that, they are particularly suited to encode informations concerning the presence or absence in the ROI of a bright centered nucleus surrounded by more heterogeneous structures characterizing normal tissue. Conversely, high ranklet resolutions encode informations concerning the symmetry of the intensity distribution which characterizes the image when analyzed at fine resolutions. Differently from low resolutions, they are hence particularly suited to encode informations concerning the presence or absence in the ROI of a boundary delimiting the bright centered nucleus.

4.3 Comparison with previous image representations

Finally, for comparison purposes, the ROC curve analysis of the three best performing pixel (**PixHRS**), discrete wavelet (**DwtHS3**) and overcomplete wavelet (**OwtS2**) image representations previously developed and evaluated is reported in Fig. 5. They are obtained on the same ROI database. By pairwise comparison, it turns out that the improvement in the A_z value with **RankletS3**$_{[16,8,4,2]}$ over that of **PixHRS** ($A_z = 0.973 \pm 0.002$) does not reach statistical significance. Conversely, by showing a two–tailed p–value < 0.0001, the improvement over that of **DwtHS3** ($A_z = 0.948 \pm 0.004$) and **OwtS2** ($A_z = 0.956 \pm 0.003$) is statistically significant.

5 Conclusions

In this study, mass detection is targeted as a supervised binary classification problem. ROIs are first encoded by means of a specific image representation. The coefficients resulting from the encoding are then classified as mass or normal tissue by a previously trained SVM classifier. To investigate whether better classification performances can be achieved with respect to pixel and wavelet image representations previously presented, a novel ranklet image representation is developed and evaluated. Results show an average A_z value of 0.978 ± 0.003 and demonstrate an improvement over the average A_z values reached by previous image representations. With respect to the wavelet image representations, this improvement is statistically significant with two–tailed p–value < 0.0001. In addition, being based on pixels' ranks rather than pixels' intensities, this image representation is well suited to prove tolerant also toward remarkable variations in the pixels' intensity of ROIs. It is hence well suited to prove much more robust than intensity–based image representations (e.g. pixel or wavelet image representations) in presence of different mammographic detectors or acquisition conditions.

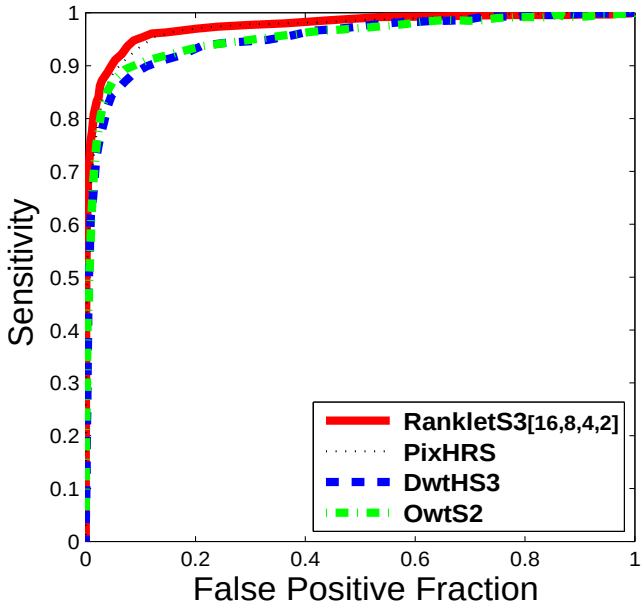


Figure 5: ROC curves comparison. **RankletS3**$_{[16,8,4,2]}$ (A_z value of 0.978 ± 0.003) is the best ranklet image representation evaluated. **PixHRS** (0.973 ± 0.002), **DwtHS3** (0.948 ± 0.004) and **OwtS2** (0.956 ± 0.003) are the best pixel, discrete wavelet and overcomplete wavelet image representations previously developed and evaluated.

References

[1] B. A. Herman C. E. Metz and J. H. Shen. Maximum–likelihood estimation of receiver operating characteristic (ROC) curves from continuously–distributed data. *Statistical Medicine*, 17:1033–1053, 1998.

[2] 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. An intelligent workstation for computer–aided diagnosis. *Radiographics*, 13:647–656, 1993.

[3] M. Heath, K. W. Bowyer, D. Copans, R. Moore, and P. Kegelmeyer. The digital database for screening mammography. In *International Workshop on Digital Mammography*, pages 212–218, 2000.

[4] R. Kohavi. A study of cross–validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Intelligence*, pages 1137–1145, 1995.

[5] C. E. Metz. ROC methodology in radiologic imaging. *Investigative Radiology*, 21:720–733, 1986.

[6] K. R. Müller, S. Mika, G. Rätsch, and K. Tsuda. An introduction to kernel–based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–201, 2001.

[7] E. Shaw de Paredes. Radiographic breast anatomy: Radiologic signs of breast cancer. *RSNA Categorical Course in Physics*, pages 35–46, 1993.

[8] F. Smeraldi. Ranklets: orientation selective non–parametric features applied to face detection. In *International Conference on Pattern Recognition*, volume 3, pages 379–382, August 2002.

[9] F. Smeraldi. A nonparametric approach to face detection using ranklets. In *International Conference on Audio and Video–based Biometric Person Authentication*, pages 351–359, June 2003.

[10] V. Vapnik. *Statistical Learning Theory*. J. Wiley, New York, 1998.