



ELSEVIER

Pattern recognition techniques for automatic detection of suspicious-looking anomalies in mammograms

Tomasz Arodz^{a,*}, Marcin Kurdziel^{a,1}, Erik O.D. Sevre^{b,2}, David A. Yuen^{b,2}

^a *Institute of Computer Science, AGH University of Science and Technology, al. Mickiewicza 30, 30-059, Kraków, Poland*

^b *Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN 55455, USA*

Received 15 March 2004; received in revised form 24 March 2005; accepted 25 March 2005

KEYWORDS

Mammogram analysis;
Computer-aided
diagnosis;
Machine learning

Summary We have employed two pattern recognition methods used commonly for face recognition in order to analyse digital mammograms. The methods are based on novel classification schemes, the AdaBoost and the support vector machines (SVM). A number of tests have been carried out to evaluate the accuracy of these two algorithms under different circumstances. Results for the AdaBoost classifier method are promising, especially for classifying mass-type lesions. In the best case the algorithm achieved accuracy of 76% for all lesion types and 90% for masses only. The SVM based algorithm did not perform as well. In order to achieve a higher accuracy for this method, we should choose image features that are better suited for analysing digital mammograms than the currently used ones.

© 2005 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Today breast cancer is the second major killer among American women [1]. Recently, each year around 40,000 women dies from breast cancer and over 200,000 develops new cases of invasive

breast cancer. Apart from the invasive cancer, at least 55,000 cases of the in situ breast cancer are also diagnosed each year. With current treatment options, the 5-year survival rate for the localised breast cancer can be as high as 97%. However, this rate drops to 78% in the case of a regionally advanced disease and to 23% for the fast growing breast cancer. Consequently, early detection of the breast cancer is crucial for an efficient therapy. The American Cancer Society recommends every woman older than around 40 years to undergo annually mammogram examination. The sensitivity of screening mammography can be improved with the use of Computer-Aided Detection (CAD) systems.

* Corresponding author. Tel.: +48 12 617 3497;
fax: +48 12 633 9406.

E-mail addresses: arodz@agh.edu.pl (T. Arodz),
kurdziel@agh.edu.pl (M. Kurdziel), esevre@msi.umn.edu
(E.O.D. Sevre), davey@krissy.geo.umn.edu (D.A. Yuen)

¹ Tel.: +48 12 617 3497; fax: +48 12 633 9406.

² Tel.: +1 612 624 9801; fax: +1 612 625 3819.

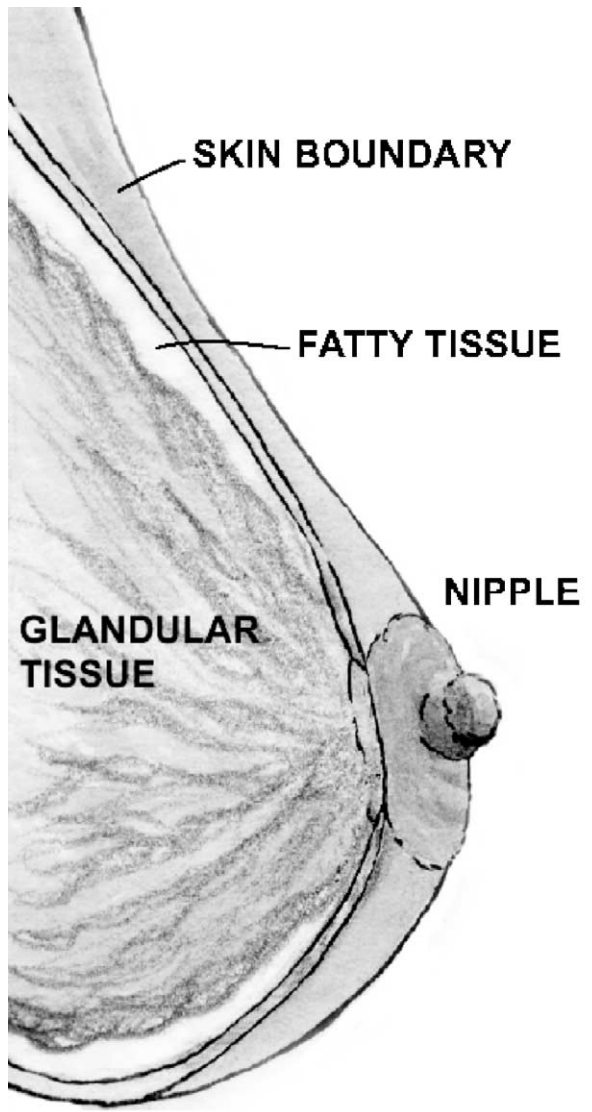


Fig. 1 Global structure of the breast. Different major regions of the breast include: nipple, fatty and glandular tissue and skin. Courtesy of Ben Holtzmann with help from Lilli Yang.

In Ref. [2] the interpretations of around 13,000 screening mammograms by two experienced radiologist supported by the CAD system were compared to the interpretations given prior to the use of CAD. These studies revealed a relative increase of the detection rate from 0.32 to 0.36%. At the same time, there were no adverse effects on the recall rate or positive predictive value for biopsy.

The global structure of the breast is presented in the Fig. 1. The breast tissue may contain two types of cancer indicators commonly evaluated by CAD systems, i.e., masses and microcalcifications. Microcalcifications appear on a mammogram image as small, bright tissue protrusions. According to Ref. [3,4] the probability of malignant process

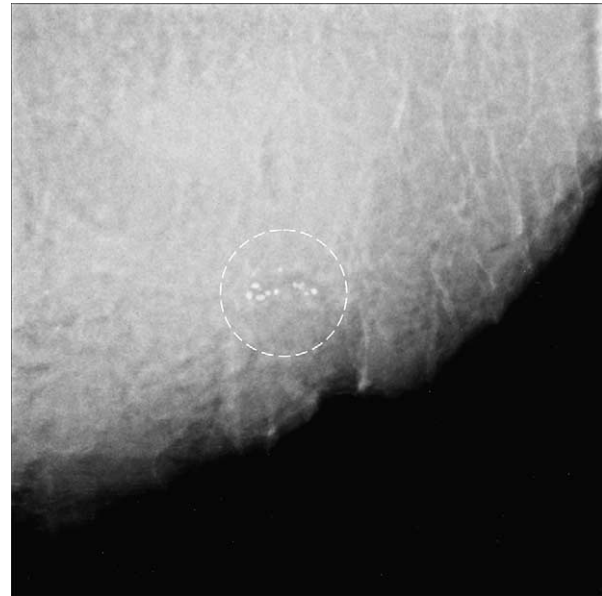


Fig. 2 A mammographic image depicting the breast region with a size 5×5 cm. On the image a cluster of microcalcifications is marked. Such clusters frequently arise from pathological processes in breast tissues and thus are considered important in breast cancer diagnosis. The contrast on the image has been manually adjusted for the best visibility of the microcalcifications. Regions with a less dense tissue have been suppressed and are depicted as black areas.

within the breast depends on the number, distribution and morphology of microcalcifications. When not in a group, microcalcifications are irrelevant from a diagnostic standpoint. If at least four or five microcalcifications are present and form a cluster, the probability of an abnormal process within the breast is significant. An example of a mammogram with one such cluster is depicted in Fig. 2. Microcalcifications vary significantly in size, shape and distribution. Typically, homogenous clusters consisting of larger, round or oval microcalcifications indicate a benign process. An example cluster of this type is pictured in Fig. 3. On the other hand, heterogeneous clusters consisting of smaller microcalcifications, and microcalcifications with irregular shapes are associated with a high risk of cancer. A good example of this type of cluster is pictured in Fig. 4. In practice it is very difficult to decide whether microcalcification clusters are benign or malignant. In a significant number of cases the clusters cannot be observed clearly. Moreover, it is common for a cluster of microcalcifications to reveal morphological features that cannot be clearly classified as being benign or malignant.

Other important breast abnormalities shown in mammograms are the masses. They may occur in various parts of breasts and have different sizes,

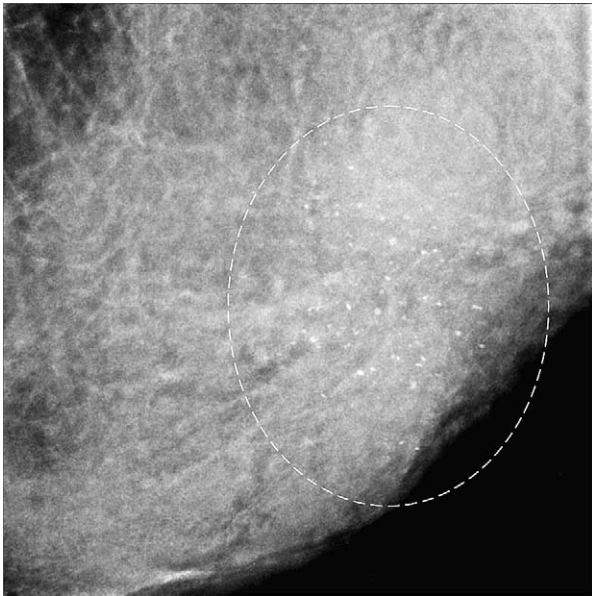


Fig. 3 A mammographic image depicting the breast region with an area of $5\text{ cm} \times 5\text{ cm}$. In this image a cluster of micro-calcifications has been marked by the white dashed oval. These micro-calcifications are small and round. Such appearances are sometimes referred to as punctuated calcifications and are rarely associated with cancer. A benign process is more probable. The image contrast has been manually adjusted for best visibility of the micro-calcifications. Consequently, regions with a less dense tissue have been suppressed and are depicted in dark areas.

shapes and boundaries. Guideline for assessing the risk of malignancy on the basis of radiological appearance are given in Ref. [4]. From a diagnostic point of view, the most important feature of a mass is the morphology of its boundary. In particular, a close examination of a mass border allows one to assess the probability whether the mass is a malignant tumour. Masses with well-defined, sharp borders are usually benign. An example of such a mass is depicted in Fig. 5. In case of masses with lobulated shapes, as in Fig. 6, the risk of malignancy increases. However, the most suspicious are the masses with ill-defined or spiculated margins. An example of a spiculated mass is pictured in Fig. 7. Such an appearance of a mass may indicate a malignant, infiltrating tumour.

Apart from micro-calcifications and anomalous masses, one can also recognize two other kinds of abnormalities in mammograms. First, the mammogram may reveal a structural distortion, i.e., a situation in which tissue in a breast region appears to be pulled towards its centre. Second, the left and right breasts may appear significantly different. This situation is referred to as an asymmetric density. These two types of breast abnormalities can

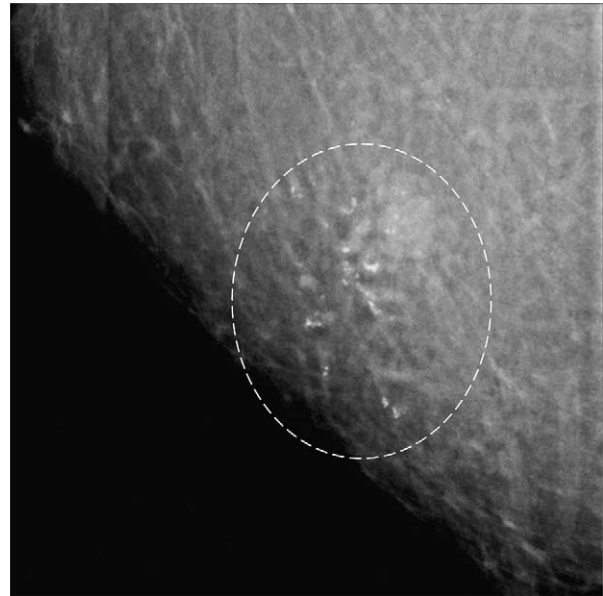


Fig. 4 A mammographic image depicting the breast region with an area of $5\text{ cm} \times 5\text{ cm}$. On the image a cluster of micro-calcifications has been marked, which are highly heterogenous in both size and shape. This is referred to as pleomorphic calcifications. Furthermore, the cluster is isolated, i.e. no micro-calcifications can be identified in the surrounding tissue. This two observations suggest a high risk of breast cancer. The contrast on the image has been manually adjusted for best visibility. Regions of a less dense tissue has been suppressed and are depicted as a black area.

rarely be targeted by the commonly used CAD systems.

2. Background

The recognition of suspicious abnormalities in digital mammograms still remains a difficult task. There are at least several reasons for this situation. First, mammography provides relatively low contrast images, especially in case of dense breasts, commonly found in young women. Second, symptoms of the presence of abnormal tissue may remain quite subtle. For example, spiculated masses that may indicate a malignancy are often difficult to detect, especially at an early stage of development. Important abnormality markers, the micro-calcification clusters, are easier to detect. However, in both cases one has to decide, with a significant level of uncertainty, whether the detected lesion is benign or malignant. For these reasons, we need robust algorithms for enhancing mammogram contrast, segmentation, detection of micro-calcifications and malignancy assessment.

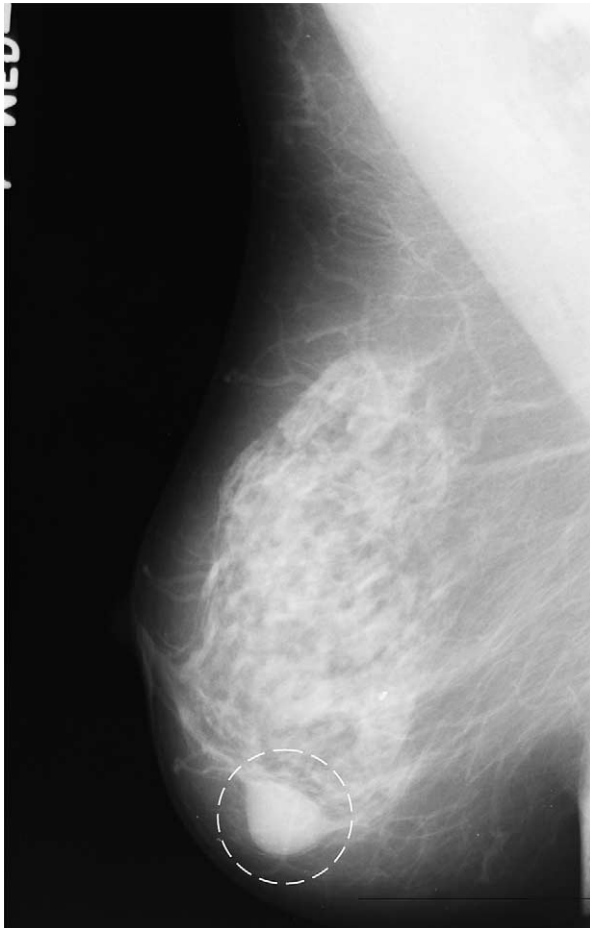


Fig. 5 A mammogram of a left breast in the mediolateral oblique (MLO) projection. On this image an anomalous mass has been marked. This mass is circumscribed and has a well-defined, sharp border. Furthermore, no other masses can be identified within the breast. These findings suggest that this lesion is benign. The image has been taken from the *mini-MIAS* database of mammograms.

2.1. State-of-the-art

Various techniques have been proposed in the literature for enhancing the contrast of digital mammograms. These include: techniques based on fractals [5], wavelet transform [6], homogeneity measures [7], and others. The segmentation of microcalcifications has been done using e.g. morphological filters [8], multiresolutional analysis [9], and fuzzy logic [10]. Furthermore, for detecting microcalcification and malignancy assessment, several classification algorithms have been used. These include: neural networks [11], nearest neighbour classifier [12], multiple expert systems [13], and support vector machines [14]. A more detailed survey on the techniques used in the automatic analysis of digital mammograms can be found in [15].

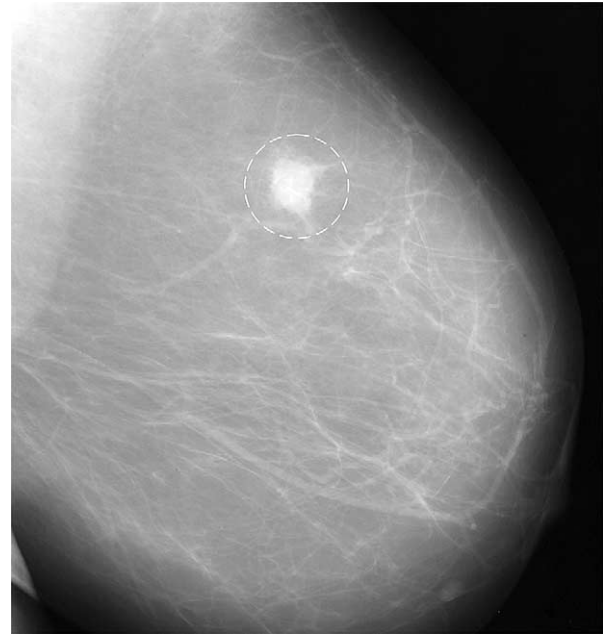


Fig. 6 A mammogram of a right breast in the MLO projection with an anomalous mass marked. The border of this mass is not sharp. It is to some degree lobulated. Such ill-defined or lobulated masses are of more serious concern than those with sharp borders, especially if the lobulations are large and there are many of them. This image was obtained from the *mini-MIAS* database.

Recently a number of information technological initiatives have been undertaken, which are devoted to the development of CAD systems and digital mammography. A good example is the *National Digital Mammography Archive* project¹ [16,17]. This is a collaborative effort between the University of Pennsylvania Medical Center, University of Chicago Department of Radiology, University of North Carolina - Chapel Hill School of Medicine Department of Radiology - Breast Imaging, Sunybrook and Women's College Health Sciences Centre of the University of Toronto, and Advanced Computing Technologies Division of BWXT Y-12 L.L.C. in Oak Ridge Tennessee. Its aim is to develop a national archive for breast imaging, which is available over the net. Furthermore, a national network and cyber-infrastructure devoted to digital mammography will be created because of the data deluge to be expected from the enormous number of high-resolution mammograms at an annual basis. A *Digital Database for Screening Mammography* has been created at the University of South Florida [18]. This voluminous database provides high-resolution digitised mammograms for developing easy-to-use cancer detection algorithms and

¹ <http://nscp.upenn.edu/NDMA>

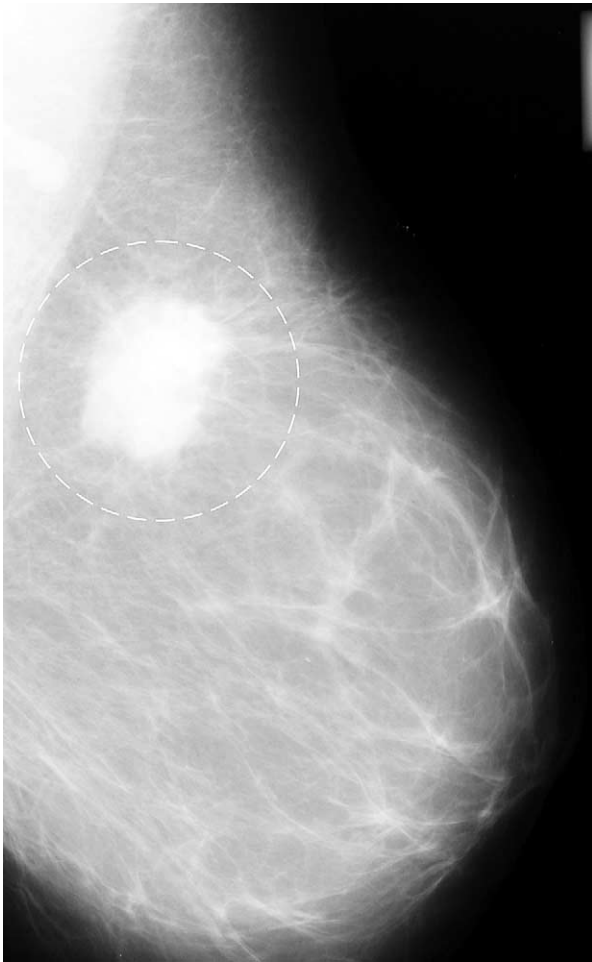


Fig. 7 A mammogram of a right breast in the MLO projection on which a mass has been marked. This example is referred to as spiculated mass and usually indicates the presence of a malignant, invasive tumour. Image from the *mini-MIAS* database.

enabling the comparative analysis of detection accuracy. Another database of digital mammograms, i.e. *mini-MIAS*, is provided by The Mammographic Image Analysis Society [19]. Devising an efficient system for handling many mammograms, with up to petabytes of accumulated data, is indeed a challenging task in computer science and information technology.

2.2. Motivation

In Ref. [20] we have used two algorithms for face detection of images, i.e., AdaBoost with simple rectangular features [21] and SVM with log-polar sampling grid [22]. Our goal is to verify whether these algorithms are suitable for distinguishing between normal and abnormal regions of breast images. However, the problem of selecting the sus-

picious regions from the entire area of the breast is beyond the scope of this study. We note that the algorithms have been adapted directly to this domain, i.e., only some parameters have been changed but the image feature extraction methods and the classifiers were the same as in [20].

The algorithms have been evaluated on the DDSM [18] database of mammogram images. This database consists of a large set of breast images in both MLO, i.e., medio-lateral and CC, i.e., cranio-caudal projections. For each patient, four images are present, including the above types of projections for the left and right breast. From the database, we have chosen the BCRP_MASS_0 and BCRP_CALC_0 subsets, including cases of malignant masses and microcalcifications, respectively. From these images, we have obtained 168 samples of breast regions with lesions and 1017 samples of breast regions without any lesion. These samples were used in the training of the classifiers and evaluating of classification accuracy.

3. Design considerations

The design of the microcalcification detection system is based on the learning, classifier-based methods [23]. The general scheme is presented below in Algorithm 1. These methods consist of several steps. First, a set of rectangular regions is selected from each of the available mammogram images. The selected regions are then partitioned into training and testing samples. Afterwards, a set of features is extracted from each of the samples. The features from training samples are used to train an algorithm capable of making a binary decision on a given image features.

After the classifier has been trained, the image features from the testing set are used to evaluate its effectiveness. In particular, a 2×2 confusion matrix [23] can be computed: $A_{i,j}$. The element $a_{i,j}$ of this matrix represents the number of samples belonging to class i , provided that the classification algorithm decides that it belongs to class j , ($i, j \in \{1 - \text{lesion}, 0 - \text{nonlesion}\}$). Since the lesion can be treated as the positive diagnosis, while lack of lesion in the sample as negative diagnosis, the confusion matrix can lead to four values quantifying the behaviour of the classifier. These are: true positive ratio specified as the $\frac{a_{1,1}}{(a_{1,0}+a_{1,1})}$, true negative ratio $\frac{a_{0,0}}{(a_{0,0}+a_{0,1})}$, false positive ratio $\frac{a_{0,1}}{(a_{0,0}+a_{0,1})}$ and false negative ratio $\frac{a_{1,0}}{(a_{1,0}+a_{1,1})}$. The true positive ratio is also referred to as sensitivity of the classifier, while the true negative as

```

INPUT:
D - Database of grey-scale, 16bit breast image regions
C - a function that represents a classification algorithm that is to be used to
    detect the micro-calcifications.
EVALUATE (C,D)
  foreach I in D
    Select Square Regions From I
    Partition the selected regions into training set TrainSet and testing set TestSet
  Learning:
    Generate Feature Vectors FTrain for TrainSet
    Learn classifier C on FTrain
  Validation:
    Generate Feature Vectors FTest for TestSet
    foreach v in FTest
      Classify v using C
    Calculate classification accuracy and the confusion matrix

```

Algorithm 1 Detection scheme for suspicious lesions

specificity. Finally, the overall accuracy, or the success rate of the classifier can be quantified as $\frac{a_{1,1}+a_{0,0}}{(a_{0,0}+a_{0,1}+a_{1,0}+a_{1,1})}$. Ideally, the confusion matrix should have values near 1.0 on the main diagonal, i.e., for true positives and negatives and values near 0 along the second diagonal, i.e., for false positives and negatives. These situations indicate that the classification algorithm is capable of properly discriminating between regions of mammograms with lesions or no lesions.

4. System description

As already noted, two classifier-based algorithms are presented and evaluated in this paper. These are the systems based on SVM and on AdaBoost classifiers.

4.1. Treatment of input images

From the DDSM database we have obtained 168 image samples with lesions and 1017 without lesions. In order to extract the samples from the original full breast DDSM images, we have used the information on the outline of the lesion. The bounding rectangle with margins of 100 pixel was extracted from each mammogram that contained a lesion. Afterwards, we have estimated the centre of the lesion. Next, each rectangle was cropped to a largest possible square region, with lesion in the centre. Finally, we estimated the diameters of the lesions. These transformations produced a set of square regions of different sizes, each one containing a lesion in the centre.

The above procedure allowed us to obtain the lesion samples. For gathering the non-lesion cases, we have used the following procedure. In most patients, the cancerous lesion was present in only one

of the two breasts. Therefore, we captured rectangular regions from the non-lesion breast. To make the non-lesion and lesion samples most informative, i.e., differing only in the presence of the lesions and non with e.g. tissue type, we have used the same region within the non-cancer breast as in the cancer breast.

Typically the ratio of non-lesion samples is much greater than that of lesion samples. Thus, we have used as additional non-lesion samples the rectangular regions adjacent to the lesion region from the top, right, bottom and left, provided they do not contain another lesion. The same procedure was carried out for the non-lesion samples in the non-lesion breast. Afterwards we discarded regions containing visible artefacts, e.g. mammogram border.

The resulting 168 abnormal, lesion samples, including 91 microcalcifications and 77 masses, and 1017 normal, non-lesion samples were divided into two sets:

- training set of 385 samples, including 297 normal samples and 88 abnormal samples (51 microcalcifications and 37 masses),
- test set of 800 samples, including 720 normal and 80 abnormal samples (40 calcifications and 40 masses).

The number of lesions in the testing set was chosen to be 10% of the total number of samples in the test set. The exact ratio of normal and abnormal samples used in training depends on the classifier. In some cases, not all non-lesion samples were used in training.

Both of the classifiers operate on a square input of fixed width. Since the extracted regions can be significantly larger in diameter than this expected input, we employ the wavelet approximation of the input samples. There are two scenarios for choosing the level of this approximation:

Fixed scaling - each sample is downsampled, using the Daubechies-4 wavelet [24] a fixed number of times. Then, the central square of a required size is extracted. The diameter of the change is not used.

Variable scaling - each sample is downsampled, using the Daubechies-4 wavelet. However, the number of successive approximations depends on the diameter of the change. It is chosen so that the diameter of the change is always smaller than the classifier window width but larger than a half of it. Then the central region of the downsampled sample is extracted.

For images of normal breasts, the diameter of the change is not valid, since no change is detected. Therefore, in the first scenario, we use the constant level of wavelet approximation. In the second scenario, the level of approximation is randomly selected from the range of scales found in the samples that contain the cancerous change. Moreover, the distribution of the scales for the abnormal samples.

The AdaBoost method uses additional filtering of input images to increase the accuracy. As the SVM classifier uses responses of Gabor filters as an input features, additional filtering is inappropriate for this method.

Finally, we use AdaBoost or SVM for classifying image windows of fixed width. Therefore, we need an efficient method of moving the window on a mammogram. The classifier determines whether a cancerous change is present in each of the windows or not. However, the method for conducting an effective search through the image is out of scope of this paper, as we will focus on the classification task.

4.2. Detection of abnormalities using support vector classification and log-polar sampling of the Gabor decomposition

4.2.1. Support vector classification

The support vector classification (SVC) was first proposed for optical character recognition in [25]. New algorithms were also proposed for regression estimation [26], novelty detection [27], operator inversion [28] and other problems.

The SVC algorithm separates the classes of input patterns with the maximal margin hyperplane. The hyperplane is constructed as:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \tag{1}$$

where: \mathbf{x} is the feature vector, \mathbf{w} is the vector that is perpendicular to the hyperplane, and $\frac{b}{\|\mathbf{w}\|}$ speci-

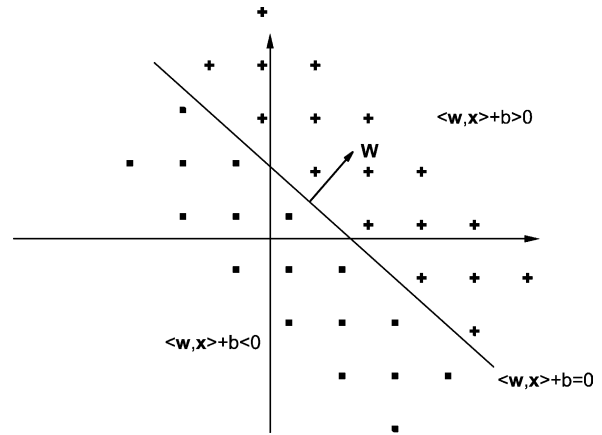


Fig. 8 A hyperplane that separates two classes of vectors. In dot-product space the hyperplane can be specified as a set of vectors \mathbf{x} that satisfies $\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$. The vector \mathbf{w} is perpendicular to the hyperplane, whereas scalar $\frac{b}{\|\mathbf{w}\|}$ specifies the offset from the beginning of the coordinate system. The hyperplane depicted on the figure is the one that maximizes the margin to the separated vectors.

fies the offset from the beginning of the coordinate system. An example of the maximal margin hyperplane given by this form is depicted in Fig. 8. In order to allow for a construction of non-linear decision boundaries, the separation is performed in a *feature space* \mathcal{F} , which is introduced by a nonlinear mapping Φ of the input patterns. As the hyperplane construction involves the computation of the inner products on the *feature space*, this mapping Φ must satisfy:

$$\langle \Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2) \rangle = k(\mathbf{x}_1, \mathbf{x}_2) \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X} \tag{2}$$

for some *kernel function* $k(\cdot, \cdot)$. The kernel function represents the non-linear transformation of the original feature space into the \mathcal{F} . Here, $\mathbf{x}_i \in \mathcal{X}$ denotes the input patterns. Maximizing the distance of this separation is equivalent to the minimization of the norm-squared: $\frac{1}{2} \|\mathbf{w}\|^2$ [29]. However, to guarantee that the resultant hyperplane separates the classes, the following constraints must be satisfied:

$$y_i \cdot (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, n \tag{3}$$

where $y_i \in \{-1, 1\}$ denotes the class label corresponding to the input pattern \mathbf{x}_i . These constraints do not impose a strict class separation. Instead, we utilized slack variables ξ_i to allow for the training of the classifier on linearly non-separable classes. The slack variables must be penalized in the minimization term. Consequently, learning of the SVC classifier is equivalent to solving a minimization problem

with the objective function of the form:

$$\min_{\mathbf{w} \in \mathcal{X}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \quad (4)$$

$$\xi \in \mathbb{R}^n$$

and the constraints are given by Eq. (3). The parameter C controls the penalty for misclassifying the training samples (see Ref. [20]). Using the Lagrange multiplier technique, we can transform this optimization problem to a dual form:

$$\min_{\alpha \in \mathbb{R}^n} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \cdot k(\mathbf{x}_i, \mathbf{x}_j)$$

subject to :

$$0 \leq \alpha_i \leq C \quad (5)$$

$$\sum_{i=1}^n \alpha_i y_i = 0$$

In the above formulation, the $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ is the vector of Lagrange multipliers. Furthermore, the feature space dot-products between input patterns are computed by using a kernel function $k(\cdot, \cdot)$. This is possible, as the mapping ϕ must satisfy Eq. (2). The Lagrange multipliers that solves the Eq. (5) can be used to compute the decision function:

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b \quad (6)$$

where:

$$b = y_i - \sum_{j=1}^n \alpha_j y_j k(\mathbf{x}_j, \mathbf{x}_i) \quad (7)$$

The solution to Eq. (5) can be found using any general purpose quadratic programming solver. Furthermore, dedicated heuristic methods have been developed that can solve large problems efficiently [30–32].

Finally, we note that the optimization problem from Eq. (5) is convex [33]. Therefore, SVC has an important advantage over neural networks, where the existence of many local minima makes the learning process rather complex, which often leads to a poor classification. Training of the support vector classifier is also much more insensitive to overfitting than the classic feed-forward neural network.

4.2.2. Log-polar sampling of the Gabor decomposition

The Gabor filters are very useful for extracting feature vectors from images. Originally, the filters

have been proposed in [34] as a Gabor elementary function and afterwards extended in [35] to two-dimensional image operators. The two-dimensional Gabor filter is defined as:

$$G(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-1/2(x^2/\sigma_x^2 + y^2/\sigma_y^2)} e^{i2\pi\nu_0 x} \quad (8)$$

This filter is a plane sinusoidal wave modulated by a Gaussian and is sensitive to the image details that within the Fourier plane corresponds to the frequencies near ν_0 . The σ_x and σ_y parameters are widths of the Gaussian function along the x- and y-axis respectively. As the wave vector of this filter is parallel to the x axis, the filter is sensitive to the vertical image details. However, for constructing a filter sensitive to image details with some orientation angle $\phi \neq 0$, it is sufficient to rotate the original filter from Eq. (8). In [22] the modified Gabor filters are employed, which are cast in log-polar coordinates:

$$\hat{G}(r, \phi) = A e^{-(r-r_0)^2/2\sigma_r^2} e^{-(\phi-\phi_0)^2/2\sigma_\phi^2} \quad (9)$$

where:

$$r = \log \sqrt{\nu^2 + \mu^2},$$

$$\phi = \arctan \frac{\mu}{\nu} \quad (10)$$

The σ_r and σ_ϕ parameters controls the radius-axis width and angle-axis width of the Gaussian function.

This approach is useful in construction of filter banks with different orientations ϕ and central frequencies ν_0 . In particular, the filters defined by Eq. (9) do not overlap at low frequencies, whereas the construction based on Eq. (8) requires a careful selection of σ_x and σ_y values for filters with small ν_0 frequency.

Finally, in [22] a log-polar, spatial grid has been proposed to sample the responses of a bank of Gabor filters. It consists of points arranged in several circles with logarithmically spaced radii (see Fig. 9). For computing the feature vector for a given image point X , the image is filtered with a bank of modified Gabor filters and magnitudes of the responses are sampled with the grid centred at X .

4.3. Boosting method for detecting abnormalities

Unlike neural networks [11], the boosting method is based on the idea of combining multiple classifiers into a single, but much more reliable, classifier. The set of weak classifiers that contribute to the final answer can be simple and erroneous to some ex-

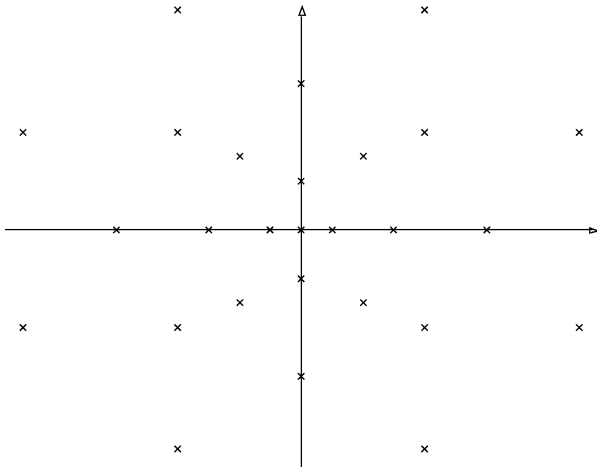


Fig. 9 A grid used to sample the responses of a bank of Gabor filters, as proposed in [22]. To sample the responses for a given point p , the grid is centred in p . Afterwards, the magnitudes of the complex responses of the bank filter are computed at each grid point, the final result is a vector, whose coordinates are the magnitudes of filter responses collected from all grid points in a pre-defined order.

tent. However, a scheme for training them has been devised in order to have a small error in the final classification. For additional information concerning the method of boosting, the reader is urged to consult e.g. Ref. [36] or Ref. [37].

From the many boosting algorithms we have favored the AdaBoost classifier [38]. The pseudocode of this classifier is presented below. This classifier is used also in Ref. [21]. In the training phase, each sample vector from the training set is weighted. Initially, the weights are uniform for all the vectors. Then, at each iteration, a weak classifier is trained to minimize a weighted error for the samples. Each

iteration changes the weights values reducing them by an amount, which depends on the error of the weak classifier, which depends on the error of the weak classifier on the entire training set. However, this reduction is made only for the examples that were correctly classified by the classifier trained in the current iteration. The weight of the weak classifier within the whole ensemble is also connected to its training error.

Assigning non-uniform, time-varying weights to the training vectors is crucial for minimizing the error rate of the final, aggregated classifier. During training, the ability of the classifier to classify training set correctly is constantly increased. The reason is that weak classifiers used by AdaBoost are complementary. Thus the samples vectors that were misclassified by some weak classifiers are classified correctly by the other weak classifiers.

The process of training the classifier is summarized in the form of Algorithm 2. In particular, in each round t of the total T rounds, a weak classifier h_t is trained on the training set Tr with weights D_t . The training set is formed by examples from domain X labelled with labels from a set C . The training of the weak classifier is left to an unspecified *WeakLearner* algorithm, which should minimize the training error ε_t of the produced weak classifier with respect to the weights D_t .

Based on the error ε_t of the weak classifier h_t , the parameters α_t and β_t are calculated. The first of the parameters defines the weight of h_t in the final, combined classifier. The second provides a multiplicative constant, which is used to reduce the weights $\{D_{t+1}(i)\}$ of the correctly classified examples $\{i\}$. The weights of the examples that were misclassified are not changed. Thus, after normalizing the new weights $\{D_{t+1}(i)\}$, the relative weights of the misclassified examples from the

```

INPUT:
  Labelled training set  $Tr$  of size  $m$ ,  $Tr = \{(x_1, c_1), \dots, (x_m, c_m)\} \subset X \times C$ 
  Algorithm WeakLearner producing a hypothesis  $h_t : X \rightarrow C$ 
  An integer number  $T$  - maximal number of training rounds
ADABOOST( $Tr$ , WeakLearner,  $T$ )
   $\forall_{1 \leq i \leq m} D_1(i) = \frac{1}{m}$ 
  for  $t = 1$  to  $T$ 
     $h_t = \text{WeakLearner}(Tr, D_t)$ 
     $\varepsilon_t = \sum_{i: h_t(x_i) \neq c_i} D_t(i)$ 
    if  $\varepsilon_t = 0$  or  $\varepsilon_t \geq 0.5$ 
       $T = t$ 
    exit loop
   $\beta_t = \frac{\varepsilon_t}{1 - \varepsilon_t}$ 
   $\alpha_t = \log \frac{1}{\beta_t}$ 
   $\forall_{i: h_t(x_i) = c_i} D_{t+1}(i) = \beta_t D_t(i)$ 
   $\forall_{i: h_t(x_i) \neq c_i} D_{t+1}(i) = D_t(i)$ 
  normalize distribution  $D_{t+1}$ 
  return  $h_{\text{fin}} = \arg \max_c \sum_{t: h_t(x) = c} \alpha_t$ 
    
```

Algorithm 2 The AdaBoost algorithm

training set are increased. Therefore, in the h_{t+1} round, the *WeakLearner* is more focused on these examples. In this way we have enhanced the chance that the classifier h_{t+1} will learn to classify them correctly.

The final, strong classifier h_{fin} employs a weighted voting scheme over the results of the weak classifiers h_t . The weights of the individual classifiers are defined by the constants α_t .

There are two special cases, which are treated individually during the algorithm execution. One is the case of ε_t equal to zero. In this case, the weights D_{t+1} would be equal to D_t , and h_{t+1} to h_t . Therefore, the algorithm does not go further ahead with training. The second case is of $\varepsilon_t \geq 0.5$. In this case, the theoretical constraints on h_t are not satisfied, and the algorithm cannot continue with new rounds of training.

One of the most important issues in using the AdaBoost scheme is the choice of the weak classifier that separates the examples into the two classes to be discriminated. Following [21], a classifier that selects a single feature from the entire feature vector is used. The training of the weak classifier consists of selecting the best feature and of choosing a threshold value for this feature, which optimally separates the examples belonging one class from examples belonging to the other class. The selection involves minimizing the weighted error for the training set. The feature set consists of the features, which are computed as differences of the sum of pixels intensities inside two, three or four adjacent rectangles. These rectangles are of various sizes and positions within the image window, as long as their contiguity is maintained. The classifier operates on an image window of the size 24×24 pixels.

4.4. Implementation note

The system was implemented in the Matlab version 6.5 environment. For the SVM classifier we used OSU SVM toolbox version 3.0. The AdaBoost has been implemented by the authors in Matlab with some functions implemented as external C libraries. For scaling of the images, we used Matlab Wavelet Toolbox. Image filtering was carried out using the Matlab Image Processing Toolbox. The images were stored in uncompressed, 16 bit per pixel, gray-scale TIFF files. The tests were carried out on the Sun Microsystems SunBlade 2000 machine running SunOS 5.9 operating system. As the tests did not require any manual evaluation of images, no special presentation or user navigation tools were developed.

5. Status report

5.1. Evaluation of the algorithm based on the SVM classifier

In this section we evaluate the method based on the SVM algorithm previously used for face detection in [20]. The evaluation was carried out on the set of images described in Section 4.1.

5.1.1. Parameters of the log-polar sampling grid and the bank of Gabor Filters

The image feature vectors were extracted using log-polar sampling grid composed of 51 points. These points were arranged in 6 circles with the radii spaced logarithmically between 5 and 40 points. The bank of Gabor filters used with the sampling grid consisted of 20 filters with a size of 85×85 points. The filters were arranged into 4 logarithmically spaced frequency channels and 5 uniformly spaced orientation channels. The lowest normalized frequency presented in the filter bank was $\frac{1}{7}$, whereas the highest was $\frac{1}{2}$. The orientation channels cover the entire spectrum, i.e., from 0 to $\frac{4}{5}\pi$ radian.

5.1.2. Training phase and the results of tests

The SVM classifier was evaluated independently for following three kernel functions:

- linear kernel: $k(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle$
- polynomial kernel of order 4: $k(\mathbf{x}, \mathbf{y}) = (\sigma \langle \mathbf{x}, \mathbf{y} \rangle + \gamma)^4$
- Gaussian RBF kernel: $k(\mathbf{x}, \mathbf{y}) = e^{-\|\mathbf{x}-\mathbf{y}\|/2\sigma^2}$

In order to select reasonable values for the SVM misclassification penalty and parameters of kernel functions, we performed a parameter study using training set with 1-level fixed scaling. For each of the kernel functions, we evaluated the specificity of the classifier on the training set using the following values for the misclassification penalty: $C \in \{1, 2, \dots, 50\}$. The parameters σ and γ were set to 1.0 in these tests. Afterwards, for each of the kernel functions we selected the value of misclassification penalty that yielded the highest sensitivity. These values were used in all further tests. Using the same approach we selected the values for the parameters σ and γ . For both parameters we evaluated the following range of values: $\sigma, \gamma \in \{0.1, 0.2, \dots, 10.0\}$. The results of the parameter study are summarized in Table 1.

To obtain the overall classification accuracy, sensitivity and specificity, we have validated the classifier on the test set. The results for various

Table 1 Values for parameters of kernel functions and misclassification penalties used in the SVM classifier

Kernel function	Parameters of the kernel function	Misclassification penalty
Linear	-	$C = 9.0$
Polynomial	$\sigma = 0.7 \gamma = 0.5$	$C = 3.0$
Gaussian RBF	$\sigma = 0.9$	$C = 7.0$

combinations of the kernel function and wavelet scaling mode are presented in Table 2.

In the tests, the highest sensitivity was achieved when using 2-level scaling and Gaussian RBF kernel. The highest specificity and accuracy was achieved with variable scaling. Also in this case the Gaussian RBF kernel performs best.

To evaluate the performance of the SVM classifier for recognizing particular types of lesions, two additional tests were performed. In the first test the training and testing sets consisted of normal samples and microcalcifications. In the second test we used training and testing set composed of normal samples and masses. The results of these tests are presented in Tables 3 and 4 respectively.

The SVM classifier has significantly higher sensitivity to microcalcifications than to masses. The

specificity and accuracy is, in general, slightly higher for masses. However, the differences in these two performance measures are small. Therefore we can conclude, that the algorithm performs better for microcalcification than for masses. The best sensitivity to microcalcifications was obtained when using 2-level fixed wavelet scaling with Gaussian RBF kernel. The best specificity and overall accuracy was obtained with 1-level scaling and linear kernel. For masses, the highest specificity, sensitivity and accuracy was obtained when using variable scaling and Gaussian RBF kernel function.

5.2. Evaluation of the AdaBoost-based classification algorithm

In this section we evaluate the method based on the AdaBoost algorithm used for face detection in [20]. The AdaBoost classifier was trained on the same database as the SVM. The algorithm used central, rectangular part of the images with size equal to 24×24 pixels.

5.2.1. Input image filtering

Before we apply the wavelet scaling, the image is filtered with various filters. The filters below rep-

Table 2 Results for SVM classifier on the testing set of 80 lesion and 720 non-lesion breast samples

Scaling type	Kernel function	Sensitivity (%)	Specificity (%)	Accuracy (%)
Fixed - 1	Linear	52.5	74.7	72.5
Fixed - 1	Polynomial	57.5	76.1	74.2
Fixed - 1	Gaussian RBF	56.2	75.7	73.8
Fixed - 2	Linear	62.5	75.3	74.0
Fixed - 2	Polynomial	67.5	75.0	74.2
Fixed - 2	Gaussian RBF	68.8	75.4	74.8
Variable	Linear	57.5	82.1	80.1
Variable	Polynomial	55.0	82.5	81.0
Variable	Gaussian RBF	57.5	83.3	82.0

Table 3 Results for SVM classifier on the testing set of 40 microcalcifications and 720 non-lesion breast samples

Scaling type	Kernel function	Sensitivity (%)	Specificity (%)	Accuracy (%)
Fixed - 1	Linear	47.5	83.3	81.4
Fixed - 1	Polynomial	55.0	81.2	79.9
Fixed - 1	Gaussian RBF	55.5	81.5	80.1
Fixed - 2	Linear	67.5	77.8	77.2
Fixed - 2	Polynomial	72.5	76.4	76.2
Fixed - 2	Gaussian RBF	72.5	76.9	76.7
Variable	Linear	67.5	73.3	73.0
Variable	Polynomial	65.0	77.9	77.2
Variable	Gaussian RBF	65.0	79.4	78.7

Table 4 Results for SVM classifier on the testing set of 40 masses and 720 non-lesion breast samples

Scaling type	Kernel function	Sensitivity (%)	Specificity (%)	Accuracy (%)
Fixed - 1	Linear	50.0	81.2	79.6
Fixed - 1	Polynomial	50.0	82.5	80.8
Fixed - 1	Gaussian RBF	50.0	82.1	80.4
Fixed - 2	Linear	52.5	78.5	77.1
Fixed - 2	Polynomial	55.0	78.5	77.2
Fixed - 2	Gaussian RBF	55.0	78.5	77.2
Variable	Linear	57.5	82.8	80.1
Variable	Polynomial	55.0	82.5	81.0
Variable	Gaussian RBF	57.5	83.3	82.0

resent a group of typical basic filters used in image processing. The following filters [39] are used:

No filtering - No filtering

Unsharp - Unsharp contrast enhancement filter, i.e., negation of the Laplacian

Sobel - Sobel horizontal filter

Laplacian - Laplacian filter

LoG - Laplacian of Gaussian filter

Dilate - Greyscale dilation using disk structuring element

5.2.2. Classifier training and results of the tests

After filtering the wavelet scaling is done and the classifier is trained. The number of rounds of the classifier is set to 200. For each different filtering type and wavelet scaling mode, a different

classifier is trained. For each configuration, an overall classification accuracy on the testing set, as well as the sensitivity and specificity of the trained classifier is given in Table 5. In the test, the Fixed scaling with no filtering achieved the highest results. The highest sensitivity is achieved for 4-level scaling. For sensitivity, four levels of scaling are the best. For overall classifier accuracy, three levels of wavelet scaling yield better results.

To find out how each type of the lesion contributed to these results, we have evaluated the classifier for a training and testing sets including either only microcalcifications and normal samples, or, in a second scenario, only masses and normal samples. The results are given in Tables 6 and 7 respectively.

Table 5 Results for AdaBoost classifier in various configurations on the testing set of 80 lesion and 720 non-lesion breast samples

Scaling type	Filtering type	Sensitivity (%)	Specificity (%)	Accuracy (%)
Fixed - 3	No filtering	73.8	77.2	76.9
Fixed - 3	Unsharp	77.5	74.6	74.9
Fixed - 3	Sobel	60.0	74.0	72.6
Fixed - 3	Laplacian	48.8	62.5	61.1
Fixed - 3	LoG	51.2	70.7	68.8
Fixed - 3	Dilate	77.5	75.8	76.0
Fixed - 4	No filtering	82.5	73.8	74.6
Fixed - 4	Unsharp	80.0	74.2	74.8
Fixed - 4	Sobel	70.0	74.0	73.6
Fixed - 4	Laplacian	41.2	78.2	74.5
Fixed - 4	LoG	53.8	78.9	76.4
Fixed - 4	Dilate	75.0	77.9	77.6
Variable	No filtering	68.8	68.5	68.5
Variable	Unsharp	66.2	72.2	71.6
Variable	Sobel	61.2	71.9	70.9
Variable	Laplacian	50.0	65.3	63.8
Variable	LoG	52.5	66.5	65.1
Variable	Dilate	72.5	67.2	67.8

For fixed wavelet scaling, the level of wavelet approximation is specified.

Table 6 Results for AdaBoost classifier on microcalcifications only for various configurations on the testing set of 40 microcalcifications and 720 non-lesion breast samples

Scaling type	Filtering type	Sensitivity (%)	Specificity (%)	Accuracy (%)
Fixed - 3	No filtering	62.5	79.4	78.6
Fixed - 3	Unsharp	65.0	73.5	73.0
Fixed - 3	Sobel	62.5	65.0	64.9
Fixed - 3	Laplacian	57.5	68.2	67.6
Fixed - 3	LoG	55.0	61.5	61.2
Fixed - 3	Dilate	70.0	76.7	76.3
Fixed - 4	No filtering	72.5	65.4	65.8
Fixed - 4	Unsharp	70.0	67.4	67.5
Fixed - 4	Sobel	67.5	69.7	69.6
Fixed - 4	Laplacian	75.0	65.6	66.1
Fixed - 4	LoG	65.0	66.5	66.5
Fixed - 4	Dilate	75.0	63.2	63.8
Variable	No filtering	60.0	54.6	54.9
Variable	Unsharp	40.0	58.2	57.2
Variable	Sobel	47.5	57.6	57.1
Variable	Laplacian	45.0	74.9	73.3
Variable	LoG	35.0	71.8	69.9
Variable	Dilate	65.0	54.2	54.7

For fixed wavelet scaling, the level of wavelet approximation is specified.

In calcification detection, the dilation of the image resulted in some improvement of the results. In particular, it increased the sensitivity, as a cost of slight decrease in specificity in all three scaling configurations.

The AdaBoost classifier obtained significantly better results for masses than for microcalcifica-

tions. This can be attributed to the nature of the features used. The features are taken directly from the face recognition domain, in which the recognized object is of similar size and shape. This type of object is more similar to masses than to microcalcification clusters, which are highly non-uniform and not as well localized.

Table 7 Results for masses only for AdaBoost classifier in various configurations on the testing set of 40 masses and 720 non-lesion breast samples

Scaling type	Filtering type	Sensitivity (%)	Specificity (%)	Accuracy (%)
Fixed - 3	No filtering	80.0	87.9	87.5
Fixed - 3	Unsharp	75.0	85.7	85.1
Fixed - 3	Sobel	60.0	80.4	79.3
Fixed - 3	Laplacian	37.5	69.0	67.4
Fixed - 3	LoG	42.5	65.4	64.2
Fixed - 3	Dilate	75.0	88.9	88.2
Fixed - 4	No filtering	75.0	89.3	88.6
Fixed - 4	Unsharp	82.5	90.0	89.6
Fixed - 4	Sobel	75.0	87.8	87.1
Fixed - 4	Laplacian	47.5	77.5	75.9
Fixed - 4	LoG	52.5	82.8	81.2
Fixed - 4	Dilate	85.0	85.7	85.7
Variable	No filtering	92.5	88.5	88.7
Variable	Unsharp	87.5	89.7	89.6
Variable	Sobel	70.0	86.9	86.1
Variable	Laplacian	72.5	63.6	64.1
Variable	LoG	80.0	62.5	63.4
Variable	Dilate	85.0	90.3	90.0

For fixed wavelet scaling, the level of wavelet approximation is specified.

6. Lessons learned and future plans

In case of the SVM-based approach, our results suggest that the algorithm fails to identify the breast image features that can be used to indicate the presence of an abnormal tissue. The overall sensitivity to abnormal breast regions is below 70%. A slightly better result was obtained when focussing only on microcalcifications. Here the highest sensitivity was equal to 72.5%. On the other hand, the sensitivity for masses only is below 60%. The specificity of the method is higher and reaches 83.3% in all three types of tests.

The most probable reason behind this behaviour is that the image feature extraction method is not suitable for classifying mammogram images. The log-polar sampling grid was proposed to detect facial landmarks, e.g. eyes, in face images [20]. Feature extraction methods, based on Gabor filtering, are appropriate for these tasks. However, in mammogram image classification more localized features are necessary as indicators of abnormality. Such features would be more adequate for detection of borders of masses or detection of small microcalcifications.

Results for the AdaBoost-based approach are promising. The classifier accuracy and sensitivity reaches 90% for masses and decreases to about 76% accuracy and sensitivity for all lesion types. This decrease is caused by inability of the classifier to recognize microcalcifications. The accuracy for only microcalcification lesion type is also 78%, however, the sensitivity is only 70%. The introduction of various filters does not lead to significant changes in the recognition accuracy.

These results can be treated as the lowest values we can expect from a mammogram detection algorithm. The tested algorithms were not originally created with cancer detection in mind. They had been directly applied from the face recognition field, which could be considered as one of the benchmark problems in the image processing and recognition field. Therefore, we emphasize here that any algorithm dedicated for the cancer detection should achieve at least this degree of accuracy, for it to be considered worthwhile.

When there are many candidate regions, our results cannot precisely identify the abnormal lesions in the breast image. In this case, our algorithm would select a significant number of normal tissue samples. Our study also suggests that the image features used are the major limitation for any further increase of the recognition accuracy.

As already noted, the problem of selecting the suspicious regions of a mammogram was beyond the scope of this study. Due to the very large size of the

mammograms digitised at high resolution, the classification algorithms are suitable only for the final decision on the presence of an abnormality. However, a fast algorithm is needed that would allow us to select suspicious locations within the image and to discard quickly the uninteresting regions.

In the future we should be focussing on the development of the following procedures based on a local processing of data:

1. Fast and accurate algorithms for the selection of suspicious regions within digitised mammograms,
2. Image feature extraction algorithms tailored to the analysis of digitised mammograms,
3. Filters that can accentuate the abnormal tissue in digitised mammograms, as e.g. in [40].

Acknowledgments

We thank Professor Robert Hollebeck, University of Pennsylvania, for his encouragement and suggestions and Dr. T. Popiela from Department of Radiology, Collegium Medicum, Jagiellonian University, Krakow (Poland) for medical consultations. We thank Ben Holtzmann and Lilli Yang for contributing to Fig. 1. This research has been supported by the Math-Geo program of National Science Foundation and the Digital Technology Center of Univ. Minnesota.

References

- [1] Cancer facts and figures, the American Cancer Society, 2003.
- [2] T. Freer, M. Ulissey, Screening Mammography with Computer-aided Detection: Prospective Study of 12,860 Patients in a Community Breast Center, *Radiology* 220 (3) (2001) 781-786.
- [3] M. LeGal, G. Chavanne, Valeur diagnostique des microcalcifications groupees decouvertes par mammographies, *Bull. Cancer* (71) (1984) 57-64.
- [4] A.C. of Radiology, BI-RADS: Mammography, in: *Breast Imaging Reporting and Data System: BI-RADS Atlas*. 4th ed., American College of Radiology, Reston, Va, 2003.
- [5] H. Li, K. Liu, S. Lo, Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms, *IEEE Trans. Med. Imag.* 16 (1997) 785-798.
- [6] A. Laine, S. Schuler, J. Fan, W. Huda, Mammographic feature enhancement by multiscale analysis, *IEEE Trans. Med. Imag.* 13 (1994) 725-740.
- [7] H. Cheng, W. Jinglia, S. Xiangjuna, Microcalcification detection using fuzzy logic and scale space approaches, *Pattern Rec.* 37 (2) (2004) 363-375.
- [8] D. Zhao, M. Shridhar, D. Daut, Morphology on detection of calcifications in mammograms, *Proc. ICASSP92*, 1992, pp. 129-132.

- [9] T. Netsch, H. Peitgen, Scale-space signatures for the detection of clustered microcalcifications in digital mammograms, *IEEE Trans. Med. Imag.* 18 (9) (1999) 774-786.
- [10] N. Pandey, Z. Salcic, J. Sivaswamy, Fuzzy logic based microcalcification detection, *Proceedings of the IEEE Neural Networks for Signal Processing Workshop X*, 2000, pp. 662-671.
- [11] S. Sehad, S. Desarnaud, A. Strauss, Artificial neural classification of clustered microcalcifications on digitized mammograms, *Proc. IEEE Int. Conf. Syst. Man Cybernet.*, 1997, pp. 4217-4222.
- [12] T. Bhangale, U. Desai, U. Sharma, An unsupervised scheme for detection of microcalcifications on mammograms, *Int. Conf. Image Proc.* vol. 1, 2000, pp. 184-187.
- [13] L. Cordella, F. Tortorella, M. Vento, Combining experts with different features for classifying clustered microcalcifications in mammograms, *International Conference on Pattern Recognition*, vol. 4, 2000.
- [14] I. El-Naqa, Y. Yang, M. Wernick, N. Galatsanos, R. Nishikawa, A support vector machine approach for detection of microcalcifications, *IEEE Trans. Med. Imag.* 21 (12) (2002) 1552-1563.
- [15] H. Cheng, X. Cai, X. Chen, L. Hu, X. Lou, Computer-aided detection and classification of microcalcifications in mammograms: a survey, *Pattern Rec.* 36 (2003) 2967-2991.
- [16] R. Hollebeek, NDMA: Collecting and Organizing a Large Scale Collection of Medical Records, Minnesota Supercomputing Institute, September 19, 2003.
- [17] R. Hollebeek, National digital mammography archive, *BIO2003 Bioinformatics Track*, Washington DC (2003).
- [18] M. Heath, K. Bowyer, D.K.R. Moore, P. Kegelmeyer, The digital database for screening mammography, *The Proceedings of the 5th International Workshop on Digital Mammography*, Medical Physics Publishing, Madison, WI, USA, 2000.
- [19] J. Suckling, J. Parker, D. Dance, S. Astley, D. Betal, N. Cerneaz, S.-L. Kok, I. Ricketts, J. Savage, E. Stamatakis, P. Taylor, The mammographic image analysis society digital mammogram database, in: *Excerpta Medica. International Congress Series*, no. 1069, 1994, 375-378.
- [20] T. Arodz, M. Kurdziel, Face recognition from still camera images, *Master of Science Thesis*, Computer Science, AGH University of Science and Technology, Kraków, Poland, 2003.
- [21] P. Viola, M.J. Jones, Robust real-time face detection, *Int. J. Comput. Vision* 57 (2) (2004) 137-154.
- [22] F. Smeraldi, O. Carmona, J. Bigün, Saccadic search with Gabor features applied to eye detection and real-time head tracking, *Image Vision Comp.* 18 (4) (2000) 323-329.
- [23] R. Duda, P. Hart, D. Stork, *Pattern Classification*, John Wiley and Sons, 2000.
- [24] I. Daubechies, Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* 41 (1988) 909-996.
- [25] C. Cortes, V. Vapnik, Support vector networks, *Machine Learning* 20 (1995) 273-297.
- [26] V. Vapnik, S. Golowich, A.J. Smola, Support vector method for function approximation, regression estimation, and signal processing, in: M. Mozer, M. Jordan, T. Petsche (Eds.), *Advances in Neural Information Processing Systems 9*, MIT Press, Cambridge, MA, 1997, pp. 281-287.
- [27] B. Schölkopf, J. Platt, J. Shawe-Taylor, A.J. Smola, R.C. Williamson, Estimating the support of a high-dimensional distribution, *Neural Comp.* 13 (2001) 1443-1471.
- [28] A.J. Smola, B. Schölkopf, On a kernel-based method for pattern recognition, regression, approximation and operator inversion, *Algorithmica* 22 (1998) 211-231 technical Report 1064, GMD FIRST, April 1997.
- [29] S. Gunn, Support vector machines for classification and regression, Technical Report, Dept. of Electronics and Computer Science, University of Southampton, Southampton, U.K. (1998). in press.
- [30] J. Platt, Fast training of support vector machines using sequential minimal optimization, in: B. Schölkopf, C.J. Burges, A.J. Smola (Eds.), *Advances in Kernel Methods – Support Vector Learning*, MIT Press, Cambridge, MA, 1999, pp. 185-220.
- [31] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, K. R. K. Murthy, Improvements to platt's smo algorithm for svm classifier design, *Tech. rep.*, Dept of CSA, IISc, Bangalore, India (1999).
- [32] E. Osuna, R. Freund, F. Girosi, An improved training algorithm for support vector machines, in: J. Principe, L. Gile, N. Morgan, E. Wilson (Eds.), *Neural Networks for Signal Processing VII*, IEEE, New York, 1998, pp. 276-285.
- [33] C.J.C. Burges, D. Crisp, Uniqueness of the SVM solution, *Advances in Neural Information Processing Systems 12*, MIT Press, Cambridge, MA, 1999 pp. 223-229.
- [34] D. Gabor, Theory of communications, *J. Int. Electr. Eng.* 93 (1946) 427-457.
- [35] G. Granlund, In search of a general picture processing operator, *Comp. Graph. Image Proc.* 8 (1978) 155-173.
- [36] R. Meir, G. Rätsch, An introduction to boosting and leveraging, in: S. Mendelson, A. Smola (Eds.), *Advanced Lectures on Machine Learning*, 2003, pp. 119-184.
- [37] Y. Freund, R. Schapire, A short introduction to boosting, *J. Jpn. Soc. Art. Intell.* 14 (5) (1999) 771-780.
- [38] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, *EuroCOLT '95: Proceedings of the Second European Conference on Computational Learning Theory*, Springer-Verlag, 1995, pp. 23-37.
- [39] R. Gonzales, R. Woods, S. Eddins, *Digital Image Processing Using Matlab*, Prentice Hall, 2004.
- [40] T. Arodz, M. Kurdziel, T.J. Popiela, E.O.D. Sevre, D.A. Yuen, A 3D Visualization System for Computer-Aided Mammogram Analysis, *Univ. Minnesota Research Rep. UMSI 2004/181*, 2004.