Pattern Recognition Techniques for Automatic Detection of Suspicious-looking Anomalies in Mammograms

Tomasz Arodź, Marcin Kurdziel, Erik O. D. Sevre, David A. Yuen

1 Institute of Computer Science, AGH, al. Mickiewicza 30, 30-059, Kraków, Poland
phone: +48 12 617 3497, fax: +48 12 633 9406
tomaro@student.uci.agh.edu.pl
marcink@student.uci.agh.edu.pl

2 Minnesota Supercomputing Institute, Univ. of Minnesota, Minneapolis, MN 55455, USA
phone: +1 612 624 9801, fax: +1 612 625 3819
esevre@msi.umn.edu
davey@krissy.geo.umn.edu

Abstract. We have employed two pattern recognition methods used commonly for face recognition in order to analyse digital mammograms. The methods are based on two novel classification schemes, the AdaBoost and the support vector machines (SVM). A number of tests have been carried out to evaluate the detection rate of these two algorithms under different circumstances. The efficiency of recognizing mammogram features appeared to be lower than that for face detection. Results based on the AdaBoost classifier method are promising. In the best case the algorithm achieved a recognition success ratio of 88.2%. The SVM based algorithm did not perform well. In order to achieve a higher efficiency for SVM, we should unveil features with e.g. a method based on projective Fourier analysis that is better suited for analysing digital mammograms than the currently used sampling, which involves the Gabor decomposition.

1 Introduction

Breast cancer is the second major killer among American women [The American Cancer Society, 2003]. In 2003 around 40,000 women died from breast cancer and over 200,000 developed new cases of invasive breast cancer. Furthermore, apart from invasive cancer, at least 55,000 cases of in situ breast cancer were also diagnosed in 2003. With current treatment options, the 5-year survival rate for the localized breast cancer can be as high as 97%. However, this rate drops to 78% in the case of a regionally advanced disease and 23% for fast growing breast cancer. Consequently, early detection of the breast cancer is crucial for efficient therapy. The American Cancer Society recommends that every woman older than about

40 years should annually undergo mammogram examination. The sensitivity of screening mammography can be improved with the use of Computer-Aided Detection (CAD) systems. In [Freer and Ulissey, 2001] the interpretations of around 13,000 screening mammograms by 2 experienced mammographers 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 of 19.5%, i.e. 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.

Despite significant recent progress, the recognition of suspicious abnormalities in digital mammograms still remains a difficult task. There are at least several reasons. First, mammography provides relatively low contrast images, especially in case of dense or heavy breasts e.g. commonly found in young women. Second, symptoms of a presence of abnormal tissue may remain quite subtle. For example, spiculated masses that may indicate a malignant tissue within the breast 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 whether the detected lesion is benign or malignant with a significant level of uncertainty. For these reasons, we need robust algorithms for enhancing mammogram contrast, segmentation, detection of micro–calcifications and malignancy assessment. Various techniques have been proposed in the literature for contrast enhancement of digital mammograms. These include: techniques based on fractals [Li et al., 1997], wavelet transform [Laine et al., 1994], homogeneity measures [Cheng et al., 2004], and others. The segmentation of micro–calcifications has been done using e.g. morphological filters [Zhao et al., 1992], multiresolutional analysis [Netsch and Peitgen, 1999], and fuzzy logic [Pandey et al., 2000]. Furthermore, for detecting micro–calcification and malignancy assessment, several classification algorithms have been used. These include (among others): neural networks [Sehad et al., 1997], nearest neighbour classifier [Bhangale et al., 2000], multiple expert systems [Cordella et al., 2000], and support vector machines [El-Naqa et al., 2002]. A more detailed survey on the techniques used in the automatic analysis of digital mammograms can be found in [Cheng et al., 2003].

Recently there are many initiatives, which are devoted to the development of CAD systems and digital mammography. A good example is the National Digital Mammography Archive project [Hollebeek, 2003b,a]. 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, Sunnybrook 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 caused by the enormous number of high–resolution mammograms. A Digital

1 http://nscp.upenn.edu/NDMA
Database for Screening Mammography has been created at the University of South Florida [Heath et al., 1998]. This voluminous database provides high-resolution digitised mammograms for developing easy-to-use cancer detection algorithms and enabling the comparative analysis of detection accuracy. Another database of digital mammograms, i.e. mini-MIAS, is provided by The Mammographic Image Analysis Society [Suckling et al., 1994]. An efficient way of handling so many mammograms with up to petabytes of accumulated data is indeed a challenging task in computer science and information technology.

2 Motivation

In [Arodź and Kurdziel, 2003] the authors used two algorithms for face detection in images, i.e. AdaBoost with simple rectangular features [Viola and Jones, 2001] and SVM with log-polar sampling grid [Smeraldi et al., 2000]. Our study will focus on the adaptation of these algorithms to map out of suspicious regions of the breast in mammogram images and to evaluate their relative efficacy. 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 suspicious 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 [Arodź and Kurdziel, 2003]. Since the tests of the algorithms were performed only once, these results should be considered preliminary.

The face-detecting algorithms have been evaluated on the mini-MIAS database of mammogram images. This database consists of 117 images of breasts with benign or malignant changes, and 213 images of normal breasts i.e. without any such changes. The resolution of the images is 1024x1024 pixels, which yields a 200-micron pixel edge. For each of the abnormal tissue sample, the approximate position and radius of the breast lesion are provided. This data was used in the training of the classifiers and evaluating of classification accuracy.

3 Breast abnormalities in mammographic images

Micro-calcification clusters are important indicators of the presence of suspicious tissue in a breast. Figure 1 (courtesy of Ben Holtzmann) shows the global structure of the breast and the major regions. Micro-calcifications are likely in the region of fatty tissue. An example of a mammogram with two such clusters is depicted in Fig. 2. Isolated micro-calcifications are usually irrelevant from a diagnostic standpoint. If at least four or five micro-calcifications are developed and they form a cluster, then it is highly probable that an abnormal and malignant process is initiating within the breast. If the cluster is isolated, i.e., only a small number of singular micro-calcifications can be found around it, this probability is even greater. Micro-calcifications vary significantly in size, shape and distribution. Typically, homogenous clusters consisting of larger, round or oval
micro–calcifications indicate a benign process. An typical cluster of this type is pictured in Fig. 3. On the other hand, heterogeneous clusters consisting of smaller micro–calcifications, and micro–calcifications 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 classify whether micro–calcification clusters are benign or malignant. In a significant number of cases the clusters cannot be observed clearly and thus cannot be evaluated. In addition, it is common for a cluster of micro–calcifications 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 various sizes, shapes and margins. From a diagnostic point of view, the most important feature of a mass is the morphology of its margin. 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 where the mass has a lobulated shape, as in Fig. 6, the risk of malignancy increases. However, the most suspicious are the masses with ill-defined or spiculated margins. An example 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 types 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 have a significantly different mammographic appearance. This situation is referred to as the asymmetric density. These two types of breast abnormalities can rarely be targeted by common CAD systems.
Fig. 1. Global structure of the breast. Different major regions of the breast are displayed. Micro-calcifications are generally found in the fatty tissue and less commonly in the nipple area and glandular tissue.
Fig. 2. A mammographic image depicting the breast region with a size 5x5cm. On the image two clustered sets of micro-calcifications are marked. These 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 micro-calcifications. Regions with a less dense tissue have been suppressed and are depicted as a black area.
Fig. 3. A mammographic image depicting the breast region with an area of 5cm x 5cm. On the image a cluster of micro-calciﬁcations has been marked by the white dashed oval. These micro-calciﬁcations are round, small and uniform in size. These 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 as a black area.
Fig. 4. A mammographic image depicting the breast region with an area of 5cm x 5cm. On the image a cluster of micro-calcifications has been marked, which are highly heterogeneous in size and shape. Such an appearance is referred to as pleomorphic calcifications. Furthermore, the cluster is isolated, i.e. no micro-calcifications can be identified in the surrounding tissue. These 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 have been suppressed and are depicted as a black area.
Fig. 5. A mammogram of a right breast in the MLO projection. We have marked the image an anomalous mass. This mass is circumscribed and has a well-defined, sharp border. Furthermore, no other masses can be identified within the breast. Such findings suggest that this lesion is benign. The image has been taken from the mini-MIAS database of mammograms.
Fig. 6. A mammogram of a left breast in the MLO projection. On the image an anomalous mass has been marked. The border of this mass is not sharp but 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. The image has been taken from the mini-MIAS database of mammograms.
Fig. 7. A mammogram of a right breast in the MLO projection on which a mass has been marked. This anomalous mass has a sharply defined border. Such an appearance is referred to as spiculated mass and usually indicates the presence of a malignant, invasive tumour.
4 Algorithms for detecting abnormalities in mammograms

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 learn a classifier, i.e. an algorithm capable of making a binary decision on a given image features (in this case is a lesion vs. not a lesion). Our algorithms that extract the image features as well as classification algorithms used in this paper are discussed in more details below. After the classifier has been learned, the image features from the testing set are used to evaluate its effectiveness. In particular, a 2x2 confusion matrix is computed: $A_{i,j}$. The element $a_{i,j}$ of this matrix represents the probability that the image belongs to class $i$, provided that the classification algorithm decides that it belongs to class $j$, ($i,j \in \{\text{lesion, non-lesion}\}$). Ideally, the confusion matrix should have values near 1.0 on the main diagonal, and values near 0 along the second diagonal. Such situations indicate that the classification algorithm is capable of properly discriminating between regions of mammograms with lesions or no lesions.

Algorithm 1 Detection scheme for suspicious lesions

INPUT:
- $D$ - Database of grey-scale, 8bit images of resolution $1024 \times 1024$ pixels containing breast mammograms.
- $C$ - 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 Rectangular Regions From $I$
  - Partition the selected regions into training set TrainSet and testing set TestSet
  - Learning:
    - Generate Feature Vectors $F_{Train}$ for TrainSet
    - Learn classifier $C$ on $F_{Train}$
  - Validation:
    - Generate Feature Vectors $F_{Test}$ for TestSet
    - foreach $v$ in $F_{Test}$
      - Classify $v$ using $C$
      - Calculate classification accuracy and the confusion matrix

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

Support vector classification. The support vector classification (SVC) was first proposed for optical character recognition in [Cortes and Vapnik, 1995]. New
algorithms were also proposed for regression estimation [Vapnik et al., 1997], novelty detection [Schölkopf et al., 1999], operator inversion [Smola and Schölkopf, 1998] and other problems.

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

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

where: \( x \) is the feature vector, \( w \) is the vector that is perpendicular to the hyperplane, and \( \frac{b}{\|w\|} \) specifies 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 \( \mathcal{F} \) introduced a nonlinear mapping \( \Phi \) of the input patterns. As the hyperplane construction involves the computation of dot-products on the feature space the mapping \( \Phi \) must satisfy:

\[ \langle \Phi(x_1), \Phi(x_2) \rangle = k(x_1, x_2) \quad \forall x_1, x_2 \in \mathcal{X} \tag{2} \]

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, \( x_i \in \mathcal{X} \) denotes the input patterns. Maximizing the distance of this separation is equivalent to the minimization of the norm-squared: \( \frac{1}{2}\|w\|^2 \) [Gunn, 1998]. However, to guarantee that the resultant hyperplane separates the classes, the following constraints must be satisfied:

\[ y_i \cdot (\langle w, x_i \rangle + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \ldots n \tag{3} \]

where \( y_i \in \{-1, 1\} \) denotes the class label corresponding to the input pattern \( x_i \). These constraints do not impose a strict class separation. Instead, slack variables \( \xi_i \) have been introduced 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 of the form:

\[ \min_{w \in \mathbb{R}^n} \min_{\xi \in \mathbb{R}^n} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} \xi_i \tag{4} \]

and the constraints are given by eq. (3). The parameter \( C \) controls the penalty for misclassification of training samples. Using the Lagrange multiplier technique, this optimisation problem can be transformed to the dual form:

\[ \min_{\alpha \in \mathbb{R}^n} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \cdot k(x_i, x_j) \]

subject to:

\[ 0 \leq \alpha_i \leq C \]
\[ \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{5} \]
In the above formulation, the \( \alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_n\} \) 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 \( \Phi \) must satisfy eq. (2). The lagrange multipliers that solves the eq. (5) can be used to compute the decision function:

\[
f(x) = \sum_{i=1}^{n} \alpha_i y_i k(x_i, x) + b
\]

where:

\[
b = y_i - \sum_{j=1}^{n} \alpha_j y_j k(x_j, x_i)
\]

The solution to eq. (5) can be found using any general purpose quadratic programming solver. Furthermore, dedicated heuristical methods have been developed that can solve large problems efficiently [Platt, 1999; Keerthi et al., 1999; Osuna et al., 1998].

Finally, it should be noticed that the optimization problem form eq. (5) is convex [Burges and Crisp, 1999]. 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 immune to overfilling than the classic feed-forward neural network.

**Log-polar sampling of the Gabor decomposition.** The Gabor filters are useful tools for extracting feature vectors from images. Originally, the filters have been proposed in [Gabor, 1946] as a Gabor Elementary Function and afterwards extended in [Granlund, 1978] to two-dimensional image operators. The two-dimensional Gabor filter is defined as:

$$G(x, y) = \frac{1}{2\pi \sigma_x \sigma_y} e^{-\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right)} e^{i2\pi \nu_0 x}$$  \hspace{1cm} (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 $\nu_0$. The $\sigma_x$ and $\sigma_y$ parameters are the x-axis and y-axis widths of the gaussian. 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 [Smeraldi et al., 2000] the modified Gabor filters are employed, which are cast in log-polar coordinates:

$$\hat{G}(r, \phi) = A e^{\frac{(r - r_0)^2}{2\sigma_r^2}} e^{-\frac{(\phi - \phi_0)^2}{2\sigma_\phi^2}}$$  \hspace{1cm} (9)

where:

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

$$\phi = \arctan \frac{\mu}{\nu}$$  \hspace{1cm} (10)

The $\sigma_r$ and $\sigma_\phi$ parameters controls the radius-axis width and angle-axis width of the gaussian.

This approach is useful in construction of filter banks with different orientations $\phi$ and central frequencies $\nu_0$. 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 $\sigma_x$ and $\sigma_y$ values for filters with small $\nu_0$ frequency.

Finally, in [Smeraldi et al., 2000] 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 $X$. 
Fig. 9. A grid used to sample the responses of a bank of Gabor filters, as proposed in [Smeraldi et al., 2000]. To sample the responses for a given point \( p \), the grid is centered 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 form all grid points in a predefined order.
4.2 Boosting method for detecting abnormalities

Unlike neural networks [Sehad et al., 1997], 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 extent. However, a scheme for learning 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. [Meir and Rätsch, 2003] or [Freund and Schapire, 1999].

Our first choice from the many boosting algorithms is the AdaBoost classifier of [Freund and Schapire, 1995]. The pseudocode of this classifier is presented below. This classifier is used also in [Viola and Jones, 2001]. 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 weights. Each iteration changes the weights values reducing them by an amount, 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 one weak classifier are classified correctly by the other ones.

The process of learning 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 \( T_r \) with weights \( D_t \). The training set is formed by examples from domain \( X \) labelled with labels from a set \( C \). The learning of the weak classifier is left to an unspecified \( WeakLearner \) algorithm, which should minimize the training error \( \varepsilon_t \) of the produced weak classifier with respect to the weights \( D_t \).

Based on the error \( \varepsilon_t \) of the weak classifier \( h_t \), the parameters \( \alpha_t \) and \( \beta_t \) 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 training set are increased. Therefore, in the \( h_{t+1} \) round, the \( WeakLearner \) is more focused on these examples. The chance that the classifier \( h_{t+1} \) will learn to classify them correctly is increased.

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 \( \alpha_t \).
There are two special cases, which are treated individually during the algorithm execution. One is the case of $\varepsilon_t$ 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 makes no further rounds of learning. 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 learning.

Algorithm 2 The AdaBoost algorithm

**INPUT:**
Labelled training set $Tr$ of size $m$, $Tr = \{(x_1, c_1), \ldots, (x_m, c_m)\} \subset X \times C$

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

$\beta_t = \frac{\alpha_t}{1-\varepsilon_t}$

$\alpha_t = \log \frac{1}{\beta_t}$

$\forall i : h_t(x_i) = c_i D_{t+1}(i) = \beta_t D_t(i)$

$\forall i : h_t(x_i) \neq c_i D_{t+1}(i) = D_t(i)$

normalize distribution $D_{t+1}$

return $h_{\text{fin}} = \arg \max_c \sum_i \alpha_t$

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 [Viola and Jones, 2001], 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 involving 24 × 24 pixels.

5 Evaluation of the algorithm based on the SVM classifier

In this section we present the results on the classification of mammograms based on the adaptation of the SVM algorithm previously used for face detection in
[Arodź and Kurdziel, 2003]. We have classified the breast tissues pictured on the mammogram images.

5.1 Dataset

For the purpose of this study, we have selected around 160 images from the mini-MIAS database have been selected. Next, abnormal breast regions from these images were manually segmented and image feature vectors were extracted using log-polar sampling grid. In particular, the grid was being centred at random locations within the segmented regions. A rectangular window of the size 85x85 pixels, containing the grid, was being extracted from the image and filtered to compute the feature vector components. In this approach several image feature vectors are extracted from various parts of an abnormal region. The resultant image feature vectors represent the abnormal samples. Similarly, image feature vectors for normal tissue were extracted from the breast lying outside the segmented abnormal regions. This image feature vectors represent normal samples.

The extracted vectors were split in two sets:

- training set of 262 normal samples and 72 abnormal samples,
- test set of 46 normal and 46 abnormal samples.

Both the training and the test sets have been collected from samples extracted from different images i.e. no single image was used in preparation of these two sets.

5.2 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 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 85x85 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 $\frac{\pi}{2}$ radian.

5.3 Training phase and the results of tests

The training set described in the previous section was used to teach a support vector classifier and to distinguish between normal and abnormal samples. The training was performed independently for the kernel functions below:

- linear kernel: $k(x, y) = \langle x, y \rangle$
- polynomial kernel: $k(x, y) = (\sigma \langle x, y \rangle + \gamma)^d$
- Gaussian RBF kernel: $k(x, y) = e^{-\frac{|x-y|^2}{2\sigma^2}}$
For each of those kernel functions, a range of the parameter values was tested. In addition, tests for different values of the misclassification penalty $C$ were performed ([Arodź and Kurdziel, 2003]).

To obtain the classification ratios and corresponding confusion matrices, we have validated the classifiers on the test set. The best results for each of the kernel functions along with corresponding parameter values are presented in tables 1, 2, 3 and 4 in section 6 below. The results indicate that the algorithm does not discriminate between normal and abnormal breast regions. In particular, the abnormal breast regions have a probability equal to 0.5 of being classified as a normal sample. At the same time, normal breast regions are classified more correctly.

6 Evaluation of the AdaBoost-based classification algorithm

In this section we evaluate the method based on the AdaBoost algorithm used for face detection in [Arodź and Kurdziel, 2003]. Similarly to the SVM classifier, the AdaBoost was trained on the mini-MIAS database of mammogram images [Suckling et al., 1994].

6.1 Preparation of the training and testing sets

Unlike SVM, the mammogram images from the mini-MIAS database were randomly divided into two sets:

- a training set of 59 abnormal and 59 normal mammograms
- a test set of 58 abnormal and 154 normal mammograms

These images were then used by the classifier both in the training and in the testing stages. We use AdaBoost for classifying image windows with size of 24x24 pixels. Therefore, we need an efficient method of moving the 24x24 pixels window on a mammogram with 1024x1024 pixels. 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. This paper focuses on the classification task. In particular, for each of the mammograms with some abnormal tissue changes, we use a window centred on the centre of the change. Since the changed region can be significantly larger than 24 pixels in the diameter, we employ the wavelet approximation of the input mammogram. There are two scenarios for choosing the level of this approximation:

**Fixed scaling** - each mammogram is downscaled, using the Daubechies-4 wavelet twice. The diameter of the change is not used.

**Variable scaling** - each mammogram is downscaled, using the Daubechies-4 wavelet. However, the number of successive approximations depends on the diameter of the change. As a result, the diameter of the change is always smaller than the window width (i.e. 24 pixels) but larger than a half of it.
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 second-level wavelet approximation. In the second scenario, the level of approximation is randomly selected from the range found in the mammograms that contain the cancerous change (i.e. range of 0 to 3). For mammograms without any changes, the centre of the window is placed randomly, as long as the window does not cross the breast boundary.

6.2 Input image filtering

Before we apply the wavelet scaling and window selection, the image is filtered using various types and combination of filters. The filters below represent a group of typical basic filters used in image processing. In addition, combinations of some filters are also used. However, this approach is employed only for filters, which by themselves yield the best results. The filters used include:

- No filtering
- Unsharp - Unsharp contrast enhancement filter, i.e. negation of the Laplacian
- Sobel - Sobel horizontal filter
- Prewitt - Prewitt horizontal filter
- Laplacian - Laplacian filter
- LoG - Laplacian of Gaussian filter
- Motion - Horizontal motion filter
- Sobel-Unsharp - Sobel horizontal filter on result of Unsharp filter
- Prewitt-Unsharp - Prewitt horizontal filter on result of Unsharp filter
- Dilate - Greyscale dilation using disk structuring element
- Sobel-Dilate - Sobel horizontal filter on the result of greyscale dilation, using disk structuring element
- Prewitt-Dilate - Prewitt horizontal filter on result of greyscale dilation using disk structuring element
- Unsharp-Dilate - Unsharp filter on result of greyscale dilation using disk structuring element

6.3 Classifier training and results of the tests

After filtering, wavelet scaling and window selection, the training set is represented by 59 images of size 24x24 pixels containing abnormal (benign or malignant) tissue, and 59 images containing normal tissue. This data are then used for training of the AdaBoost classifier with the simple, rectangular features [Viola and Jones, 2001; Arodz and Kurdziel, 2003]. 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. We use the same position of window in the normal mammograms, i.e. the random selection of position is carried only once, at the beginning of training cycle. For each configuration we have partitioned the images in the same way into training and test sets.
The trained classifier is tested on the dataset composed of 58 abnormal and 154 normal images. Each of them is filtered and has a wavelet scaled window 24x24 pixels in size. Results for various types of filtering and wavelet scaling are obtained. For each different filtering type and wavelet scaling mode, we use the same position of window in the normal mammogram. Thus, at the beginning of the test cycle, we carry out a random selection of position only once.

For each configuration, an overall classification accuracy, i.e. the ratio of correctly classified samples to the total number of samples is given in Table 5. For the configuration that uses the Fixed Scaling, no filtering is tested, as this configuration is inferior to the Variable Scaling configuration. In Table 6 we present the confusion matrix for a Fixed Scaling configuration, which specified the classification ratio for each abnormal and normal tissue.

For most of the Variable Scaling configurations, the ratios of classification for abnormal tissue and normal tissue separately is similar to the overall classifier accuracy (e.g. see confusion matrix in Table 7). However, in some cases, assembled in Tables 8, 9, the confusion matrices are not so uniform.

6.4 Tests for more realistic scenario

For the scheme used above, in which the samples of normal tissue for training and testing are collected randomly from the entire breast, the results may lead to an unrealistically high classification accuracy. The reason is that large number of samples may contain fatty or muscle tissue, which is easily classified. Therefore, we must discard regions of the breast, which have a very low probability of being cancerous.

Watershed-based breast segmentation In order to allow for more realistic evaluation of the classifier, now we propose a new way of segmenting the images. The segmentation is used before selection of the tissue samples. The result of the segmentation, the region of interest (ROI), is a part of the breast, which has higher probability of containing cancerous tissue than the whole breast. The method for selection of abnormal tissue remains the same, since the position of the abnormal region is given. However, the position of the cancerous region in the abnormal breasts constrains the choice of the segmentation method. In particular, when the segmentation is carried out for an abnormal breast, the ROI given by method should contain the centre of the abnormal tissue.

The simple segmentation method proposed in this paper is based on the watershed transform [Vincent and Soille, 1991]. First, the image is negated, so that the brightest pixels in the image have the lowest intensity values. Then, the watershed transform is performed, splitting the breast into a large number of watershed regions. The regions are numbered according to the immersion scheme. In principle, the lower the number associated with the region, the lower is the minimal pixel intensity in the region. The final ROI consists of the regions that have a range of numbers below a certain threshold, as well as of the watershed
boundaries between those regions. We assume that the threshold is fixed to be 30% of the maximal number in the given mammogram, although some adaptive method might yield better results.

The area of the mammogram containing the *pectoralis major muscle* was excluded from the region of interest, as this region is highly homogenous and none of the abnormal changes was situated in this part of the mammogram. In principle there is a positive, albeit very low probability of a cancerous tissue situated in this region of mammogram. The original breast image and the mask of the region selected by the segmentation algorithms are depicted in Fig. 10 and Fig. 11.

![Fig. 10](image.jpg)

**Fig. 10.** Results of watershed-based segmentation. Breast image comes from the min *i-MIAS* database [Suckling et al., 1994].
Fig. 11. Result of watershed-based segmentation for a more glandular breast. Breast image from the mini-MIAS database [Suckling et al., 1994].

We show in Table 10 the ratio of the area of the selected region of interest to the area of the whole breast, as well as to the area of the breast without the removed muscle. For the abnormal breasts, 87% of the centres of the abnormal tissue regions is situated in the selected ROI.

Classification results Several algorithm configurations have been tested on the new training and test sets. These include no filtering, as well as unsharp, Sobel, Prewitt and dilate filtering of the image, all within the Variable Scaling scheme. We summarize all of the results in Table 11 12, 13 for a configuration with relatively high difference in the real positive and negative ratios.
7 Concluding remarks

In the 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 the abnormal tissue. This is because the abnormal breast regions are often confused with normal regions, while the opposite case does not hold. The most probable reason of 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 [Arodź and K urdziel, 2003]. 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, because the micro-calcifications are very small. The imbalance between normal and abnormal samples in the training set does not explain the observed misclassifications. If the image feature vectors were derived from two separate classes, the SVM algorithm would be able to construct a proper decision boundary regardless of the disproportionality between the number of samples in each class.

Results for the AdaBoost-based approach are promising. The classifier accuracy is almost 90% when the normal tissue is collected randomly from the whole area of breast and decreases to about 80% when recognition ratio is more constrained. The introduction of various filters leads to significant changes in the recognition accuracy. In general, the ratios of false negatives and false positives are similar, reaching from 10% to 20%. Such a large number of false negatives are unacceptable in medical applications. However, these results can be treated as the lowest values we can expect from a mammogram detection algorithm. The tested algorithms were not 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, in order 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 out of 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 the uninteresting regions.

We would recommend that future work be focused on the development of the procedures focused on a local processing of data such as:

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 accentuate the abnormal tissue in digitised mammograms.

Our results show the feasibility of detecting clusters of micro-calcifications using modern pattern recognition methods. In comparison to other types of abnormalities found in mammograms, it is relatively easy to detect micro-calcifications. Last by not least, the isolated clusters of micro-calcifications strongly indicate the presence of an abnormal tissue.

Acknowledgments
We thank Professor Robert Hollebeck, University of Pennsylvania, for 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 Figure 1. This research has been supported by the Math-Geo program of National Science Foundation.
Table 1. Results for support vector machines with different kernel functions [%]

<table>
<thead>
<tr>
<th>Kernel Function</th>
<th>Parameters values</th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>$C = 30$</td>
<td>66.3%</td>
</tr>
<tr>
<td>Polynomial</td>
<td>$d = 4, \sigma = 1.5$</td>
<td>70.6%</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 0.9, C = 20$</td>
<td></td>
</tr>
<tr>
<td>Gaussian RBF</td>
<td>$\sigma = 2, C = 25$</td>
<td>66.3%</td>
</tr>
</tbody>
</table>

Table 2. Confusion Matrix - support vector machines with linear kernel [%]

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Actual</th>
<th>Abnormal</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal</td>
<td>54.3%</td>
<td>45.7%</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>13.1%</td>
<td>86.9%</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Confusion Matrix - support vector machines with polynomial kernel [%]

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Actual</th>
<th>Abnormal</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal</td>
<td>52.1%</td>
<td>47.9%</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>10.9%</td>
<td>89.1%</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Confusion Matrix - support vector machines with Gaussian RBF kernel [%]

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Actual</th>
<th>Abnormal</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal</td>
<td>39.1%</td>
<td>60.9%</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>6.6%</td>
<td>93.4%</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Results for AdaBoost and various filters

<table>
<thead>
<tr>
<th>Scaling type</th>
<th>Filtering type</th>
<th>Classification accuracy [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td>No filtering</td>
<td>72.6</td>
</tr>
<tr>
<td>Variable</td>
<td>No filtering</td>
<td>82.1</td>
</tr>
<tr>
<td></td>
<td>Unsharp</td>
<td>83.0</td>
</tr>
<tr>
<td></td>
<td>Sobel</td>
<td>86.3</td>
</tr>
<tr>
<td></td>
<td>Prewitt</td>
<td>86.3</td>
</tr>
<tr>
<td></td>
<td>Laplacian</td>
<td>60.8</td>
</tr>
<tr>
<td></td>
<td>LoG</td>
<td>65.6</td>
</tr>
<tr>
<td></td>
<td>Motion</td>
<td>65.6</td>
</tr>
<tr>
<td></td>
<td>Sobel-Unsharp</td>
<td>88.2</td>
</tr>
<tr>
<td></td>
<td>Prewitt-Unsharp</td>
<td>87.3</td>
</tr>
<tr>
<td></td>
<td>Dilate</td>
<td>86.3</td>
</tr>
<tr>
<td></td>
<td>Sobel-Dilate</td>
<td>87.3</td>
</tr>
<tr>
<td></td>
<td>Prewitt-Dilate</td>
<td>87.3</td>
</tr>
<tr>
<td></td>
<td>Unsharp-Dilate</td>
<td>84.9</td>
</tr>
</tbody>
</table>
Table 6. Confusion Matrix - AdaBoost, Constant Scaling, No filtering

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal</td>
<td>Abnormal</td>
<td>82.8%</td>
</tr>
<tr>
<td>Abnormal</td>
<td>Normal</td>
<td>17.2%</td>
</tr>
<tr>
<td>Normal</td>
<td>Abnormal</td>
<td>31.2%</td>
</tr>
<tr>
<td>Normal</td>
<td>Normal</td>
<td>68.8%</td>
</tr>
</tbody>
</table>

Table 7. Confusion Matrix - AdaBoost, Variable Scaling, No filtering [%]

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal</td>
<td>Abnormal</td>
<td>87.9%</td>
</tr>
<tr>
<td>Abnormal</td>
<td>Normal</td>
<td>12.1%</td>
</tr>
<tr>
<td>Normal</td>
<td>Abnormal</td>
<td>14.3%</td>
</tr>
<tr>
<td>Normal</td>
<td>Normal</td>
<td>85.7%</td>
</tr>
</tbody>
</table>

Table 8. Confusion Matrix - AdaBoost, Variable Scaling, Prewitt [%]

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal</td>
<td>Abnormal</td>
<td>91.4%</td>
</tr>
<tr>
<td>Abnormal</td>
<td>Normal</td>
<td>8.6%</td>
</tr>
<tr>
<td>Normal</td>
<td>Abnormal</td>
<td>15.6%</td>
</tr>
<tr>
<td>Normal</td>
<td>Normal</td>
<td>84.4%</td>
</tr>
</tbody>
</table>

Table 9. Confusion Matrix - AdaBoost, Variable Scaling, Sobel-Unsharp [%]

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal</td>
<td>Abnormal</td>
<td>84.5%</td>
</tr>
<tr>
<td>Abnormal</td>
<td>Normal</td>
<td>15.5%</td>
</tr>
<tr>
<td>Normal</td>
<td>Abnormal</td>
<td>10.4%</td>
</tr>
<tr>
<td>Normal</td>
<td>Normal</td>
<td>89.6%</td>
</tr>
</tbody>
</table>

Table 10. Segmentation results - area in the region of interest

<table>
<thead>
<tr>
<th>Breast type</th>
<th>Value type</th>
<th>ROI / Breast</th>
<th>ROI / (Breast - muscle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal</td>
<td>Mean</td>
<td>32%</td>
<td>37%</td>
</tr>
<tr>
<td>Abnormal</td>
<td>StdDev</td>
<td>12%</td>
<td>12%</td>
</tr>
<tr>
<td>Abnormal</td>
<td>Max</td>
<td>69%</td>
<td>69%</td>
</tr>
<tr>
<td>Abnormal</td>
<td>Min</td>
<td>6%</td>
<td>7%</td>
</tr>
<tr>
<td>Normal</td>
<td>Mean</td>
<td>34%</td>
<td>38%</td>
</tr>
<tr>
<td>Normal</td>
<td>StdDev</td>
<td>13%</td>
<td>12%</td>
</tr>
<tr>
<td>Normal</td>
<td>Max</td>
<td>96%</td>
<td>96%</td>
</tr>
<tr>
<td>Normal</td>
<td>Min</td>
<td>7%</td>
<td>9%</td>
</tr>
</tbody>
</table>
Table 11. Results for AdaBoost with Variable Scaling and various filters

<table>
<thead>
<tr>
<th>Filtering type</th>
<th>Classification accuracy [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>No filtering</td>
<td>81.1</td>
</tr>
<tr>
<td>Sobel</td>
<td>82.6</td>
</tr>
<tr>
<td>Prewitt</td>
<td>82.0</td>
</tr>
<tr>
<td>Dilate</td>
<td>83.9</td>
</tr>
</tbody>
</table>

Table 12. Confusion Matrix - AdaBoost, Variable Scaling, Unsharp, Segmentation [%]

<table>
<thead>
<tr>
<th></th>
<th>Predicted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abnormal</td>
<td>Normal</td>
</tr>
<tr>
<td>Actual</td>
<td>Abnormal</td>
<td>74.1%</td>
</tr>
<tr>
<td>Actual</td>
<td>Normal</td>
<td>14.3%</td>
</tr>
</tbody>
</table>

Table 13. Confusion Matrix - AdaBoost, Variable Scaling, Unsharp, Segmentation [%]

<table>
<thead>
<tr>
<th></th>
<th>Predicted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abnormal</td>
<td>Normal</td>
</tr>
<tr>
<td>Actual</td>
<td>Abnormal</td>
<td>74.1%</td>
</tr>
<tr>
<td>Actual</td>
<td>Normal</td>
<td>14.9%</td>
</tr>
</tbody>
</table>
Bibliography


URL http://research.microsoft.com/users/cburges/papers/nips_unique.ps.gz


URL http://radiology.rsna.org/cgi/content/abstract/220/3/781


URL http://www.boosting.org/papers/FreSch95.ps.gz


URL http://nscp.upenn.edu/hollebeek/talks/msi/technicalv3.htm


URL http://guppy.mpe.nus.edu.sg/ mpessk/npa_tr.ps.gz


URL http://www.boosting.org/papers/MeiRae03.ps.gz


URL http://www.kernel-machines.org/papers/oneclass-tr.ps.gz
