

A Computer Aided Detection (CAD) System for Microcalcifications in Mammograms - MammoScan μ CaD

Kjersti Engan, Thor Ole Gulsrud, Karl Fredrik Fretheim, Barbro Furebotten Iversen, Liv Eriksen

Abstract—Clusters of microcalcifications in mammograms are an important sign of breast cancer. This paper presents a complete Computer Aided Detection (CAD) scheme for automatic detection of clustered microcalcifications in digital mammograms. The proposed system, MammoScan μ CaD, consists of three main steps. Firstly all potential microcalcifications are detected using a method for feature extraction, VarMet, and adaptive thresholding. This will also give a number of false detections. The goal of the second step, Classifier level 1, is to remove everything but microcalcifications. The last step, Classifier level 2, uses learned dictionaries and sparse representations as a texture classification technique to distinguish single, benign microcalcifications from clustered microcalcifications, in addition to remove some remaining false detections. The system is trained and tested on true digital data from Stavanger University Hospital, and the results are evaluated by radiologists. The overall results are promising, with a sensitivity $> 90\%$ and a low false detection rate (approx 1 *unwanted* pr. image, or 0.3 *false* pr. image).

Keywords—mammogram, microcalcifications, detection, CAD, MammoScan μ CaD, VarMet, dictionary learning, texture, FTCM, classification, adaptive thresholding

I. INTRODUCTION

BREAST cancer is the most frequent cancer disease, and a leading cause of cancer death among women. In Norway, populated by 4.5 million people, as many as 2644 women developed breast cancer and 715 died from breast cancer in 2003 [1]. The survival rate is greatly influenced by how early the cancer is treated, thus it is important to discover the disease at an early stage. There are two main types mammographic findings indicating a possibly cancer in the breast tissue. Soft tissue lesions and clusters of microcalcifications. Soft tissue lesions can be subdivided in different groups, they can be malignant or benign, and are sometimes palpable. Clusters of microcalcifications are early sign of possibly cancer, or precancerous changes, and are in general *not* palpable. Microcalcifications are small calcifications of different shape and density, approx 0.1 - 1 mm in diameter. Single microcalcifications are not dangerous, but clusters of microcalcifications might be malignant or benign. An example of a cluster of microcalcifications is seen in Figure 5 a). In

the following, (a) microcalcification(s) will often be denoted μ Ca for simplicity. When we are referring to a *cluster of microcalcifications*/ μ Ca this will be specified.

Mammograms are X-ray projections of the breast tissue onto a detector array or a film plate. The exposure of the mammographic film or detector array is proportional to the intensity of the X-ray photons transmitted through the breast. Tumors and microcalcifications are denser than the surrounding tissue, thus absorb more of the incident X-rays. They can therefore often be seen as bright spots/regions in the mammograms. Most western countries today have *mammographic screening* programs. A screening is defined as the presumptive identification of unrecognized disease or defect by application of tests, examinations, or other procedures. The procedures should be easy to carry out and not to painful. In Norway every women between 50 and 69 years of age are invited to screening every second year. Evaluating screening data is a very labor and time demanding process. Only a few of the cases will actually be present with cancer (approx. 0.6 % in Norway), and due to the vast amount of images relatively rapid interpretation is done. Hence there is a risk that subtle abnormalities can be overlooked. In Norway today, two independent radiologists evaluate all screening data. This is recognized as the best way of evaluating screening data and many other countries do the same. However, due to lack of qualified radiologists, or for economical reasons, some places screening data are evaluated by a single radiologist.

A well working Computer-Aided Detection (CAD) system could be used as a second opinion. Whether CAD can replace *one* of the two independent radiologists is under discussion and still needs to be proven. However, CAD would, no doubt, be helpful where no second radiologist is at hand. A CAD system can also be helpful in clinical analysis and in training of radiologists. As early as in 1990 Chan, Doi et.al. showed that CAD (state of the art in 1990) did significantly improved radiologists' accuracy in detecting clustered μ Ca under conditions that simulated the rapid interpretation of screening mammograms [2].

Computer-aided methods for detecting clustered μ Ca have been investigated using many different techniques. Some of the important and/or later work is briefly explained in the following: Cheng, Lui, and Freimanis suggest a fuzzy logic technique in [3], showing good enhancement capabilities for μ Ca. Veldkamp and Karssemeijer use a two step approach in [4]. The first step is initial detection using a statistical method, based on Bayesian techniques, and application of a

Manuscript received May, 2007

Kjersti Engan is with the dept. of Electrical and Computer Eng. at University of Stavanger, N-4036 Stavanger, Norway, (email:kjersti.engan@uis.no)

Thor Ole Gulsrud is with Roxar Flow Measurement AS, Stavanger, Norway (email: Thor.Ole.Gulsrud@roxar.com)

Karl Fredrik Fretheim is with dva Optical Networking AS, Oslo, Norway

Barbro Furebotten Iversen and Liv Eriksen are with the Breastdiagnostic Center at Stavanger University Hospital (SUS), Stavanger, Norway

Markov random field model for segmentation. The second step starts by calculating a number of features for each detection, and classification between true and false cluster of μCa is done by a k-nearest neighbor method. Yu and Guan [5] also use a two step approach. Firstly, potential μCa are segmented from the mammogram using mixed features consisting of wavelet features and gray level statistical features. Second, a large number of features (31) are calculated from the original image for all the potential μCa . The discriminatory power of these features is analyzed using general regression neural networks. A learned Support Vector Machine (SVM) approach is investigated by El-Naqa and Yang et.al. [6]. An SVM is trained using supervised learning, and thereafter used to detect if a μCa is present at each location in the image. The pixels classified as μCa are grouped together to form μCa objects by using morphological processing as described in [7]. Clusters of μCa were identified by grouping the μCa objects, with the following rules: 1) The objects are connected with nearest-neighbor distance less than 0.2 cm. 2) Three μCa should be detected within an area of 1 cm^2 . Lemaury and Drouiche et.al. develop new, highly regular, wavelets for the detection of clustered μCa [8]. The detail coefficients at the first scale level is hard thresholded. The highly regular wavelets are shown to have better capabilities to detect clustered μCa compared to ordinary wavelets. Related work is done by Heinlein and Drexler et.al. in [9], where integrated wavelets for enhancement of multiscale structures in images are developed. The integrated wavelets are shown to have good enhancement capabilities for μCa . Both these work focus on the feature extraction, or image enhancement part, and do not include a complete CAD system. A neural-genetic algorithm for feature selection and breast abnormality classification is presented by Zhang, Verma and Kumar in [10]. This would be the "last part" of a complete CAD system, where they use already segmented areas as input, and report their *classification capabilities* (i.e. *decide for benign or malign*). Yoshida presents a Matching Pursuit (MP) based method for extraction of μCa inside predefined ROI's in [11]. A predefined wavelet packet dictionary is used (10-tap Symmlet), and the different wavelets are associated with individual weights, trained using a training set. This can also be regarded as a "last part" of a full CAD system, since predefined ROI's are expected. Horváth, Valyon et. al. are working on a complete CAD system called, intelligent advisory system for screening mammography [12], where they plan to take into account comparing of views as well as detecting asymmetry (comparing breasts), and changing (comparing history with new images). Their μCa system is based on region of interest extraction after texture analysis followed by neural network to classify the extracted regions as clusters or not. A method to extract and select features for μCa detection using multiscale image processing and artificial neural networks is presented by Vega-Corona and Andina in [13]. They apply a generalized regression neural network in feature selection and identification.

Our system, MammoScan μCaD , is built as a three step approach. Firstly, detection of suspicious areas using a method we have named VarMet [14]. Second step removes most false detections (but no μCa). Third step removes single/spread

out benign μCa , so that only clusters of μCa are kept using *learned dictionaries and sparse representation as a classifier technique*. The novelty of this work lies in the algorithm of step 1, the use of learned dictionaries for classification of mammograms in step 3 as well as putting it all together as a *complete CAD system* for microcalcifications and testing it on true digital data. Step 2 uses more or less standard image processing techniques.

The outline of the paper is as follows: In Section II, an overview of the system is presented. Details of the detection part is presented in Section III, details of the Classifier level 1, removing false detections, is found in Section IV, and details of the Classifier level 2, removing single, benign μCa , is found in Section V. Section VI includes a case study as well as volume experiments and results. Section VII concludes the paper.

II. OVERVIEW OF THE SYSTEM

In this section we will give an overview of the system for detection and classification of μCa in mammograms, MammoScan μCaD . Figure 1 shows a block diagram of the system with the different blocks numbered. Since the digital

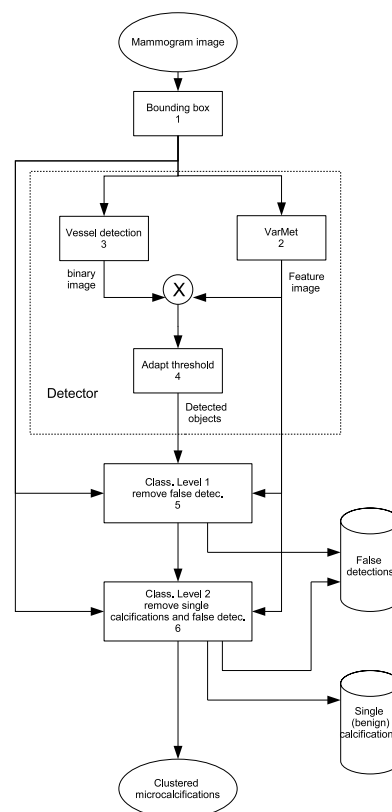


Fig. 1. Schematic overview of the system, MammoScan μCaD .

images from the digital mammograph all have the same size, but the size of the breast region varies, the first step, *box 1*, is to shrink the image to contain only a bounding box around

the breast region. On fully digital images as used in this work, finding the breast region is an easy task, and can be done by simple thresholding followed by an erode operation. The erode operation is done to avoid searching for μCa in the skinline area. This procedure results in a binary image of the breast region. Both the original mammogram, and the binary image, are reduced to the bounding box around the breast region, resulting in I_{orig} and I_b respectively.

Box 2 is a μCa detection algorithm. The algorithm used is called VarMet, and is a method that emphasizes edges and objects of a certain size. μCa are small but with relatively sharp edges compared to the background tissue, and this is exploited. VarMet is a method based on median filtering in different scales, followed by variance calculations. The algorithm is explained thoroughly with various experiments in [14]. The output from VarMet (and *box 2*) is a feature image where all μCa are emphasized, unfortunately in addition to some other objects, like μCa in vessels and connective tissue lines. In *box 3* some of the vessels (containing μCa) are detected. Such areas will often result in very many false detections in many CAD systems, thus it is desirable to locate these areas and remove them from further investigations. The objects that are, with very high probability, vessels form a binary image, with the breast region as binary one, and the background as well as the detected vessel areas as binary zero. This binary image is multiplied with the feature image from VarMet so that the vessel areas are removed, i.e. set to zero, in the feature image. The last step of the detector is the adaptive thresholding of the feature image (*box 4*). The adaptive thresholding algorithm starts by setting a (too) high threshold, calculate some parameters, and reduce the threshold before calculating the parameters again. This is repeated until some predefined conditions are fulfilled. All parameters are chosen empirically. More details on *box 2,3,4* can be found in Section III. The output from the adaptive thresholding algorithm is binary boxes containing the suspicious objects, as well as the coordinates of the boxes relative to I_{orig} .

To make sure that all μCa are detected, even very small ones, there are unfortunately a relatively large number of false detections after the adaptive thresholding. This can typically be μCa in vessels that were not removed by the vessel detection, connective tissue lines, edge between the pectoral muscle and the breast, dark holes in the tissue etc. The next task is to remove as much as possible of what is considered as false detections. *Box 5* is called *Classifier level 1* and consists of a collection of tests done to remove as many false detections as possible without removing any μCa . To be able to do this the classifier needs I_{orig} as input as well as the coordinates of the suspicious regions and the thresholded binary boxes of the suspicious regions. Details on the Classifier level 1 is found in Section IV. After Classifier level 1 many of the false detections are removed, but unfortunately not all. In addition there are many of the detected μCa that are single μCa , and these are always benign. The last step in the MammoScan μCaD , *box 6*, is a second classifier we call *Classifier level 2*. In this classifier the main task is to remove everything that is not a *cluster* of μCa , thus we remove both some false detections and many single μCa . This is further described in Section V.

All the parameters in the system are chosen empirically/trained using a training set of approximately 200 cases *not* included in the test sets in the experiments. The training set as well as the test sets include both cases with and without calcifications.

In a perfect world the output of the system would be solely clusters of μCa . Unfortunately, to make sure that true clusters of μCa are not discarded, some false detections and some single μCa manage to go through the system. This system shows, however, a very good true positive rate with relatively few false detections. Some experiments and results will be shown in Section VI.

III. DETECTION DETAILS

In this section the part inside the dotted box in Figure 1 is explained in more detail.

A. VarMet - Algorithm for detection of calcifications

Details of the VarMet algorithm can be found in [14]. In short the VarMet scheme relies on computing statistical features (local variances) based on the coefficients obtained from the multiscale median transform (MMT) [15] of the input image. In the following we first present the basic MMT algorithm. Thereafter, the extension is described.

The MMT algorithm: In the following we define the median transform of the image \mathbf{X} , with square kernel of dimensions $N \times N$, as $\text{med}_N\{\mathbf{X}\}$. The iteration index is denoted as j , and J is the number of resolution levels or scales. The MMT algorithm is then given as:

- 1) Let $N = 2s + 1$, $s = 1$.
- 2) Let $\mathbf{M}_j = \mathbf{X}$ with $j = 0$.
- 3) Compute $\mathbf{M}_{j+1} = \text{med}_{2s+1}(\mathbf{M}_j)$.
- 4) Compute the multiresolution coefficients $\mathbf{W}_{j+1} = \mathbf{M}_j - \mathbf{M}_{j+1}$.
- 5) Let $j \leftarrow j + 1$; $s \leftarrow 2s$. Go to Step 3 if $j < J$.

The MMT algorithm produces at each scale, j , a set $\{\mathbf{W}_j\}$ having the same number of pixels as the input image, \mathbf{X} . The input image \mathbf{X} can be expressed as the sum of all the scales and the smoothed residual image \mathbf{X}_J (equivalent to \mathbf{M}_J): $\mathbf{X} = \mathbf{X}_J + \sum_{j=1}^J \mathbf{W}_j$.

The VarMet scheme: The output from the original MMT algorithm is $\{\mathbf{W}_j\}_{j=1}^J$, and singularities in the images are often found using \mathbf{W}_J . However, in the application of detecting microcalcifications we found it more beneficial to proceed from the MMT algorithm with a variable $\mathbf{Y} = \mathbf{M}_1 - \mathbf{M}_3$. Let \mathbf{h} represent a vector with N^2 elements with value $h(i) = \frac{1}{N^2}$ for all i . In addition, we define

$$\mathbf{y}(m, n) = [y(m - K, n - K) \ y(m - K, n - K + 1) \dots y(m, n) \dots y(m + K, n + K)]^T,$$

where $K = \frac{N}{2}$. Then, an estimate of the local mean at position (m, n) in \mathbf{Y} may be written as the vector expression $\mathbf{h}^T \mathbf{y}(m, n)$. Finally we can estimate an image of local variances, \mathbf{V} . The variance at position $v(m, n)$ is estimated as:

$$v(m, n) = \frac{1}{N^2 - 1} \|\mathbf{y}(m, n) - \mathbf{h}^T \mathbf{y}(m, n)\|^2. \quad (1)$$

The variance image, V , is the output from the VarMet algorithm. It can be looked upon as a feature image, I_f , in which the microcalcifications, along with other strong singularities, are enhanced.

B. Vessel detection

Calcifications contained inside blood vessels are never malign, thus they can be discarded. Often this type of calcification is obvious to the human eye when looking at the entire breast, because it has a characteristic appearance that looks like a railway track through the breast tissue, as can be seen on the example image (mammogram) shown in Figure 2 a). Unfortunately, looking closely at such calcifications inside a vessel we see that they have an irregular appearance that can be mistaken for a cluster of malignant μCa , especially when only a small part of the vessel has calcifications. Since most CAD systems, including MammoScan μCaD , do local processing it is often difficult to distinguish the μCa inside the vessels from other types of μCa . The vessel detector in MammoScan μCaD finds relatively large and relatively straight objects with a detectable edge. Note that the edge of the breast/skin-line area is not searched for vessels for two reasons: There are no visible vessels in the skin-line, and the skin-line produces a lot of relatively long detectable edges that would be a problem for this vessel detector. The first step of the vessel detector is to perform an edge detection on the entire image, using Sobel operators [16], resulting in a binary image, I_{edge} . An example can be seen in Figure 2 in different parts of the process. The reduced original mammogram, I_{orig} , is shown in a), with the detail depicted in b), and the resulting I_{edge} in c). In this image the edges of the blood vessels are visible, but some places the line is broken. To fill the entire vessel, and also close some of the broken lines, a morphological dilation is performed on I_{edge} : $I_{obj} = I_{edge} \oplus Se_d$, where Se_d is a square 15×15 pixels structure element. The resulting I_{obj} for the example image can be seen in Figure 2 d).

Two properties of all the objects in I_{obj} are calculated: 1) The major axis length (in pixels) of the ellipse that has the same normalized second central moments as the object. 2) The eccentricity of the ellipse that has the same normalized second central moments as the object. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1 (0 and 1 are degenerate cases; an ellipse whose eccentricity is 0 is actually a circle, while an ellipse whose eccentricity is 1 is a line segment). To be classified as a vessel the object should be quite straight and relatively large, or quite large and relatively straight. Empirically the following decision rule is made:

- (If eccentricity > 0.92 AND major axis length > 90) OR (If major axis length > 150 AND eccentricity > 0.8) THEN the object is classified as a vessel.

The dilation was helpful in finding the blood vessels, but the vessel objects are now too large due to the dilation. Therefore an erosion is done before the objects are removed from the part of the mammogram that is investigated further: $I_{ero} = I_{obj} \ominus Se_e$, where Se_e is a square 11×11 pixels structure element. I_{ero} from the example is shown in Figure 2 e). The

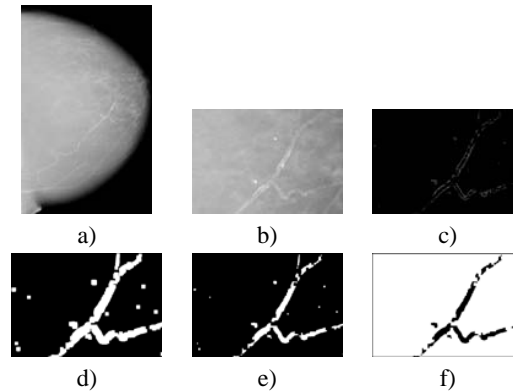


Fig. 2. a) Original image, I_{orig} . Detail from: b) I_{orig} , c) Edge image, I_{edge} , d) Dilated edge image containing objects, I_{obj} , e) Eroded object image, I_{ero} , f) Binary decision image containing areas classified as vessels, I_{ves}

objects from I_{ero} that were classified as a vessel according to the decision rule (performed on the objects from I_{obj}) should now be removed from the area that is searched for μCa . A binary image, I_{ves} , with the search area as binary one and the background as well as the detected blood vessels as binary zero is the result, and can be seen in Figure 2 f). To make sure that no areas that are *not* vessels are discarded the decision rule is quite strict. Thus not all vessels are detected. Some of them are successfully detected, however, and thereby removed from the rest of the CAD scheme. Due to the strict decision rule no objects that are *not* a vessel have been classified as one in our tests. There is another problem, however, that can be seen in the example of Figure 2. There are two single μCa lying very close to the vessel. After the dilation these become a part of the vessel object, so that they are removed as well as the vessel, as can be seen in Figure 2 f). A single μCa is always regarded as benign anyway, so if these μCa was not removed here they would have been removed in the Classifier level 2. Hopefully, if a cluster lies close to the vessel, not all of the μCa of the cluster would be removed. It is still worth noticing as a potential problem if a malign and small cluster should lie very close to a vessel.

C. Adaptive thresholding

The feature image from the VarMet algorithm, I_f , is multiplied pixel by pixel with the binary image I_{ves} , resulting in I_{f2} . The next step is to perform a thresholding on I_{f2} . There can be large variations from one image to the next, so a constant threshold will not be good enough. An adaptive thresholding strategy, to find a threshold for a particular image, was developed. The algorithm is presented in Algorithm 1.

The main idea is to find a threshold that gives a predefined number of detections in each image. Starting with a large threshold, the threshold is lowered iteratively. The algorithm is forced to give a number of detections for each image, even if many of the images have no μCa . Since some images can have many μCa and the most dangerous kind is usually the most subtle one, we have to make sure that we do further investigation on a relatively large number of the enhanced areas

```

1 Algorithm for adaptive thresholding
input :  $I_{f2}$ 
output:  $Obj, I_{det}$ 
2  $I_{fn} \leftarrow I_{f2} \times \frac{1}{\arg\max_{i,j}(I_{f2}(i,j))}$  //The feature image is normalized
3  $min_{areal} = 10, no_{max} = 60, no = 30, T = 1.1$ 
   //Initializing parameters (empirically decided)
4  $Th = 0.3, no_{areal} = 0$  //Initializing variables
5  $I_{th} \leftarrow I_{fn} > Th$ 
6  $Obj \leftarrow \{ (Objects\ in\ I_{th}) > min_{areal} \}$ 
7  $no_{areal} \leftarrow |Obj|$ 
8 while  $no_{areal} < no$  AND  $Th > T \cdot 10^{-3}$  do
9   if  $Th < T \cdot 10^{-2}$  then
10     $\delta \leftarrow 10^{-3}$ 
11   else if  $Th < T \cdot 10^{-1}$  then
12     $\delta \leftarrow 10^{-2}$ 
13   else
14     $\delta \leftarrow 5 \cdot 10^{-2}$ 
15   end
16    $Th \leftarrow abs(Th - \delta)$ 
17    $I_{th} \leftarrow I_{fn} > Th$ 
18    $Obj \leftarrow \{ (Objects\ in\ I_{th}) > min_{areal} \}$ 
19    $no_{areal} \leftarrow |Obj|$ 
20   if  $no_{areal} = 0$  then
21     exception handler
22   end
23 end
24 while  $no_{areal} < no$  do
25    $Obj \leftarrow Obj + \text{largest obj. from } I_{th} \text{ not yet in } Obj$ 
26    $no_{areal} \leftarrow |Obj|$ 
27 end
28 while  $no_{areal} > no_{max}$  do
29    $min_{areal} \leftarrow min_{areal} + 4$ 
30    $Obj \leftarrow \{ (Objects\ in\ I_{th}) > min_{areal} \}$ 
31    $no_{areal} \leftarrow |Obj|$ 
32 end
33  $I_{det} \leftarrow (I_{th} - \text{objects not in } Obj)$ 

```

Algorithm 1: Adaptive thresholding algorithm.

of the feature image I_f . The algorithm demands minimum 30 and maximum 60 objects per image, and these parameters are decided empirically so that all μCa in the training set were detected at this stage. The strategy of finding 30 to 60 objects in every image obviously leads to very many “false” detections at this stage. The Classifier level 1 removes the most obvious false detections and is the subject of the next section.

IV. CLASSIFIER LEVEL 1

The purpose of Classifier level 1 is to remove as many obviously false detections as possible without removing *any* μCa . This is done by a collection of more or less simple tests. The input to the test is the thresholded VarMet image, I_{det} , as well as the VarMet feature image, I_{f2} , and I_{orig} . A flow chart of the classifier level 1 is shown in Figure 3. Every object

in I_{det} is treated as a detection. A small area/box centered around each object is ran trough various tests, along with the corresponding area from I_{orig} , and sometimes in I_{f2} . The different tests will be explained in the rest of this section.

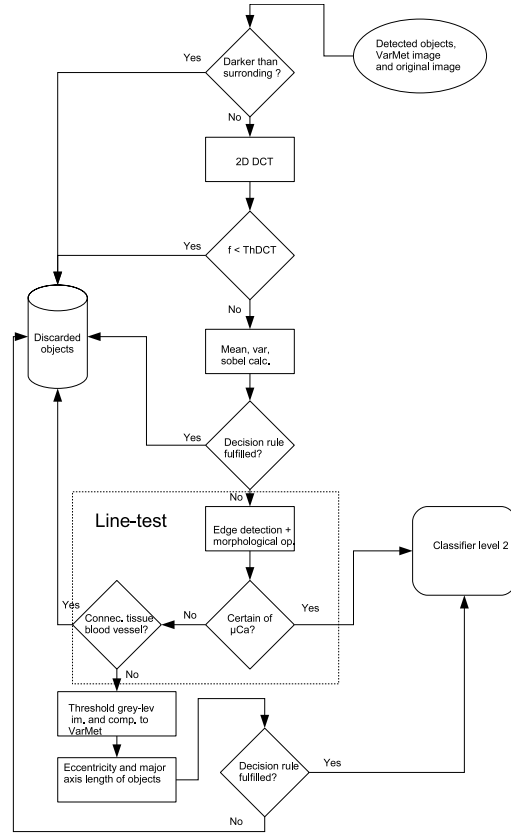


Fig. 3. Flow chart of classifier level 1

The first step is to test if the edge/area that the VarMet algorithm enhanced is a dark(er) area/spot. In that case it is *not* a μCa because a μCa is more dense than the breast tissue, thus it appears brighter on the X-ray image than the surrounding tissue. The second test is a simplified spectral analysis. A 2D Discrete Cosine Transform (DCT) is performed on a 64×64 block with the center of the detection as the center of the block: $\mathbf{W} = \mathbf{C} \cdot \mathbf{X} \cdot \mathbf{C}^T$, where \mathbf{C} is the DCT matrix, \mathbf{X} is the data matrix (image pixels in a 64×64 block), and \mathbf{W} is the coefficient matrix:

$$\mathbf{W} = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,64} \\ w_{2,1} & w_{2,2} & \dots & \\ \vdots & & \ddots & \\ w_{64,1} & \dots & & w_{64,64} \end{bmatrix} \quad (2)$$

One feature, f_{det} , is calculated based on the DCT coefficients as follows:

$$f_{det} = \frac{\sum_{i=1}^{64} \sum_{j=1}^{64} |w_{i,j}| - (|w_{1,1}| + |w_{1,2}| + |w_{2,1}|)}{|w_{1,2}| + |w_{2,1}|} \quad (3)$$

As seen from the equation, the DC coefficient ($w_{1,1}$) is not used when calculating the feature, f_{det} . If the image block \mathbf{X}

is more or less homogenous, the value of f_{det} will be quite low because most energy will be located in the lower frequencies, $w_{1,1}(DC), w_{1,2}, w_{2,1}$. If the image block contains a μCa (or other sharper edges/objects) there will be more energy in the higher frequencies as well, and f_{det} will be higher. Thus f_{det} is compared to an empirically decided threshold (8.5), and if $f_{det} < threshold$ the detection is discarded because it is so homogenous that it is not caused by a μCa .

This test is followed by another version of a homogeneity test. The first and second order moment for the image box around the detection is calculated (mean and variance). In addition an edge detection using Sobel operator [16] is performed, and the number of edge-pixels is counted. The detection is discarded as too homogenous if the following decision rule is fulfilled: ($edge-pixels < par1$) AND ($mean < par2$) AND ($var < par3$).

By now we have hopefully discarded all detections due to darker objects as well as all (most) of the detections where the image is too homogenous to be a μCa . There are, however, other types of objects that often lead to detection from the VarMet algorithm, as connective tissue appearing as bright lines or μCa inside blood vessels, also appearing as lines or “railway tracks”. The next test, called *line-test*, is referring to the part inside the dotted line in Figure 3. The line-test use edge detection as well as morphological operations on a box around the detection from I_{orig} to try to decide whether the detection is due to a connective tissue line, blood vessel, or a genuine μCa . The test gives three possible answers:

- 1) The detection is due to a genuine μCa , and no more testing in Classifier level 1 is necessary. The detection goes straight to Classifier level 2.
- 2) The detection is due to a connective tissue line or a blood vessel. The detection is discarded.
- 3) The detection is probably not due to a connective tissue line or a blood vessel, but it is uncertain if the detection is due to a genuine μCa . The detection is tested further in Classifier level 1.

A more detailed flow chart of the line-test algorithm is depicted in Figure 4. The last test of Classifier level 1 compares a thresholded version of a box, I_{box} , from a normalized version of the original image, I_{orig} , centered around the detection, to the corresponding area after the VarMet algorithm, i.e. in I_{det} . First a threshold, Th_{class1} , is found as a function of the maximum and minimum values in the box from I_{orig} . A binary image is found by thresholding: $I_{box} = I_{box} > Th_{class1}$. Each object in I_{box} is compared to the corresponding region in I_{det} . If the corresponding region in I_{det} is black (no object) that particular object from I_{box} is discarded. Two features are calculated for all objects > 1 pixel that overlaps (totally or partly) with an object in I_{det} : The eccentricity of the ellipse that has the same second order moment as the object, and the major axis length of this ellipse. A last decision rule was made: IF $eccentricity < Th_{ecc}$ AND $major\ axis\ length < Th_{mal}$ the object can be a μCa and should continue to classifier level 2. If not it is discarded. The output of Classifier level 1 is I_{det2} , where all the objects from I_{det} that were discarded as false are removed.

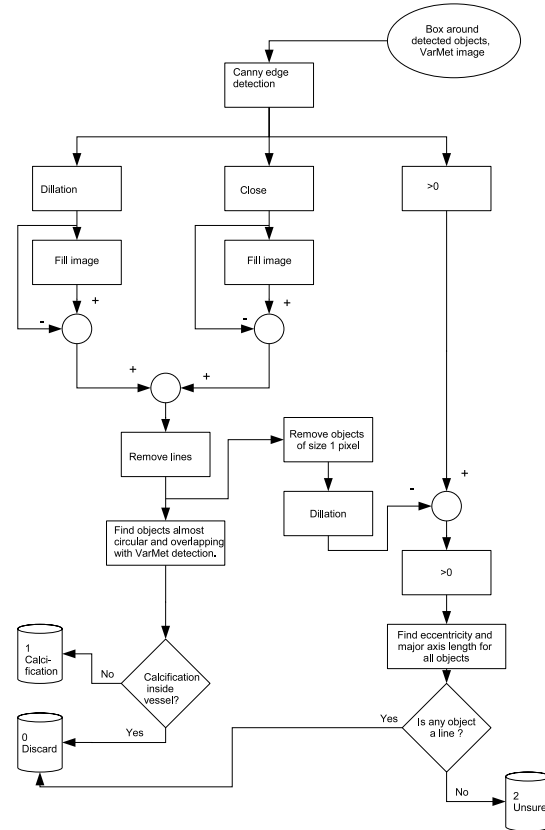


Fig. 4. Flow chart of the line-test algorithm.

V. CLASSIFIER LEVEL 2

After Classifier level 1 most (but unfortunately not all) false detections are removed. Thus most of the objects/detections that enters Classifier level 2 are μCa . Some of them are single, and thereby benign, calcifications. Clusters of μCa are much more suspicious. Some clusters are benign and others are malignant. This is, however, often hard to decide from the mammogram image, and a biopsy is often needed. In Classifier level 1 we defined our task as to separate μCa from false detections. We define our task in Classifier level 2 as to separate single μCa from clusters of μCa . Unfortunately, the single and benign μCa are not a homogenous group. Some of these single benign calcifications are very large and bright, and some of them are small and subtle. The large and bright type of calcification is always benign, thus before doing anything else in Classifier level 2 we want to remove this class. This can be done by a simple thresholding of the normalized version of a box from I_{orig} , centered around the detection. Morphological closing and opening is performed on the binary image for denoising. If there exist a large object centered in the box, the detection is (most probably) due to a large single calcification, and it is discarded as benign. The decision parameters are decided empirically so that only large, single benign calcifications are discarded on the training set.

The main part of the Classifier level 2 is based on a

texture classification technique where sparse representation using learned overcomplete frames/dictionaries is the main tool. In the following sections *sparse representations, learned dictionaries, and texture classification using sparse representations and learned dictionaries* will be explained briefly. More details can be found in [17], [18], [19].

A. Sparse representation and dictionary learning

Signal expansions using *frames*, or *overcomplete dictionaries*, can be regarded as generalizations of signal expansions based on transforms, filter banks, and wavelets. A finite *frame* or an *overcomplete dictionary* is a set of K vectors of dimension N , spanning the N -dimensional vector space, with $K > N$ [20]. In many applications we are interested in a *sparse approximation* of a signal rather than an exact representation. The reconstructed N -dimensional signal vector (often an approximation of \mathbf{x}) can be written as: $\hat{\mathbf{x}} = \mathcal{F}\mathbf{w} = \sum_{i=1}^K w_i \mathbf{f}_i$, where \mathcal{F} denotes the $K \times N$ matrix, $K \geq N$, with the vectors, \mathbf{f}_j , as its columns. Since the number of dictionary vectors in general is larger than the dimension, the vectors \mathbf{f}_j are not linearly independent. Finding the *best* sparse approximation is NP-hard [21], and suboptimal vector selection algorithms are used. We use Order Recursive Matching Pursuit (ORMP) throughout this work [22], [23]. Sparse representations can be useful for many applications, like signal compression [24], [25], feature extraction, denoising [26], blind source separation (of more sources than mixtures) [27], [28], [29] and texture classification [17].

The quality of a sparse representation of a class of signals is highly dependent on the dictionary. If the dictionary is well suited for the class of signals, a good quality of fit (low mean square error) can be obtained with a very sparse representation (approximation). Thus finding a good dictionary for a specific class of signals is important. The algorithm used for learning dictionaries in this work is the algorithm called Method of Optimal Directions (MOD) by Engan et al. [30], [31]. This algorithm is the core algorithm of a larger family of algorithms named Iterative Least Squares based Dictionary Learning Algorithms (ILS-DLA) by Engan et al. [18]. MOD is an algorithm for learning unrestricted block based dictionaries, whereas ILS-DLA includes block based and overlapping dictionaries, with or without different kinds of restrictions/constraints.

B. Texture classification using sparse representation

In image processing the texture of a region describes the pattern of spatial variation of grey tones in a neighborhood where the neighborhood is small compared to the region. By definition, texture classification is to identify the texture class in a region, whereas texture segmentation means finding a boundary map between different texture regions of an image [32]. Still there is an obvious borderline here since classification can be used for segmentation. We use the term texture classification in the following even though the classification often leads to a segmentation. The algorithm used in this work is similar to the Frame Texture Classification Method (FTMC) of [17], and the main idea of the algorithm is

as follows: Each pixel in the texture image should be classified as belonging to a certain texture/class. For each pixel in the (possible preprocessed) texture image a vector, \mathbf{y} , is made from the specific pixel and a number of neighbourhood pixels. The size and shape of the neighbourhood is predefined. For each possible texture class, a dictionary, \mathbf{F}_i , is learned using a training set, where the training vectors are constituted the same way as \mathbf{y} . The vector to be classified, \mathbf{y} , is then represented sparsely (with a predefined sparseness factor, i.e. number of nonzero entries in \mathbf{w}) by all the different dictionaries, and the different residuals are calculated:

$$r_i = \|\mathbf{y} - \mathbf{F}_i \mathbf{w}_i\|. \quad (4)$$

This is done for every single pixel in the (possible preprocessed) texture image, and results in a new image, a residual image, of the same size as the texture image for each class, \mathbf{R}_i . These residual images are now smoothed using a Gaussian filter, as done in FTMC, resulting in \mathbf{R}_s . We use a 5×5 Gaussian filter with $\sigma = 8$. The last part is a simple classification: for each single pixel in the texture image, the class is decided as the class with the lowest value at the corresponding position in \mathbf{R}_s .

C. Texture classification in Classifier level 2

The classification scheme in Classifier level 2 has two classes: μCa and *False*. Before the dictionaries can be learned, a training set corresponding to the classes; μCa and *background* (one training set per. class), has to be arranged. Parts of training images corresponding to the different classes are selected carefully. Training vectors are arranged from a neighborhood around each pixel in the same manner as is done when using the dictionaries for classification. We use a square 5×5 neighborhood in this work, and no preprocessing of the image data is done. The dictionary learning is done using MOD [31], [18].

The input to Classifier level 2 is I_{det2} from Classifier level 1 along with the VarMet feature image, I_{f2} , and I_{orig} . For each detection in I_{det2} , a box from I_{orig} , corresponding to the area around the detection, is investigated. 75 pixels (7.5 mm) in each direction of the edges of the detection makes the box, thus the size of the box varies since the size of the detection varies. The image box is transformed into test vectors using a square 5×5 neighborhood around every pixel. The classification is done, as described in Section V-B, resulting in a binary image, I_{cb} , at the same size as the image box where "1" corresponds to the class μCa and "0" corresponds to the class *background*.

Let I_{boxf2} be the corresponding box of the VarMet image. I_{boxf2} is thresholded with a threshold as a function of the max and min value in I_{boxf2} , and the resulting binary image is 5×5 median filtered, giving $I_{boxf2_{bm}}$. If, for an image box, there are very many objects in $I_{boxf2_{bm}}$ (> 30) this is a good indication that the detection is not due to a (possible cluster of) μCa . This is because a μCa would be so much more enhanced in I_{boxf2} so that a thresholding based on min and max values would result in only the μCa being present in $I_{boxf2_{bm}}$. Thus if this situation occurs, the detection is discarded. Else, I_{cb}

is compared to $I_{boxf2bm}$, and whenever there is an object in I_{cb} not overlapping with an object in $I_{boxf2bm}$, the object is removed from I_{cb} resulting in I_{cb2} .

If there are no objects left in I_{cb2} , that particular detection is discarded. We are now quite sure that the remaining objects in I_{cb2} are due to calcifications. The last question is if there is *one* single μ Ca, which we then want to discard as benign, or if there are several μ Ca which can be a dangerous cluster. If there are several (more than one) object at this point, this detection is regarded as a cluster of μ Ca. However, it can not be assumed that only *one* object necessarily means only one calcification. The μ Ca can have very different sizes, and they can be lying so close so that a small cluster can become *one* object. The area of the object (number of pixels) is found, A_{obj} , as well as the area of the convex hull of the object, A_{chull} . Since a μ Ca is more or less circular, A_{obj} and A_{chull} will be very close in value if there is only *one* μ Ca in the object. If the object is truly a cluster, the convex hull will be larger than the area of the object, sometimes much larger. Since the size of a μ Ca can vary, we test both the absolute difference, as well as the ratio between these two areas. If the difference is more than a certain threshold in both these tests, the object is considered a cluster. If not it is considered a single μ Ca, and discarded as benign.

VI. EXPERIMENTS AND RESULTS

All experiments are performed on true digital mammograms supplied by the Breastdiagnostic Center of Stavanger University Hospital (SUS). The images are originally of size 2294×1914 with a resolution of 100 microns (0.1 mm/pixel) and recorded on a GE Senograph DS as a part of the daily clinical work at SUS. All CAD results are evaluated carefully by radiologists from SUS, and all diagnosis (benign/malign) of the test sets are verified by biopsy.

All the parameters in the system are chosen empirically/trained using a training set of approximately 200 cases *not* included in the test sets in the experiments. The training set as well as the test sets include both cases with and without calcifications.

A. A case study

To illustrate the different parts of the system an image with one cluster of μ Ca is shown in Figure 5 and 6 at various stages through the system. In Figure 5 b) the corresponding feature image, I_{f2} , is depicted with the cluster marked. From this image it can be seen that the feature extraction method, VarMet, does indeed emphasize μ Ca. The image resulting after I_{f2} has been adaptively thresholded, I_{det} , is seen in Figure 5 c). There are a number of detections from inside the cluster (inside the square box that is there for illustration), as well as a number of detections from the edge area of the breast, close to the skin-line, in total 53 detections. Every single one of these detections are treated separately in Classifier level 1, see Section IV. 46 detections are removed in Classifier level 1. 18 are removed on the first test. 21 is removed by the DCT test, and 1 by the second homogeneity test using second order moments and edge detection. None is removed

as being calcifications inside vessels, 2 are removed as being connective tissue lines, and finally, 4 is removed on the last test of Classifier level 1.

After all the false detections discovered in Classifier level 1 are removed we get I_{det2} depicted in Figure 5 d). In addition to 4 detections in the area of the cluster of μ Ca, i.e. inside the square marked on the image, there are three more small detections, let them be called detection 5, 6, and 7, in I_{det2} . All these 7 detections will proceed to Classifier level 2. The size of detection 5, 6, and 7 are all only a couple of pixels, thus they are hard to see on this reduced size image. A zoomed version of the breast tissue centered around detection 5 is seen in Figure 6 b). This is, no doubt, a false detection that hopefully will be removed in Classifier level 2. Detection 6 and 7 are quite close to each other, so a zoomed version of the breast tissue surrounding both these two detections is seen in Figure 6 c), the arrows pointing at the two detection points. The uppermost arrow corresponds to detection 6, and the lowermost to detection 7. For illustration a contrast enhanced version of the same tissue area is seen in Figure 6 d). The two detection points are slightly denser than the surrounding tissue (the gray-level is higher), and this is why they were enhanced by the VarMet algorithm. They *could* be caused by very small μ Ca, but in that case they are single μ Ca (not close enough to be considered a cluster, i.e. more than approx. 7.5 mm apart) and should be discarded in Classifier level 2. However, both detections looks more like false detections (not μ Ca). Detections 1, 2, 3, and 4 are treated separately, and all classified as cluster of μ Ca by Classifier level 2. There are three different important variables (binary images) calculated in Classifier level 2: 1) The thresholded version of a gaussian filtered locally normalized version of the VarMet image, $I_{boxf2bm}$, 2) the result after classifying using learned dictionaries, I_{cb} , 3) the result after the objects in I_{cb} not overlapping with objects in $I_{boxf2bm}$ are removed from I_{cb} , I_{cb2} . The three binary images are all depicted in Figure 7. When there are more than one object in I_{cb2} the detection is classified as being part of a μ Ca cluster. In the end these three detections will be marked as *one* cluster since the detections are so close that they obviously come from the same cluster (less than approx. 7.5 mm apart). Sometimes the system might miss some of the μ Ca in a cluster, but there is a very high probability that at least some of the μ Ca will be detected. As in this case we have 4 true detections (one of them covering two close μ Ca) within the cluster of (some of them very small) μ Ca (see Figure 6 a). Following detections 5, 6, and 7 through Classifier level 2, we find that all detections are discarded due to the large number of elements in $I_{boxf2bm}$ (see Section V for explanation), depicted for the different detections in Figure 8. I_{cb} and I_{cb2} are also shown for illustration purposes, however they will not be calculated for these detections since they are discarded due to the large number of objects in $I_{boxf2bm}$.

B. Volume studies

The digital mammograph at SUS was mainly used for clinical investigations at the time of our study, thus we have not tried our system on vast amounts of screening data where most of the breasts are normal.

Volume study 1 was done on images from 51 selected patients, with in average approx. 3 images pr. patient. The number of images pr. patient varies from 1 to 8, giving a total of 155 images. All of the patients have μCa in some form. Some of the patients have only spread out/single benign μCa (17), others have clusters of μCa (34). Of the latter type some have been tested (cytological or histological) to be malignant (12), the rest is benign (22). Some patients have more than one cluster (malignant or benign), and the clusters can be visible on several of the images since there are often images from different views of the breast. Since these data are from clinical investigation there are often images zoomed in on the region of interest, however these type of images are *not* a part of this test. Our 155 images are all images of a complete breast, but from different views (cranio caudal, medio lateral oblique).

Ideally a CAD system should only mark *malignant* clusters of μCa , but it is often hard to distinguish between benign and malignant clusters (biopsy is often performed as to be sure of the diagnosis). Thus we aim for our system to detect *all clusters of μCa* . All the single/spread out benign μCa should be removed by classifier level 2. However, not all of the single/spread out μCa are removed and it is difficult to decide if these should count as *false* detections. Therefore we have calculated the number of *unwanted* detections pr. image, including all these single/spread out μCa , as well as the number of *false* detections pr. image, including detections that have nothing to do with μCa as well as detections due to μCa inside blood vessels. However, from a clinical (user) point of view, all the unwanted detections will count as false detections.

The question of sensitivity, i.e. the number of true detections relative to the number of clusters, is not straightforward either. In most cases there are more than one image taken from different views. Sometimes the cluster is more pronounced in one view than in another, thus sometimes it is detected perfectly in one view and lost in the other. We have calculated the number of true detections relative to the number of clusters looking at all images independently of the case, and the number of true detected clusters relative to the total number of clusters looking at a case (patient) at the time. The latter means that if a cluster is detected in one view, it is counted as a success even if it is (possibly) lost in another. The results from this study is seen in Table I.

TABLE I
RESULTS AFTER VOLUME STUDY 1.

True detections (image based) 84 %	True detections (case based) 94 %
Unwanted detections pr. image 1.1	False detections pr. image 0.3

A different study, volume study 2, was done on 39 different cases with a total of 139 images. This is a mixture of images that contain clusters of μCa (some malignant and some benign), images of normal breasts, and images that contains soft tissue lesions. However, detecting soft tissue lesions is a completely different task and is *not* included in the system described in this paper. The results from this study (for clusters

of μCa) is seen in Table II.

TABLE II
RESULTS AFTER VOLUME STUDY 2.

True detect. total (image based) 95 %	True detect. malign (image based) 90 %
True detect. malign (case based) 100 %	Unwanted detect. pr. image 1.3

C. Comparing with reported results

Many reported detection schemes are not complete CAD systems. Some of them only focus on the enhancement of μCa , which often will be one of the first steps in a complete CAD scheme [9], [8]. Yu and Guan [5] test their system on a mixture of training data and test data, which make their results uninteresting. Other reported schemes appears to be complete systems, but the testing is done on Region Of Interest (ROI) of the image, i.e. a $N \times M$ area around the μCa cluster, considerable smaller than the entire image [3], [13]. Reporting false positive pr. image is thereby *not* comparable to a complete CAD system where whole images are used. There are a lot of details to take into account when dealing with complete real world images that can be overlooked when dealing with carefully selected ROI images. For example the skin-line area that is often enhanced when calcifications are enhanced, or the calcifications inside blood vessels that is a significant problem since it usually leads to many "false positives" (this is actually μCa , but considered false positives because calcifications inside blood vessels are always benign. In addition there can be a lot of this in some breasts and the radiologists would be disturbed by a CAD system that marks all such calcifications).

Almost all reported tests are conducted on digitized mammograms, originally from film. As the fully digital mammograms are getting more used in hospitals (and is, no doubt, the future), we wanted to train and test our system on fully digital images. There are no public database of true digital mammograms, and we got our data from Stavanger University Hospital (SUS) in Norway. This makes direct comparison impossible, but some comparable numbers can be suggested, and are listed in Table III.

TABLE III
COMPARING WITH REPORTED RESULTS. FOR EL-NAQA, YANG ET. AL. THE SENSITIVITY IS DEPENDENT ON HOW THEY DEFINE A CLUSTER OF μCa . FOR R2 THE SENSITIVITY INCLUDES SOFT TISSUE LESIONS AS WELL, AND IS CALCULATED CASE BASED. A CASE BASED SENSITIVITY NUMBER FOR CLUSTERED MICROCALCIFICATIONS CAN BE FOUND TO BE 98 %, BUT NO FALSE DETECTION RATE IS GIVEN RELATIVE TO THIS NUMBER.

Author	Sensitivity	False detection pr. image
Horváth, Vályon et. al. [12]	> 90 %	2-3
El-Naqa, Yang et. al. [6]	84-94 %	approx. 1
Veldkamp and Karssemeijer [4]	90 %	approx. 1
Company	Sensitivity (total)	False detection
R2 [33]	91 %	1.5 pr. normal case

VII. DISCUSSION AND CONCLUSION

We have presented a complete system for detection of clustered μCa in digital mammograms, including: breast detector/bounding box, vessel detector, a novel method for feature extraction for calcification detection (VarMet), adaptive thresholding of feature image, classifier level 1 for discarding everything but calcifications using standard image processing techniques, classifier level 2 for discarding single (thus benign) calcifications and remaining false detections. The classifier level 2 uses learned dictionaries and sparse representation as a novel classifier technique for microcalcifications. The scheme, named MammoScan μCaD , is trained, and tested carefully by radiologists, on true digital data from clinical examination at SUS. The lack of a large database of true digital data makes our test set a bit small, but a couple of different tests are conducted and reported. The proposed scheme works well compared with the reported results from different schemes. Future work is to incorporate a mass detection/classification part to fulfill the system.

REFERENCES

- [1] www.kreftregisteret.no.
- [2] H. P. Chan, K. Doi, C. Vyborny, R. Schmidt, C. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. Macmahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. the potential of computer-aided diagnosis," *Investigative Radiology*, vol. 25, no. 10, pp. 1102–1110, 1990.
- [3] H. D. Cheng, Y. M. Lui, and R. Freimanis, "A novel approach to microcalcification detection using fuzzy logic technique," *IEEE Trans. Medical Imaging*, vol. 17, no. 3, pp. 442–450, 1998.
- [4] W. J. H. Veldkamp and N. Karssemeijer, "An improved method for detection of microcalcification clusters in digital mammograms," in *Proc. of SPIE International Symposium Medical Imaging, Image Processing 1999*, vol. 3661, San Diego, Ca, USA, May 1999, pp. 512–522.
- [5] S. Yu and L. Guan, "A cad system for the automatic detection of clustered microcalcifications in digitized mammogram films," *IEEE Trans. Medical Imaging*, vol. 19, no. 2, pp. 115–126, 2000.
- [6] I. El-Naqa, Y. Yang, M. W. N.P., Galatsanos, and R. Nishikawa, "A support vector machine approach for detection of microcalcification," *IEEE Trans. Medical Imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.
- [7] R. Nishikawa, M. Giger, K. Doi, C. Vyborny, and R. A. Schmidt, "Computer aided detection of clustered microcalcifications in digital mammograms," *Med. Biol. Eng. Comput.*, vol. 33, pp. 174–178, 1995.
- [8] G. Lemaire, K. Drouiche, and J. DeConinck, "Highly regular wavelets for the detection of clustered microcalcification in mammograms," *IEEE Trans. Medical Imaging*, vol. 22, no. 3, pp. 393–401, 2003.
- [9] P. Heinlein, J. Drexler, and W. Schneider, "Integrated wavelets for enhancement of microcalcifications in digital mammography," *IEEE Trans. Medical Imaging*, vol. 22, no. 3, pp. 402–413, 2003.
- [10] P. Zhang, B. Verma, and K. Kumar, "A neural-genetic algorithm for feature selection and breast abnormality classification in digital mammography," in *Proceedings of IEEE International Joint Conference on Neural Networks, 2004*, vol. 3, 2004, pp. 2003–2008.
- [11] H. Yoshida, "Matching pursuit with optimally weighted wavelet packets for extraction of microcalcifications in mammograms," *Applied Signal Processing*, vol. 5, no. 3, pp. 127–141, 1998.
- [12] G. Horvath, J. Vallyon, G. Strausz, M. Pataki, L. Sragner, L. Lasztovicza, and N. Szekely, "Intelligent advisory system for screening mammography," in *Proceedings of IEEE Instrumentation and Measurement Technology Conference, IMTC 04*, vol. 3, May 2004, pp. 2071–2076.
- [13] D. A. A. Vega-Corona, "Cad system for identification of microcalcifications in digitized mammography applying grnn neural networks," in *Proceedings of World Automation Congress, 2004*, vol. 17, 2004.
- [14] K. Engan and T. O. Gulsrud, "VarMet - a method for detection of image singularities with application to mammography," *WSEAS Transactions on Signal Processing*, vol. 2, no. 9, pp. 1222–1229, Sept. 2006.
- [15] J.-L. Starck, "Non Linear Multiscale Transforms," in *Multiscale and Multiresolution Methods*, T. Barth, T. Chan, and R. Haimes, Eds. Springer-Verlag, 2002, pp. 239–278.
- [16] R. Gonzalez and R. Woods, *Digital Image Processing*. USA: Addison Wesley, 1993.
- [17] K. Skretting and J. H. Husøy, "Texture classification using sparse frame-based representations," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. Article ID 52561, 11 pages, 2006, doi:10.1155/ASP/2006/52561.
- [18] K. Engan, K. Skretting, and J. Husøy, "A family of iterative LS-based dictionary learning algorithms, ILS-DLA, for sparse signal representation," *Digital Signal Processing, Elsevier*, vol. 17, no. 1, pp. 32–49, 2007, doi:10.1016/j.dsp.2006.02.002.
- [19] K. Engan, K. Skretting, J. Herredsvella, and T. Gulsrud, "Frame texture classification method (FTCM) applied on mammograms for detection of abnormalities," *WASET International Journal of Signal Processing (IJSP)*, vol. 4, no. 2, 2007, ISSN = 1304-4478, http://www.waset.org/ijsp/v4/v4-2-16.pdf.
- [20] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs: Prentice-Hall, 1995.
- [21] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM journal on computing*, vol. 24, pp. 227–234, Apr. 1995.
- [22] M. Gharavi-Alkhansari and T. S. Huang, "A fast orthogonal matching pursuit algorithm," in *Int. Conf. on Acoust. Speech and Signal Proc.*, Seattle, U.S.A., May 1998, pp. 1389–1392.
- [23] S. F. Cotter, J. Adler, B. D. Rao, and K. Kreutz-Delgado, "Forward sequential algorithms for best basis selection," *IEEE Proc. Vis. Image Signal Process*, vol. 146, no. 5, pp. 235–244, Oct. 1999.
- [24] K. Skretting, K. Engan, and J. Husøy, "Ecg compression using signal dependent frames and matching pursuit," in *Proc. Int. Conf. Acoust. Speech, Signal Proc.*, Philadelphia, Pennsylvania, USA, 2005.
- [25] A. Rahmoune, P. Vanderghenst, and P. Frossard, "MP3D: Highly scalable video coding scheme based on matching pursuit," in *Proc. Int. Conf. Acoust. Speech, Signal Proc.*, Montreal, Canada, May 2004.
- [26] K. Engan, K. Skretting, and J. Husøy, "Denoising of images using signal dependent frames and matching pursuit," in *Proc. Int. Conf. Acoust. Speech, Signal Proc.*, Philadelphia, Pennsylvania, USA, 2005.
- [27] T. W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE Signal Processing Letters*, vol. 6, no. 4, pp. 87–90, Apr. 1999.
- [28] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, 2001.
- [29] M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation*, no. 13, pp. 863–882, 2001.
- [30] K. Engan, S. O. Aase, and J. H. Husøy, "Method of optimal directions for frame design," in *Proc. ICASSP '99*, Phoenix, USA, Mar. 1999, pp. 2443–2446.
- [31] —, "Multi-frame compression: Theory and design," *Signal Processing*, vol. 80, pp. 2121–2140, Oct. 2000.
- [32] M. Tuceryan and A. K. Jain, "Texture analysis," in *Handbook of Pattern Recognition and Computer Vision*, C. H. Chen, L. F. Pau, and P. S. P. Wang, Eds. Singapore: World Scientific Publishing Co, 1998, ch. 2.1, pp. 207–248.
- [33] www.r2tech.com.



Kjersti Engan (M'01) was born in Bergen, Norway, in 1971. She received the Ing. (B.E.) degree in electrical engineering from Bergen University College in 1994 and the Siv.Ing. (M.S.) and Dr.Ing. (Ph.D.) degrees in 1996 and 2000 respectively, both in electrical engineering from the University of Stavanger, Stavanger, Norway.

She is an associate professor with the department of Electrical and Computer Engineering at University of Stavanger, Norway. Her research interests include signal and image representation and compression, sparse signal representation, image analysis, and medical signal and image segmentation and classification.

Dr. Engan is a member of NORSIG (Norwegian Signal Processing Society) and currently on the board of NOBIM (Norwegian Image and Pattern Recognition Society).

Thor Ole Gulsrud was born in Stavanger, Norway, in 1965. He received the Ing. (B.E.) degree from the University of Stavanger, Stavanger, Norway, in 1986 and the Siv.Ing. (M.S.) degree from the University of Strathclyde, Glasgow, Scotland, in 1991, both in electrical engineering. In 2001 Dr. Gulsrud received his Ph.D. degree in digital image processing from Aalborg University, Aalborg, Denmark.

In the period 1999-2006 he was an associate professor with the department of Electrical and Computer Engineering at the University of Stavanger, Norway. He is currently a technical product manager at Roxar Flow Measurement AS, Stavanger, Norway. His research interests include image analysis, medical image segmentation and classification, and ultrasonic techniques for fluids characterization.

Dr. Gulsrud is a member of NOBIM (Norwegian Image and Pattern Recognition Society).

Barbro Furebotten Iversen was born 1948 in Saltdal, Norway. She got her MD 1975 in Lbeck, Germany and the speciality in Radiology 1984 at Stavanger University Hospital.

She has been working with mammography since 1986. She is the head of the Breastdiagnostic Center at The University Hospital of Stavanger.

Since 1998 she has been a member of different national groups looking into the use of digital mammography including use of CAD. Some of them resulting in published reports: The Norwegian Cancer registry: report 1-2007 and 2-2007.

Liv Eriksen was born 1942 in Fjaler, Norway. She got her MD in Oslo, Norway 1970. Specialist in Radiology 1982 at Stavanger University Hospital.

She has been working with mammography since 1985 and was the head of the Breastdiagnostic Center at The University Hospital of Stavanger until 2006.

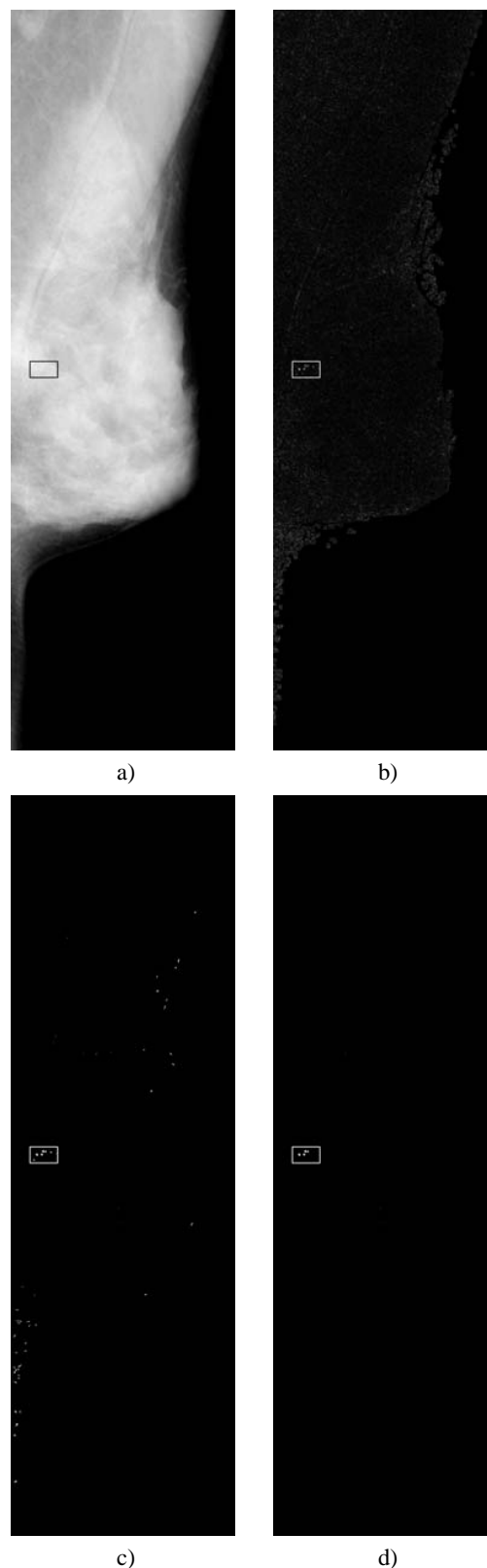


Fig. 5. a) Original image in case study, b) Feature image, I_{f2} , c) Thresholded image, I_{det} , d) After Class. level 1, I_{det2}

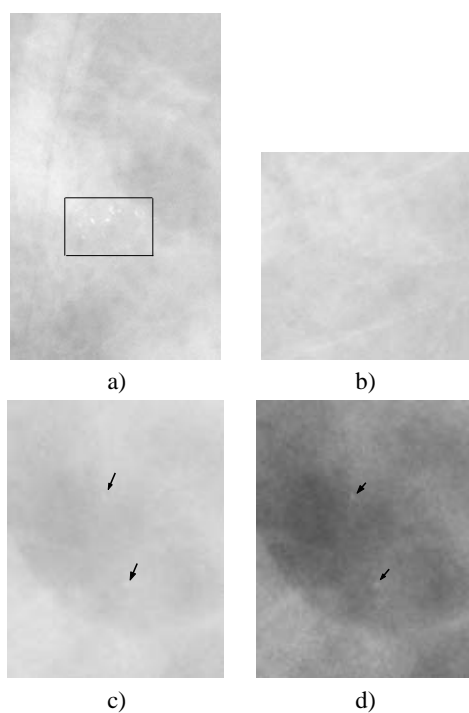


Fig. 6. a) Zoomed version of clustered μCa , b) detection 5, c) and d) detection 6 and 7.

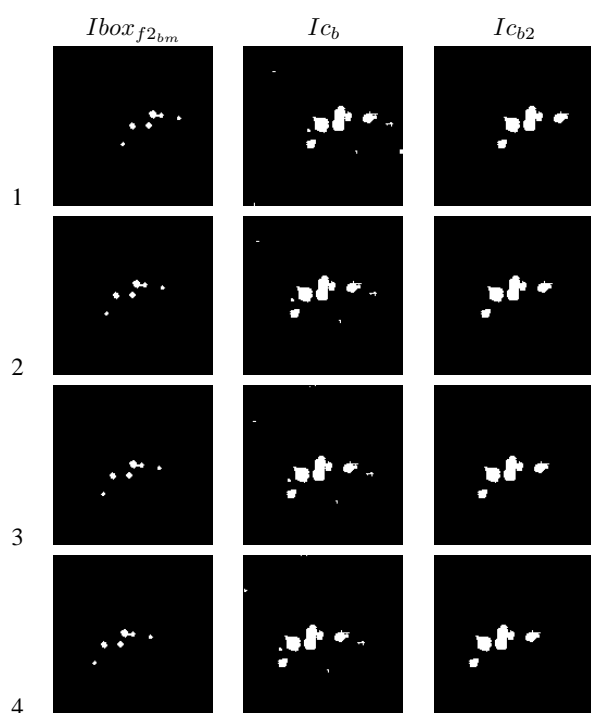


Fig. 7. Classifier level 2 results for the true detections. First column shows $I\text{box}f_{2bm}$ for the detections indicated by the number on the rows. Second column shows the corresponding Ic_b , and last column shows Ic_{b2} .

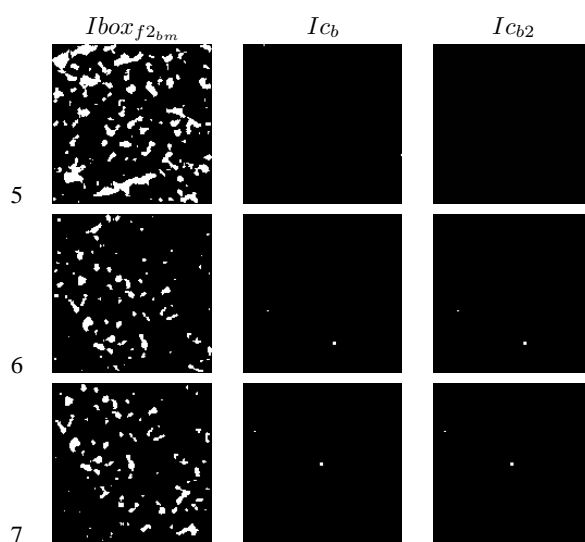


Fig. 8. Potential classifier level 2 results for the false detections. First column shows $I\text{box}f_{2bm}$ for the detections indicated by the number on the rows. Second column shows the corresponding Ic_b , and last column shows Ic_{b2} (if these had been calculated, which they will not be in this case since the detections are discarded after calculating $I\text{box}f_{2bm}$).