Online ISSN 1848-3380, Print ISSN 0005-1144
ATKAFF 53(4), 362–372(2012)

Mario Muštra, Mislav Grgić, Krešimir Delač

Breast Density Classification Using Multiple Feature Selection

DOI 10.7305/automatika.53-4.281
UDK 004.932.02:618.19-073.7; 025.4
IFAC 2.8; 1.1.8; 5.9.2

Original scientific paper

Mammography as an x-ray method usually gives good results for lower density breasts while higher breast tissue densities significantly reduce the overall detection sensitivity and can lead to false negative results. In automatic detection algorithms knowledge about breast density can be useful for setting an appropriate decision threshold in order to produce more accurate detection. Because the overall intensity of mammograms is not directly correlated with the breast density we have decided to observe breast density as a texture classification problem. In this paper we propose breast density classification using feature selection process for different classifiers based on grayscale features of first and second order. In feature selection process different selection methods were used and obtained results show the improvement on overall classification by choosing the appropriate method and classifier. The classification accuracy has been tested on the mini-MIAS database and KBD-FER digital mammography database with different number of categories for each database. Obtained accuracy stretches between 97.2 % and 76.4 % for different number of categories.

Key words: Breast Density, Feature Selection, Haralick Features, Soh Features, Classification

Klasifikacija dojki prema gustoći izborom značajki. Mamografija je rendgenska metoda koja daje dobre rezultate pri slikanju dojki koje imaju manju gustoću, dok joj osjetljivost značajno opada pri snimanju dojki veće gustoće i time može doći do lažno pozitivnih rezultata. Poznavanje gustoće dojke može biti korisno kod algoritama za automatsku detekciju zbog mogućnosti određivanja praga odluke na osnovi tog znanja. S obzirom na to da ukupni intenzitet pojedinog mamograma nije izravno povezan s gustoćom, odlučili smo se promatrati gustoću kao problem klasifikacije teksture. U ovom radu predlažemo klasifikaciju dojki prema gustoći izborom izdvojenih značajki intenziteta prvog i drugog reda za različite klasifikatore. Za određivanje prikladnih značajki koristili smo različite metode i tako dobivene značajke pokazale su bolju točnost klasifikacije za odabrane klasifikatore. Točnost klasifikacije testirali smo na bazi mamografskih slika mini-MIAS i bazi digitalnih mamografskih slika KBD-FER s različitim brojem kategorija u koje su slike bile podijeljene. Postignuta točnost klasifikacije proteže se između 97,2 % i 76,4 % za različit broj kategorija u koje su mamogrami podijeljeni.

Ključne riječi: Gustoća dojke, izbor značajki, Haralickove značajke, Sohove značajke, klasifikacija

1 INTRODUCTION

Breast density is an important measure which shows the possibility for the detection of abnormalities in mammograms. Higher breast density usually indicates a higher possibility for the presence of malignant tissue. Higher breast density tends to mask abnormal tissue and because of that it is more difficult to detect malignant tissue [1]. Mammograms are captured in two different projections; Medio-Lateral Oblique (MLO) and Cranio-Caudal (CC). Because of the projection nature of mammograms, higher breast density can present a severe detection problem, because dense tissue layers produce a masking effect. Breast density is usually correlated with the woman's age in a way that younger women usually have denser breasts than older women. In the past, there were many different approaches in dividing breast tissue into well distinguished categories. Wolfe was one of the first researchers who presented the correlation between different breast densities and the probability for the development of breast cancer [2]. He proposed that breast should be divided into four density categories before BI-RADS division was established. Today radiologists classify breast according to their density into four categories. Categories according to the American College of Radiology (ACR) BI-RADS are [3]:

- BI-RADS I: almost entirely fatty breast (0 - 25 %);
- BI-RADS II: some fibroglandular tissue (26 % - 50 %);
- BI-RADS III: heterogeneously dense breast (51 % - 75 %);
- BI-RADS IV: extremely dense breast (76 % - 100 %).

A human observer can distinguish different structures very well without the information of their overall brightness. In automatic breast density classification it is important to decide which parameters give the best separability between categories. Because of different imaging conditions and different breast volume, two images from different categories can have very similar intensity properties. This makes the breast density classification somewhat more difficult to perform. Therefore, to be able to achieve as good as possible classification results it is necessary to extract as many image features as possible and then try to select the ones that are least correlated with each other and produce best classification results. With a large number of features, manual feature selection is not possible and it is necessary to use an automatic feature selection method. There have been many attempts to perform automatic breast density classification with various achieved results on different datasets. The usage of publicly available datasets is very important because it makes future results from different researchers comparable to the presented ones. We think that two publicly available mammography databases are the best ground to make objective comparisons between presented methods. Those two databases are the MIAS database [4] and the DDSM database [5]. The mini-MIAS database is publicly available and it contains the same number of images as the original MIAS database. The difference is that the mini-MIAS images are resized to the image size of 1024×1024 pixels with 8 bits per pixel. For testing the performance of our classification method we have used mini-MIAS and KBD-FER mammograms. The reason behind that is a very large number of DDSM images which require preselection in order to make a decent comparison. In our opinion, the only fair comparison between two or more methods can be achieved when using completely the same dataset. KBD-FER consists of digital mammograms stored according to the DICOM standard, contains images of high (diagnostic) quality and is not yet publicly available. Comparison of mini-MIAS and KBD-FER should give a clue whether image quality influences algorithm's performance or not.

In this paper we propose an automatic breast density classification method. We are using different statistical features extracted directly from the region of interest (ROI), from histograms and from gray level co-occurrence matrices (GLCMs). From GLCMs we have extracted 18 Haralick and Soh [6,7] texture features for each direction at four different distances trying to cover as much as possible variation in spatial intensity. After the feature extraction we have used different feature selection algorithms to select features that give the best classification results and finally tested them with different classifiers. The feature evaluation, selection and classification were done using WEKA data mining software [8]. Obtaining the best feature selection results is not possible solely by selecting the features which provide best individual classification results but with the optimal combination of features, as explained in [9].

This paper is organized as follows. Section II gives the overview of the previous work in the field of automatic breast density classification. In Section III the entire preprocessing stage for images from both used databases is described as well as the feature extraction selection and classification. Section IV brings the experimental results and the discussion of the results. Section V draws the conclusion.

2 OVERVIEW OF THE PREVIOUS WORK

There are many papers that consider automatic breast density classification with different approaches in feature extraction and used classifiers. Karssemeijer used 615 digitized images from publicly unavailable Nijmegen database for testing the performance of the proposed method for breast density detection [10]. As a ground truth he used radiologist's classification. Because images he worked with were scanned film mammograms, he used segmentation according to the global threshold and straight line approximation for the pectoral muscle removal. Distance mapping and histogram calculation is being applied for different distances according to distance from the breast skin-air interface. Extracted set of eight features include standard deviation and skewness calculated from histograms and comparison of the tissue density with the pectoral muscle density. Finally, he used k-nearest-neighbor (k-NN) classifier for classification with reported accuracy of around 65 %.

Muhimmah and Zwiggelaar proposed a multiresolution histogram based method which uses no segmentation at all [11]. This method uses comparison of histogram properties on different scales of each image. The proposed method has been tested on the entire MIAS database. The overall correct classification into three categories of 77.57 % was reported.

Oliver et al. proposed in [12] a different approach. First they divided the image into two clusters using fuzzy C-means algorithm. For initializing seeds they used two gray values that represented 15 % and 85 % of the cumulative histogram of the whole breast. From the co-occurrence matrix [6] they have extracted 9 features and classified them using two different classifiers: the k-NN algorithm and a Decision Tree (ID3) classifier. As an experimental dataset they used 300 Medio-Lateral Oblique (MLO) right mammograms taken from the DDSM database. With this method, the classification accuracy of 47 % was achieved using combined classifiers, as opposed to 43.3 % for ID3, and 40.3 % for k-NN individually.

Oliver et al. [13] have used a similar approach to classify 270 mammograms from the MIAS database. The obtained results for the leave-one-out classification method and the k-NN classifier are 67 % and 73 % of correct classification for the ID3 classifier.

Torrent et al. presented a comparison of different approaches for clustering of fatty and dense breast tissue [14].

Oliver et al. [15] have used similar approach (fuzzy C-means algorithm combined with Haralick texture features [6]) in their later work. They have also used two-stage classification, meaning that results from different classifiers were used as an input to the Bayesian classifier. The proposed algorithm was tested on the MIAS and the DDSM databases and the reported overall classification result into four categories was 86 % for the MIAS database and 77 % for the DDSM database.

Petroudi et al. [16] proposed an algorithm based on defining texture classes as statistical distributions (histograms) over texton dictionaries developed from a training set. Classification was done using an appropriate distance measure for the data that is obtained from a training set. Results of correct classification are: 91 %, 64 %, 70 % and 78 % for categories BI-RADS I to BI-RADS IV, respectively.

Subashini et al. presented a method [17] which uses 9 statistical features and the support vector machine classifier obtaining correct classification of 95.44 % on the MIAS database mammograms. Such high classification accuracy can be explained by the fact that they used only 43 out of 322 available images from the MIAS database and the result of correct classification would probably decrease if the entire database was considered.

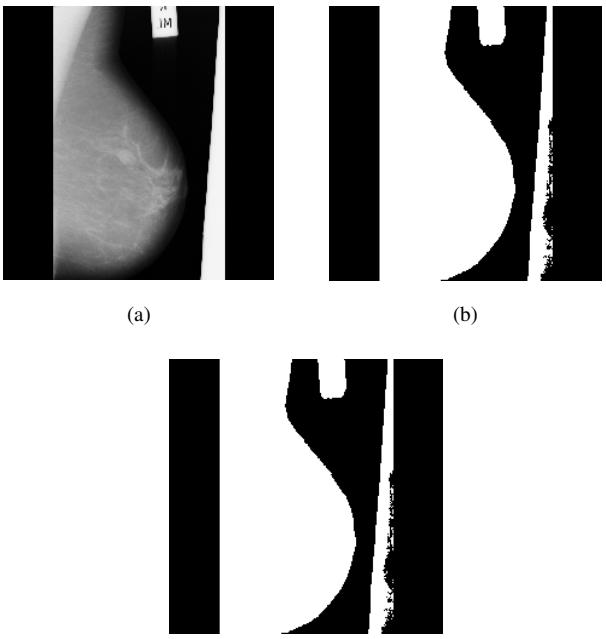
3 METHOD FOR AUTOMATIC BREAST DENSITY CLASSIFICATION

3.1 Preprocessing stage

Before feature extraction it is necessary to complete the preprocessing stage. Because we are using two different datasets it is necessary to make some adjustments in the preprocessing of different images. The mini-MIAS database has 322 images of the same dimension (1024×1024 pixels) with 8 bits per pixel. Because all images are scanned films, it is necessary to properly segment the breast tissue from the background. The segmentation process also removes the artifacts, leaving only the breast tissue area. Segmentation of the breast tissue is done with the mixture of k-means thresholding and morphology operators, in this case, erosion and dilation. Opening, which is accomplished by erosion and dilation, proves to be a good method to eliminate objects around the breast tissue such as orientation tags and adhesive tape. First we divide each image into 10 clusters using k-means algorithm. For creating initial binary mask we use clusters [2-10] in order to remove low intensity objects and light leakage from the scanning process. This step also suppresses high intensity components such are orientation tags which have intensity close or equal to the maximal intensity in an image. After the binary mask with the proposed threshold is created it is necessary to remove everything that lies outside the breast tissue area. Because the breast tissue area is the largest object in the binary image it is possible to use morphological opening operator. The binary mask is first eroded with a square structuring element of 103×103 pixels. Erosion is then being followed by dilatation, with the same structuring element, completing the function of opening. This experimentally determined size of a structuring element which is 1/10 of the image size gives a good result in preserving the breast tissue region and elimination of object outside the breast tissue region. This experimental setup works well with mini-MIAS images or mammograms with the same size and similar properties. The example of these preprocessing steps is shown in Fig 1. From Fig. 1 (b) it is obvious that the breast tissue area has the largest area of all objects in the binary image. Because of that opening with the structuring element significantly smaller than the tissue area will not affect its final shape.

To be able to resize all images in the same way it is important to align them according to the chosen criterion. We have chosen to align images so that the first top pixel is situated in the top left corner of the image window. Since in mammography left and right breast images have different orientation according to the vertical axis, it is necessary to flip images of the right breasts to be the same orientation as the left breast images. Creating binary segmentation masks as explained before is shown in Fig. 1. After flipping and aligning of images from the mini-MIAS database we get the segmentation results as shown in Fig. 2.

The other database which we have used is the KBD-FER digital mammography database. This database consists of 144 digital mammograms. All mammograms have the same dimension (4084×3328 pixels) with 12 bits per pixel. Images are originally stored in DICOM format and do not have to be segmented because that step is done during the acquisition process. The only thing that had to be done was flipping and aligning images to the top left corner. The next step is resizing images to make all the breast areas approximately the same size. The reason for doing this is to be able to extract ROI more easily, concentrating on the same area for each image. That would not be possible if the breast dimensions would vary in size. Since breast density is a quantitative measure, resizing should not affect the classification accuracy. In this step we try to fit all the images in the window which height is 1024

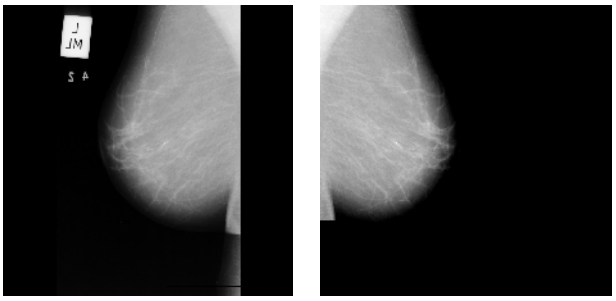


(a) (b)

(c)

Fig. 1: Original image "mdb012" (1a); binary mask after thresholding (1b); binary mask after opening (1c)

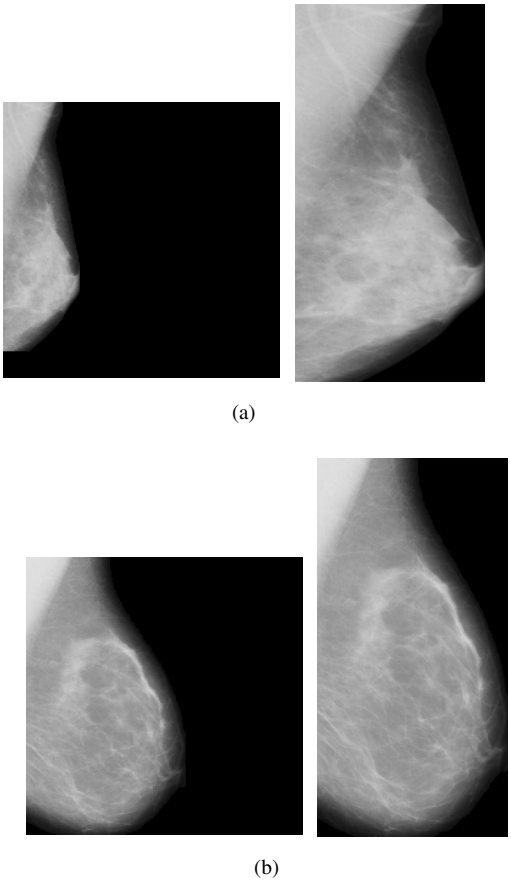
pixels and width 512 pixels. These dimensions are chosen in order to try to minimize the interpolation error, because the size of images from the mini-MIAS database is 1024×1024 pixels. If the window would have a larger width, it would be necessary to interpolate some images with a higher factor and in the case of a smaller window, some images would have to be downsized with a higher factor, resulting in a more significant loss of detail. The interpolation has been done in two dimensions separately. First, the image is being resized in the horizontal dimension to fit a window with height of 512 pixels. Afterwards, vertical interpolation fits the image into a window of 1024



(a) (b)

Fig. 2: Original image "mdb011" (2a); segmented and aligned image (2b)

pixels. For resizing we have used bicubic interpolation method, where the output pixel value is a weighted average of pixels in the nearest 4-by-4 neighborhood, for each resizing step. The result of resizing two initially different breast tissue sizes is shown in Fig. 3. There are two different scenarios shown; Figure 3 (a) shows the result of resizing a breast smaller than the window to fit the window and Fig. 3 (b) shows the result of resizing the breast larger than the window to fit the window. After the interpolation all images have a size of 1024×512 pixels and breast tissue region touches all image borders. This step is the same for both image databases we have used and it results in images of the same final size. Isolating the ap-



(a)

(b)

Fig. 3: The result of resizing a breast smaller than the target window (a); the result of resizing the breast larger than window to fit the target window (b)

propriate ROI for breast density estimation is what has to be done next. It would be wrong to consider the entire breast area or entire mammogram as a ROI because breast density is evaluated only in area around the fibroglandular disc. Isolating the appropriate ROI removes a large number of background pixels and a pectoral muscle in MLO images which we have used for testing performance of the

proposed method. The resizing process explained above now comes handy because there is no need for the pectoral muscle removal and we are able to observe a similar part of fibroglandular disc from all breasts. Figure 4 shows the dimension in pixels of the extracted ROI.

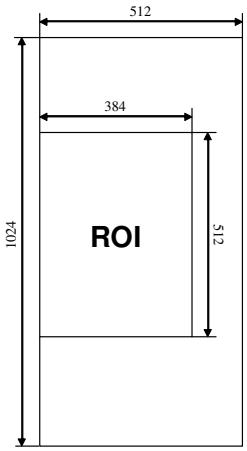


Fig. 4: Dimensions in pixels of the image and extracted ROI

Example of the extracted ROI from an image is shown in Fig. 5. It can be seen that only a very small portion of pectoral muscle and the tissue close to the skin-air interface lies inside ROI. The entire preprocessing step in a block chart form for two different databases is shown in Fig. 6. It can be seen that the preprocessing of digital mammograms is easier because there is no need for segmentation. Not having to scan and segment images gives the possibility of a more frequent usage for CAD purposes, because it significantly shortens the amount of work radiologists have to put into.

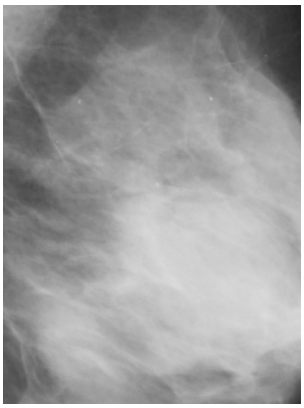


Fig. 5: Example of the extracted ROI from "mdb001"

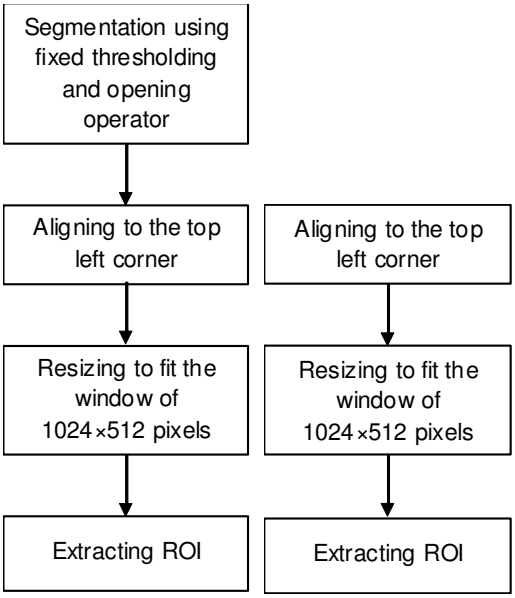


Fig. 6: The entire preprocessing step for the mini-MIAS database (a); the entire preprocessing step for the KBD-FER database (b)

3.2 Feature Extraction

Feature extraction has been done on the GLCM and directly on the ROI of each image. GLCM is a matrix assembled in a way to show the distribution of co-occurring values at a given offset. It is a square matrix which size is defined by the bit depth of the image as 2 to the power of number of bits in an image. From GLCMs we have extracted the Haralick and Soh features [6], [7]. Let $p(i,j)$ be the (i,j)th entry in a normalized GLCM. The mean values for the rows and columns of the matrix are:

$$\begin{aligned} \mu_x &= \textstyle\sum_i \sum_j i \cdot p\,(i,j)\,, \\ \mu_y &= \textstyle\sum_i \sum_j j \cdot p\,(i,j)\,. \end{aligned} \tag{1}$$

The standard deviations for the rows and columns of the matrix are:

$$\begin{aligned} \sigma_x &= \textstyle\sum_i \sum_j (i-\mu_x)^2 \cdot p\,(i,j)\,, \\ \sigma_y &= \textstyle\sum_i \sum_j (j-\mu_y)^2 \cdot p\,(i,j)\,. \end{aligned} \tag{2}$$

$p_x(i)$ is the ith entry in the marginal-probability matrix obtained by summing the rows of $p(i,j)$, $p_y(i)$ is the ith entry in the marginal-probability matrix obtained by summing the columns of $p(i,j)$, N_g is the number of distinct gray levels in GLCM, *HX* and *HY* are entropies of p_x and p_y. The feature extraction has been done for the angles of $0°$, $45°$, $90°$ and $135°$ and distances between the pixel of interest and its neighbor equal to 1, 3, 5 and 7. This gives

a feature vector of 288 elements. Features extracted from the GLCMs are shown in appendix.

From the ROI of each image we have extracted the following set of features: normalized number of pixels with higher intensity than the mean intensity of the muscle region, mean histogram value, normalized number of pixels according to Otsu's thresholding method [18], standard deviation of positive elements, entropy, kurtosis, skewness, 32-bin histogram and 16-bin histogram as well as mean, standard deviation, skewness and kurtosis of 256-bin histogram. Table 1 shows the list of features extracted from ROI. $p_{ix}(i,j)$ is pixel value at the (i,j)th position in the ROI, m and n are the dimension of ROI and $Mpix$ is the mean pixel intensity inside the muscle region. From full 256-bin histograms we have extracted four intensity features: mean, standard deviation, skewness and kurtosis. The feature vector for each mammogram includes the following: Haralick and Soh features extracted from co-occurrence matrices in four orientations with distances of 1, 3, 5 and 7 pixels, multiresolution histograms with 8, 16, 32 and 64 bins, 7 features extracted from pixel intensities in ROI and 4 features extracted from 256-bin histograms. Each feature vector is consisted of 419 elements. Before feature selection and classification stages, all features were normalized using min-max normalization method.

Table 1: Features extracted from ROIs

Feature Name	Mathematical Expression
Higher intensity than muscle region	$N_M = \frac{\sum_i \sum_j p_{ix}(i,j) > \overline{Mpix}}{m \cdot n}$
Mean intensity	$\overline{H} = \frac{\sum_i \sum_j p_{ix}(i,j)}{m \cdot n}$
Higher intensity than Otsu's threshold	$N_O = \frac{\sum_i \sum_j p_{ix}(i,j) > Otsu\_threshold}{m \cdot n}$
Standard deviation	$S = \sqrt{\frac{1}{m \cdot n} \sum_{i=1}^{m \cdot n} \left(p_{ix_i} - \overline{p_{ix}}\right)^2}$
Entropy	$E_n = -\sum_i \sum_j p_{ix}(i,j) \log\left(p_{ix}(i,j)\right)$
Skewness	$S_k = \frac{(p_{ix}(i,j) - \overline{p_{ix}})^3}{S^3}$
Kurtosis	$K = \frac{(p_{ix}(i,j) - \overline{p_{ix}})^4}{S^4}$

3.3 Feature Selection and Classification

After the features have been extracted, we have proceeded with the feature selection step. Our goal has been achieving the highest possible classification accuracy. In the feature selection process all the images from the mini-MIAS and KBD-FER databases have been used with previously described preprocessing procedure. For the feature evaluation we have used wrappers [19] with optimization for different classifiers. Wrappers are feature selection algorithms that search through the space of all possible features and evaluate each subset using the chosen classifier [20]. Subsets that give the best results are being further modified until there is no further improvement for the desired criterion. Wrapper search algorithm implemented in WEKA [8] gives the possibility of choosing different search methods. We have chosen Best First with forward, backward and bi-directional search, Genetic search and Random search with 25 % of all subsets [21]. Each of these methods has been tested in conjunction with k-NN classifier with k equal to one (IB1 [22]) and five (IBk [22]) and Naive Bayesian classifiers [23]. For the evaluation we have used 10-fold cross validation inside the wrapper evaluation algorithm. Originally, the mini-MIAS database mammograms are divided into three density categories: fatty (F), fatty-glandular (G) and dense-glandular (D). There are 106 images belonging to fatty category, 104 images are marked as fatty-glandular and 112 are marked as dense-glandular. Because of the non-standard original division of the MIAS database it is impossible to test the classification accuracy according to the standard BI-RADS density classification. To be able to accomplish that task we have used the trained radiologist classification of the MIAS database. According to the ACR BI-RADS density classification standard the MIAS database is divided into four density categories as follows: BI-RADS I category contains 87 images, BI-RADS II contains 103 images, BI-RADS III contains 95 images and BI-RADS IV contains 37 images. The expert classification was kindly provided by Mr. Zwiggelaar and is the same as the one used in work presented by Oliver et al. [15]. The division into two density categories has been done by merging BI-RADS categories I and II and categories III and IV. This results in division into fatty (F) and dense (D) categories. There are 190 images belonging to fatty category and 132 belonging to dense category. Table 2 shows the division of the mini-MIAS database into different categories according to the tissue density.

Table 2: Division of the mini-MIAS database into different density categories

Density Category	Number of images	Density Category	Number of images
F	106	**BI-RADS I (F)**	87
G	104	**BI-RADS II (F)**	103
D	112	**BI-RADS III (D)**	95
		BI-RADS IV (D)	37

The KBD-FER digital mammography database is obtained from the University Hospital Dubrava, Zagreb,

Table 3: Division of the KBD-FER database into BI-RADS density categories

Density Category	Number of images
BI-RADS I (F)	68
BI-RADS II (F)	36
BI-RADS III (D)	32
BI-RADS IV (D)	8

Table 4: Division of the KBD-FER database into BI-RADS density categories

Number of categories	Classifier	Classification Accuracy
2	IB1 (1-NN)	247/322 (76.7 %)
	IBk (5-NN)	262/322 (81.4 %)
	Naive Bayesian	251/322 (78.0 %)
3	IB1 (1-NN)	227/322 (70.5 %)
	IBk (5-NN)	234/322 (72.7 %)
	Naive Bayesian	219/322 (68.0 %)
4	IB1 (1-NN)	182/322 (56.5 %)
	IBk (5-NN)	197/322 (61.2 %)
	Naive Bayesian	174/322 (54.0 %)

Croatia. This KBD-FER database consists of 144 MLO images, from 72 patients. For the breast density representation, the BI-RADS notation is being used and radiologist's evaluation is considered as ground truth. The KBD-FER database is divided into four categories according to ACR BI-RADS. There are 68 images belonging to BI-RADS I, 36 belonging to BI-RADS II, 32 belonging to BI-RADS III and 8 belonging to BI-RADS IV category. Table 3 shows the division of the KBD-FER database.

From the Table 3 it is obvious that the KBD-FER database does not contain uniform distribution of all four density categories, and the original division of MIAS database in three categories presents a better choice for feature extraction and classifier training according to the uniformity of distribution among categories.

4 EXPERIMENTAL RESULTS AND DISCUSSION

To show the difference between classification accuracy with and without feature selection, we have shown the comparison between them on the same dataset. Without any feature selection, classification for the mini-MIAS database gave the results according to different number of categories and used classifiers shown in Table 4. The classification has been made using the leave-one-out method. From the results shown in Table 4 we can see that the number of correctly classified images reduces as the number of categories increases. Different feature selection methods resulted in significantly different number of features used for classification. The number of features was reduced from roughly 1/3 of the initial feature set to less than 1/6 of the initial feature set. This reduction in number of features is not based on overall feature ranking but on combination of features which provides best classification results. The reason for such a large feature reduction is the high correlation between them. Standard deviation calculated on pixels from ROI gives the best class separability according to χ^2 feature evaluation. When comparing features extracted from co-occurrence matrices, the best separability is achieved using homogeneity (9) and Information Measure of Correlation 1 & 2 (17, 18). When considering different distances between observed pixels in co-occurrence matrices, for the division into 3 categories, distance of 1 pixel gives better results and for the division into 2 and 4 categories, distances of 5 and 7 pixels give better separability. The reason for that may be in the initial division of images into categories, because 4-category division was made according on BI-RADS standard and 2-category division was made by simple grouping of non-dense and dense categories. Classification accuracy using different feature selection methods and different classifiers according to different division of the mini-MIAS database into different density categories is shown in Table 5. The best classification accuracy for the division of the mini-MIAS database into three categories which we have obtained was 264/322 (82.0 %). The corresponding confusion matrix for this classification is shown in Table 6.

Table 6: Confusion matrix for the classification of the mini-MIAS database divided into three categories

		Automatic Classification		
		F	***G***	***D***
Expert's readings	***F***	93	9	4
	G	8	78	18
	D	1	18	93

Classification accuracy for the division of the mini-MIAS database into four categories using the IB1 classifier and Best First backward selection was 255/322 (79.2 %). The best classification accuracy for the division of the mini-MIAS database into two categories which we obtained was 295/322 (91.6 %). It has been obtained using Naive Bayesian classifier and Best First Backward selection feature selection method.

Without any feature selection, classification for the KBD-FER digital mammography database, gave the results according to different number of categories and used classifiers as shown in Table 7. The classification has been done using the leave-one-out method, same as for the mini-MIAS database. The results of the KBD-FER database classification are somewhat less consistent than those of the mini-MIAS database. The reason for that lies in significantly different number of images in each category. The

Table 5: Classification accuracy for the mini-MIAS database divided into different density categories

Classifier	Feature selection method, (Wrapper+...)	Classification Accuracy, 2 categories	Classification Accuracy, 3 categories	Classification Accuracy, 4 categories
IB1 (1-NN)	Best First, Forward selection	285/322 (88.6 %)	**264/322 (82.0 %)**	241/322 (74.8 %)
	Best First, Backward selection	291/322 (90.4 %)	253/322 (78.6 %)	**255/322 (79.2 %)**
	Best First, Bi-directional selection	290/322 (90.1 %)	253/322 (78.6 %)	245/322 (76.1 %)
	Genetic search	287/322 (89.1 %)	250/322 (77.6 %)	243/322 (75.5 %)
	Random search (25 %)	290/322 (90.1 %)	252/322 (78.3 %)	244/322 (75.8 %)
IBk (5-NN)	Best First, Forward selection	282/322 (87.6 %)	254/322 (78.9 %)	234/322 (72.7 %)
	Best First, Backward selection	287/322 (89.1 %)	255/322 (79.2 %)	238/322 (73.9 %)
	Best First, Bi-directional selection	286/322 (88.9 %)	230/322 (71.4 %)	231/322 (71.7 %)
	Genetic search	281/322 (87.3 %)	245/322 (76.1 %)	240/322 (74.5 %)
	Random search (25 %)	288/322 (89.4 %)	254/322 (78.9 %)	241/322 (74.8 %)
Naive Bayesian	Best First, Forward selection	288/322 (89.4 %)	244/322 (75.6 %)	237/322 (73.6 %)
	Best First, Backward selection	**295/322 (91.6 %)**	246/322 (76.4 %)	239/322 (74.2 %)
	Best First, Bi-directional selection	287/322 (89.1 %)	244/322 (75.6 %)	236/322 (73.3 %)
	Genetic search	282/322 (87.6 %)	231/322 (71.7 %)	240/322 (74.5 %)
	Random search (25 %)	289/322 (89.8 %)	246/322 (76.4 %)	243/322 (75.5 %)

Table 7: Classification results of the KBD-FER database without feature selection

Number of categories	Classifier	Classification Accuracy
2	IB1 (1-NN)	131/144 (91.0 %)
	IBk (5-NN)	130/144 (90.3 %)
	Naive Bayesian	129/144 (89.6 %)
4	IB1 (1-NN)	96/144 (66.7 %)
	IBk (5-NN)	87/144 (60.4 %)
	Naive Bayesian	77/144 (53.5 %)

Table 8: The best classification results for the KBD-FER database

Number of categories	Classification accuracy
2	140/144 (97.2 %)
4	110/144 (76.4 %)

best classification accuracy of 140/144 (97.2 %) has been obtained using Best First Forward selection feature selection method and IB1 classifier, for the classification into two categories. Because of significantly more images belonging to non-dense category, the classification into non-dense category was 100 % while the classification for the dense category was 90 %. The best classification accuracy for the classification into four categories is 110/144 (76.4 %). These results are shown in Table 8. This classification result could be improved with the equal division of categories in the database, but we did not have equally divided digital mammography database according to density at the time of testing. Good automatic classification results into two categories could be very useful for later automatic detection of microcalcifications and masses in mammograms. That knowledge could also be used for a better elimination of false positive and false negative results by setting the appropriate detection thresholds.

The results shown in Table 9 give the overall comparison of different recently proposed methods which use substantial number of images from the publicly available MIAS database. The method proposed in this paper out-

performs all the proposed methods despite the number of categories in which database mammograms are divided. Only Oliver et al. achieved better classification results using the combination of results obtained using k-NN and C4.5 classifier estimated with Bayesian classifier. Results of the proposed method show some improvement over the recently proposed methods which use similar classification techniques.

Table 9: Comparison of the proposed classification methods for the MIAS database divided into different categories

	Automatic classification		
	MIAS 3-category	***Method characteristics***	
Oliver et al. [13]	67 %	k-NN	k-means + GLCMs (270 images)
	73 %	ID3	
Muh. & Zwigg. [11]	75.57 %	DAG-SVM	Multi resolution histogram (322 images)
Our method	82.5 %	k-NN (k=1)	Multi features (322 images)
	4-category	***Method characteristics***	
Oliver et al. [15]	**MIAS** 77 %	SFS+k-NN	k-means + GLCMs + morphological features (322 images)
	72 %	C4.5	
	MIAS 86 %	Combination using Bayesian estimation method	
Karssemeijer [10]	**Nijmegen** 65 %	k-NN	Histogram statistics with different distances according to skin-air interface (615 images)
Petroudi et al. [16]	**Non-public database** 75.5 %	NN using χ^2 distribution	Texton dictionary (120 images)
Our method	**MIAS** 79.3 %	k-NN	Multi features (322 images)
	MIAS 2-category	***Method characteristics***	
Oliver et al. [15]	91 %	Combination using Bayesian estimation method	k-means + GLCMs + morphological features (322 images)
Our method	91.6 %	k-NN	Multi features (322 images)

Our goal was to outperform the accuracy of single classifiers and we did not test how a set of features which we have extracted performs when using classifier combination. Results achieved by Oliver et al. using k-NN and C4.5 classifiers separately, which are part of the most successful method published in recent papers, show slightly lower classification accuracy than the method proposed in this paper.

5 CONCLUSION

The proposed method consists of an image preprocessing stage, feature extraction and finally classification. Although we have used two completely different image databases, with different properties, our idea was to have similar ROIs extracted from each image. The feature extraction process is based on extraction of Haralick and Soh textural features form GLCMs, statistical features from the ROI and histogram features from the ROI. We have extracted 419 features for each image and used different feature selection methods with the adaptation to different classifiers. Results of the overall classification accuracy, after the feature selection, were improved around 3 % to 12 %. We have used the following approaches integrated into wrapper feature evaluators for the feature selection: forward selection, backward selection, bi-directional selection, genetic search and random search with the random search over 25 % of all possible feature sets. The best classification accuracy was generally achieved when using forward and backward feature selection. Classification accuracy has been tested using different classifiers and the best results were not always correlated with the usage of the same classifier. For the mini-MIAS database divided into three density categories the best classification result of 82.5 % was achieved using the k-NN classifier with k equal to 1. For the division of the mini-MIAS base into four categories we got 79.2 % correctly classified instances. The proposed classification method gave the correct classification of 97.2 % for the division into two categories and 76.3 % for the division into four categories according to BI-RADS standard on the KBD-FER digital mammography database. Classification results were overall slightly better for KBD-FER database than for the mini-MIAS database probably because of the raw image quality. It is important to stress out that we have not used any image selection and the testing results have been obtained from the test sets which size is sufficient to avoid possible overfitting of the used model. The obtained results show a slight improvement over the proposed methods on the dataset which is publicly available. Our further work will be based on accuracy improvement by extracting different features and using different feature selection and classification algorithms. We will also focus on extracting new features that better correspond with different breast density categories and therefore give better classification results.

FEATURES EXTRACTED FROM THE GLCMS

Feature Name	Mathematical Expression
1. Autocorrelation	$f_1 = \sum_i \sum_j (ij)\, p(i,j)$
2. Contrast	$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \,\vert\vert i-j\vert = n \right\}$
3. Correlation	$f_3 = \frac{\sum_i \sum_j (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
4. Cluster Prominence	$f_4 = \sum_i \sum_j (i + j - \mu_x - \mu_y)^4 \, p(i,j)$
5. Cluster Shade	$f_5 = \sum_i \sum_j (i + j - \mu_x - \mu_y)^3 \, p(i,j)$
6. Dissimilarity	$f_6 = \sum_i \sum_j \vert i-j\vert \, p(i,j)$
7. Energy	$f_7 = \sum_i \sum_j p(i,j)^2$
8. Entropy	$f_8 = -\sum_i \sum_j p(i,j) \log\left(p(i,j)\right)$
9. Homogeneity	$f_9 = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$
10. Maximum Probability	$f_{10} = \max\left(p(i,j)\right)$
11. Variance	$f_{11} = \sum_i \sum_j (i-j)^2 \, p(i,j)$
12. Sum Average	$f_{12} = \sum_{i=2}^{2N_g} i p_{x+y}(i)$
13. Sum Variance	$f_{13} = \sum_{i=2}^{2N_g} (i - f_{14})^2 \, p_{x+y}(i)$
14. Sum Entropy	$f_{14} = -\sum_{i=2}^{2N_g} p_{x+y}(i) \log\left(p_{x+y}(i)\right)$
15. Difference Variance	$f_{15} = -\sum_{i=0}^{N_g-1} (i - f_6)^2 \, p_{x-y}(i)$
16. Difference Entropy	$f_{16} = -\sum_{i=0}^{N_g-1} p_{x-y}(i) \log\left(p_{x-y}(i)\right)$
17. Information Measure of Correlation - 1	$f_{17} = \frac{-\sum_i \sum_j p(i,j)\left(\log(p(i,j)) - \log\left(p_x(i) p_y(i)\right)\right)}{\max(HX, HY)}$
18. Information Measure of Correlation - 2	$f_{18} = \sqrt{1 - e^{-2(a-b)}}$, $a = -\sum_i \sum_j p_x(i)\, p_y(i) \log\left(p_x(i)\, p_y(i)\right)$, $b = -\sum_i \sum_j p(i,j) \log\left(p_x(i)\, p_y(i)\right)$.

ACKNOWLEDGMENT

We would like to thank Mr. Reyer Zwiggelaar and his research team for providing us with the expert classification of the MIAS database into four density categories according to ACR BI-RADS. The authors would like to thank J. Pont and E. Pérez from the Department of Radiology of the Girona University Hospital "Dr. Josep Trueta" (Spain) and E.R.E. Denton from the Department of Breast Imaging of the Norwich and Norfolk University Hospital (UK) for the BIRADS classification of the MIAS database used in this work. We would also like to thank the Radiology Department of the University Hospital Dubrava, Zagreb, Croatia, for providing us with the expert classification of KBD-FER digital database. The work described in this paper was conducted under the research project: "Intelligent Image Features Extraction in Knowledge Discovery Systems" (036-0982560-1643), supported by the Ministry of Science, Education and Sports of the Republic of Croatia.

REFERENCES

[1] S. Obenauer, C. Sohns, C. Werner, E. Grabbe, Impact of breast density on computer-aided detection in full-field digital mammography, Journal of Digital Imaging, Vol. 19, No. 3, pp. 258-263 (2006).

[2] J.N. Wolfe, Risk for Breast Cancer Development Determined by Mammographic Parenchymal Pattern, Cancer, Vol. 37, Issue 5, pp. 2486-2492 (May 1976).

[3] American College of Radiology, American College of Radiology Breast Imaging Reporting and Data System (BI-RADS), (American College of Radiology, 4th Edition, 2003).

[4] J. Suckling, J. Parker, D.R. Dance, S. Astley, I.Hutt, C.R.M. Boggis, I. Ricketts, E. Stamatakis, N. Cernaez, S.L. Kok, P. Taylor, D. Betal, J. Savage, The Mammographic Image Analysis Society Digital Mammogram Database, Proceedings of the 2nd International Workshop on Digital Mammography, York, England, pp. 375-378 (10-12 July 1994).

[5] M. Heath, K. Bowyer, D. Kopans, R. Moore, W.P. Kegelmeyer, The Digital Database for Screening Mammography, Fifth International Workshop on Digital Mammography, IWDM 2000, pp. 212-218 (Medical Physics Publishing, 2001).

[6] R.M. Haralick, K.S. Shanmugan, I. Dunstein, Textural Features for Image Classification, IEEE Trans. Syst., Man, Cybern., Vol. SMC-3, No. 6, pp. 610-621 (November 1973).

[7] L. Soh and C. Tsatsoulis, Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices, IEEE Transactions on Geoscience and Remote Sensing, Vol. 37, No. 2, pp. 780-795 (March 1999).

[8] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA Data Mining Software: An Update, SIGKDD Explorations, Vol. 11, Issue 1, pp. 10-18 (July 2009).

[9] I. Guyon and A. Elissee, An Introduction to Variable and Feature Selection, Journal of Machine Learning Research, Vol. 3, pp. 1157-1182 (March 2003).

[10] N. Karssemeijer, Automated classification of parenchymal patterns in mammograms, Phys. Med. Biol., Vol. 43, pp. 365-389 (1998).

[11] I. Muhimmah, R. Zwiggelaar, Mammographic Density Classification using Multiresolution Histogram Information, Proceedings of the International Special Topic Conference on Information Technology in Biomedicine, ITAB 2006, Ioannina - Epirus, Greece, 6 pages (26-28 October 2006).

[12] A. Oliver, J. Freixenet, R. Zwiggelaar, Automatic Classification of Breast Density, Proceedings of the IEEE International Conference on Image Processing, ICIP 2005, Vol. 2, pp. 1258-1261 (11-14 September 2005).

[13] A. Oliver, J. Freixenet, A. Bosch, D. Raba, and R. Zwiggelaar, Automatic Classification of Breast Tissue, Lect. Notes Comput. Sci., Vol. 3523, pp. 431-438 (2005).

[14] A. Torrent, A. Bardera, A. Oliver, J. Freixenet, I. Boada, M. Feixes, R. Martí, X. Lladó, J. Pont, E. Pérez, S. Pedraza and J. Martí, Breast Density Segmentation: A Comparison of Clustering and Region Based Techniques, IWDM 2008, LNCS 5116, pp. 9-16 (2008).

[15] A. Oliver, J. Freixenet, R. Martí, J. Pont, E. Pérez, E.R.E. Denton, R. Zwiggelaar, A Novel Breast Tissue Density Classification Methodology, IEEE Transactions on Information Technology in Biomedicine, Vol. 12, Issue 1, pp. 55-65 (January 2008).

[16] S. Petroudi, T. Kadir, M. Brady, Automatic Classification of Mammographic Parenchymal Patterns: a Statistical Approach, Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Cancun, Mexico, Vol. 1, pp. 798-801 (17-21 September 2003).

[17] T.S. Subashini, V. Ramalingam, S. Palanivel, Automated Assessment of Breast Tissue Density in Digital Mammograms, Computer Vision and Image Understanding, Vol. 114, Issue 1, pp. 33-43 (January 2010).

[18] N. Otsu, A Threshold Selection Method from Gray-Level Histograms, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, pp. 62-66 (1979).

[19] R. Kohavi, G.H. John, Wrappers for Feature Subset Selection, Artificial Intelligence, Vol. 97, Issue 1-2, pp. 273-324 (December 1997).

[20] M. Gütlein, E. Frank, M. Hall, A. Karwath, Large-scale Attribute Selection Using Wrappers, Proceedings IEEE Symposium on Computational Intelligence and Data Mining, pp. 332-339 (March 30-April 2 2009).

[21] M. Mustra, M. Grgic, K. Delac, Feature Selection for Automatic Breast Density Classification, Proceedings ELMAR 2010, pp. 9-16 (15-17 September 2010).

[22] D.W. Aha, D. Kibler, M.K. Albert, Instance-based Learning Algorithms, Machine Learning, Vol. 6, pp. 37-66 (January 1991).

[23] G.H. John, P. Langley, Estimating Continuous Distributions in Bayesian Classifiers, 11th Conference on Uncertainty in Artificial Intelligence, San Mateo, pp. 338-345 (18-20 August 1995).

Mario Muštra received the B.Sc. degree in electrical engineering from the University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia in 2007. Since 2007 he is with the University of Zagreb, FER, Department of Wireless Communications and Video Communication Laboratory where he is a Research Assistant. His current research interests are image processing techniques in medical imaging and development of automatic algorithms for detection of tissue types in mammograms.

Mislav Grgić received his BSc, MSc, and PhD in electrical engineering from University of Zagreb, Faculty of Electrical Engineering and Computing (FER), Zagreb, Croatia, in 1997, 1998, and 2000, respectively. He is currently a full professor in the Department of Wireless Communications at FER. His research interests include multimedia communications and image processing. He has had more than 120 scientific papers published in international journals and conference proceedings.

Krešimir Delač received the PhD. degree in electrical engineering from the University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia in 2007. He is with the Video Communication Laboratory where he is a Collaborator Researcher. His current research interests are image processing techniques in face recognition and medical imaging, pattern classification and recognition and data mining.

AUTHORS' ADDRESSES

Mario Muštra, B.Sc.
Prof. Mislav Grgić, Ph.D.
Krešimir Delač, Ph.D.
Faculty of Electrical Engineering and Computing,
University of Zagreb,
Unska 3, HR-10000 Zagreb, Croatia
email: {mario.mustra, mislav.grgic}@fer.hr
kdelac@ieee.org

Received: 2012-05-14
Accepted: 2012-09-08