

Feature Selection for Automatic Breast Density Classification

Mario Mustra, Mislav Grgic, Kresimir Delac

Department of Wireless Communications
University of Zagreb, FER
Zagreb, Croatia
mario.mustra@fer.hr

Abstract - Mammography is probably the best method for early detection of abnormalities in the breast tissue. 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 also be useful for setting an appropriate threshold. It is impossible to produce satisfactory classification results by knowledge of overall intensity because exposure and breast volume are different. Because of that we observe breast density as a texture classification problem. In this paper we propose feature selection process based on Haralick and Soh feature set with optimization for k-nearest neighbor classifier. Feature selection was done by individual feature ranking, using linear forward selection and finally using wrappers. The best feature selection results were obtained using wrappers. The improvement on overall classification is 6.8% in comparison to the classification without feature selection on the same dataset.

Keywords - Breast Density; Feature Selection; Haralick and Soh Features; Classification

I. INTRODUCTION

Breast density is an important measure which describes the possibility for the detection of abnormalities in mammograms. Higher breast density usually indicates a higher possibility for the presence of malignant tissue. Because of high breast density, it is more difficult to detect malignant tissue. Breast density is usually correlated with the woman's age in a way that younger women usually have denser breasts than older women. Wolfe was one of the first researchers who presented the correlation between different breast densities and the probability for the development of breast cancer [1]. Today radiologists classify breast according to their density into four categories. Categories according to the American College of Radiology (ACR) BI-RADS are [2]:

- 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 (71%-100%).

Automatic breast density is somewhat more difficult to perform, because human observer can distinguish different

structures very well, despite those structures can have very similar intensity properties. Therefore, for accurate classification it is necessary to extract as many image features as possible and then try to select ones that are least correlated with each other and produce best classification results. When dealing with many features, it is hard to make manual feature selection and for that purpose one needs to have an automatic feature selection method. Automatic breast density classification achieved various results on different datasets. Two publicly available mammography databases are the best ground to make objective comparison between presented methods. Those two databases are the MIAS database [3] and the DDSM database [4]. The MIAS database, although outdated, presents a good choice for automatic breast detection. The MIAS database is publicly available as the mini-MIAS database, containing all the images as original MIAS. The difference is that the mini-MIAS images are downsampled to the resolution of 1024×1024 pixels with 8 bits per pixel. The main disadvantage of the MIAS database, in making it suitable for breast density classification, lies in the lack of ACR BI-RADS standardized classification into four classes. Instead, the MIAS database has mammograms divided into three classes.

There are many papers that consider automatic breast density classification. Muhimmah and Zwigelaar proposed a multiresolution histogram based method which uses no segmentation at all [5]. 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 [6] a classification method that uses a fuzzy C-means algorithm to group pixels into separate categories. For initializing seeds they used two gray values that represented 15% and 85% of the cumulative histogram of the whole breast. Extracted features were classified 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, a classification accuracy of 47% was achieved using combined classifiers, as opposed to 43.3% for ID3, and 40.3% for k-NN. Oliver et al. [7] have used fuzzy C-means algorithm combined with Haralick texture features [8] in their later work. They have also used two-stage

classification, meaning that results from different classifiers were used as an input of 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. Subashini et al. presented a method [9] which uses Haralick features and the support vector machine classifier obtaining correct classification of 95.44% on the MIAS database mammograms. Such high classification accuracy can be biased by selecting only 43 of 322 mammograms from the MIAS database.

In this paper we try to test classification accuracy in automatic breast density classification using solely Haralick and Soh [8, 10] texture features and k-NN classifier with different feature selection schemes. The idea is to obtain the highest possible correct classification rate and to find out which texture features should be further investigated. The Haralick and Soh features are extracted from the gray-level co-occurrence matrix (GLCM) [8] of each image. The feature evaluation, selection and classification were done using Weka data mining software [11]. The idea behind the feature selection process is in our case to get the best possible classification results from the extracted features. That is not possible solely by selecting the features which provide best individual classification results but with the optimal combination of features, as explained in [12]. Good feature subsets should contain features that are highly correlated with the class but uncorrelated with each other.

This paper is organized as follows. In Section II the entire feature extraction procedure with the preprocessing stage and the feature extraction from the GLCMs is explained. Section III presents the feature selection methods and brings the experimental results for different methods. Section IV draws the conclusion.

II. FEATURE EXTRACTION

Before we could extract all the features, it was necessary to complete the entire preprocessing stage. The entire experiment is based on the mini-MIAS database which has its own specificities. All images in the mini-MIAS database have the same dimension of 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 is done with the mixture of fixed gray-level thresholding and morphology operators, in this case, erosion and dilation. Erosion and dilation prove to be a good method to eliminate objects around the breast tissue such as orientation tags and adhesive tape. The segmentation step begins with creation of the binary mask according to the threshold. In this step all gray values lower than 15 and higher than 245 are removed. The binary mask is then eroded with a square structuring element of 103×103 pixels. Erosion is being followed by dilatation, with the same structuring element, completing the function of opening. The example of these preprocessing steps is shown in Fig. 1.

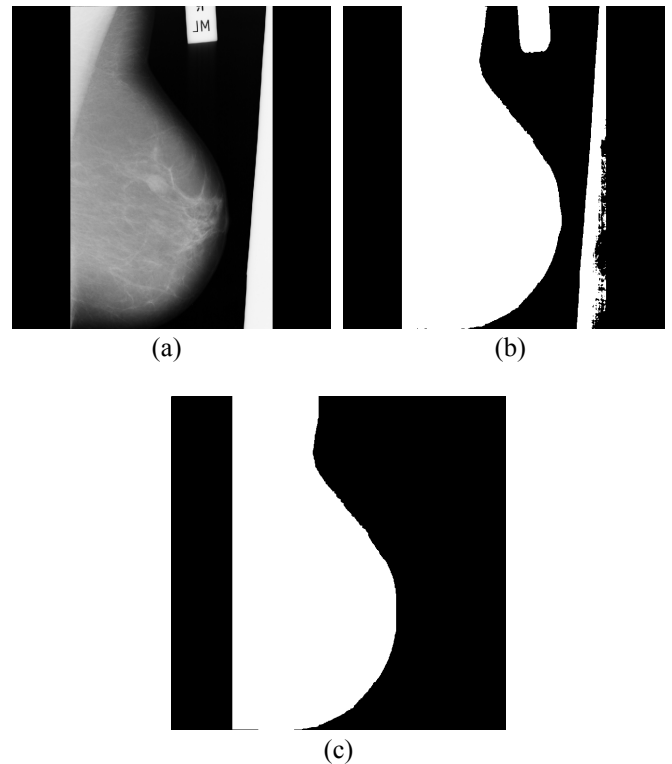


Figure 1. (a) Original image "mdb012". (b) Binary mask after thresholding. (c) Binary mask after opening.

After the segmentation from the background all images showing left breasts in medio-lateral oblique view are automatically flipped to be the same orientation as images showing right breasts. This is followed by aligning all the images to the left border in order to proceed with the registration process. Flipping and aligning of binary masks is shown in Fig. 2.

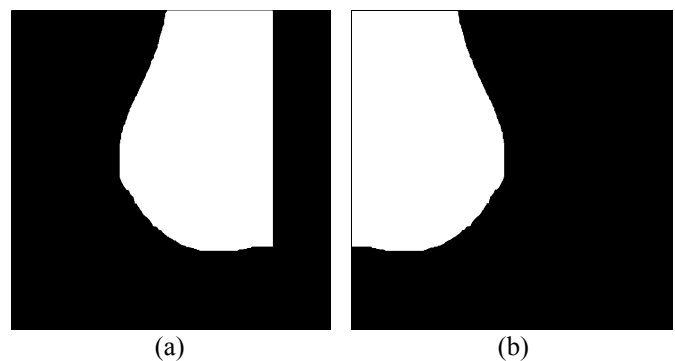


Figure 2. (a) Binary mask after opening "mdb011". (b) Binary mask automatically flipped and aligned to the top left corner.

The result of image segmentation and alignment is shown in Fig. 3.

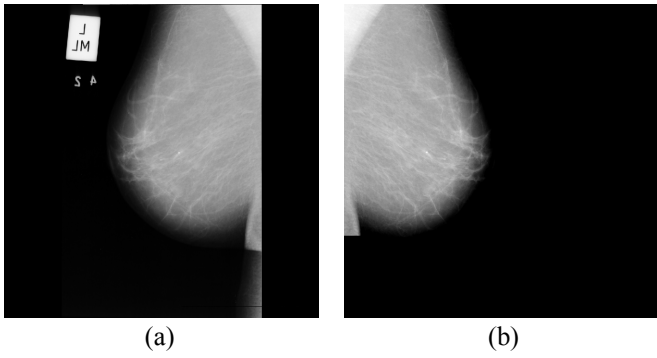


Figure 3. (a) Original image "mdb011". (b) Segmented and aligned image.

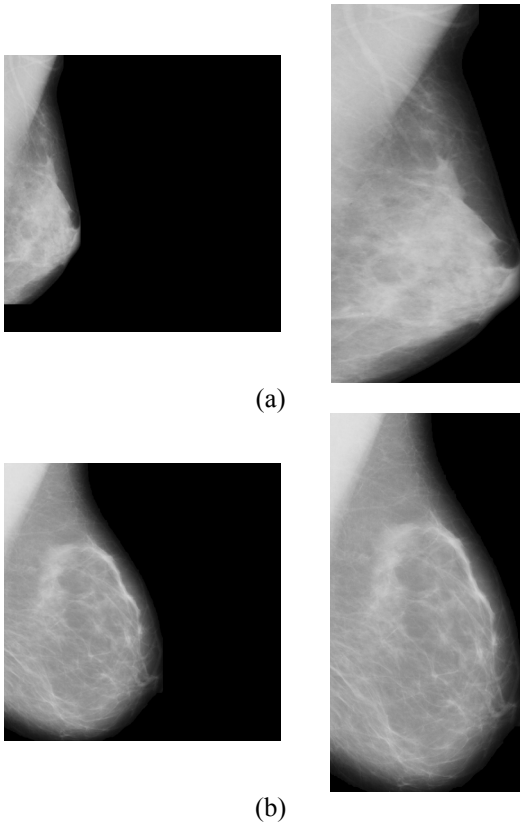


Figure 4. (a) The result of resizing a breast smaller than the target window. (b) The result of resizing the breast larger than window to fit the target window.

The next step is resizing images to make all the breast areas of approximately the same size. In this process we try to fit all the images in the window which height is 1024 pixels and width 512 pixels. These dimensions are chosen in order to try to minimize the interpolation error. If the window would have larger width, it would be necessary to interpolate some images with the higher factor and with the smaller window, some images would have to be downsampled with higher factor, resulting in more significant loss of detail. The interpolation is done in two dimensions separately. First, the image is resized in the vertical dimensions to fit the window with height of 1024 pixels. Afterwards, horizontal interpolation fits the

image into window of 512 pixels. 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 the resizing step. The result of this step is shown in Fig. 4. There are two different scenarios; Fig. 4 (a) shows the result of resizing a breast smaller than window to fit the window and Fig. 4 (b) shows the result of resizing the breast larger than window to fit the window. After the interpolation all images have size of 1024×512 pixels and breast tissue region touches all image borders.

After the preprocessing steps are finished it is necessary to isolate the region of interest (ROI). It would be wrong to consider the entire breast area as ROI because breast density is evaluated only in the fibroglandular disc. The resizing process now comes handy because there is no need for the pectoral muscle removal and we are able to observe the similar part of fibroglandular disc from all breasts. Resizing of images will not have a large influence on extracted features because breast density is a quantitative measure. Fig. 5 shows the dimension in pixels of the extracted ROI.

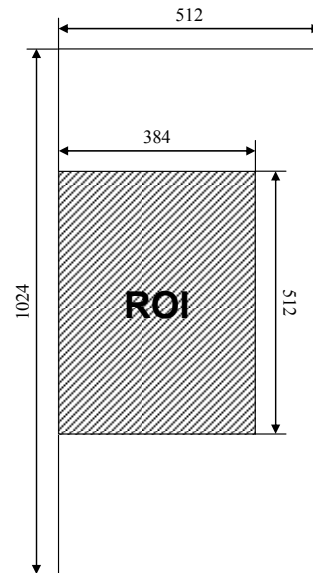


Figure 5. Dimensions in pixels of the image and extracted ROI.

The example of the extracted ROI from the image is shown in Fig. 6.

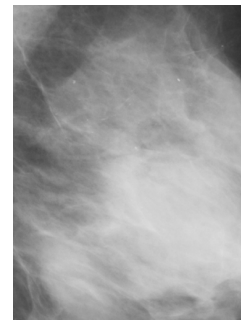


Figure 6. Example of the extracted ROI from "mdb001".

The entire preprocessing step in a block chart form is shown in Fig. 7.

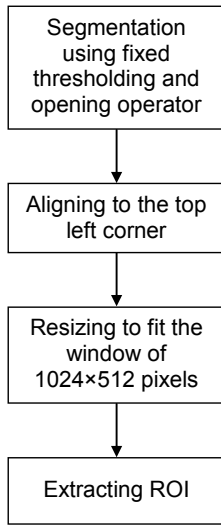


Figure 7. The entire preprocessing step.

The next step is feature extraction. We have chosen to extract the following Haralick and Soh features. 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:

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

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

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

Table I shows the extracted features. $p(i,j)$ is the (i,j) th entry in a normalized GLCM, $p_x(i)$ is the i th entry in the marginal-probability matrix obtained by summing the rows 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 only for the angle of 0° and distance between the pixel of interest and its neighbor equal to 1.

TABLE I. EXTRACTED FEATURES

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

III. FEATURE SELECTION

For feature selection process all images from the mini-MIAS database have been used with previously described preprocessing procedure. We have used three different feature selection methods with the adaptation for IB1 (k-NN) classifier. The first method we have tested is feature ranking. Feature ranking performs feature evaluation on isolated features. It is important to stress out that the combination of features with the best isolated ranking does not necessarily provide the best classification result or the feature number minimization because features can be highly correlated. We have used two different feature evaluators implemented in Weka, the Chi-Square feature evaluator and Classifier feature evaluator. The Chi-Square feature evaluator evaluates features individually by measuring the chi-squared statistic with respect to the class. The result obtained using the Chi-Square attribute evaluator is shown in Table II. Higher rank means that the feature should solely provide better classification result and attribute number corresponds to features from the Table I.

TABLE II. LIST OF ATTRIBUTES ACCORDING TO THEIR RANK

Rank	Chi-Square statistic	Feature number
1.	185.8129	17
2.	184.8239	18
3.	162.9306	3
4.	161.7668	14
5.	118.4767	11
6.	118.4767	1
7.	116.7354	13
8.	115.0909	7
9.	104.9851	12
10.	88.3584	4
11.	87.4099	8
12.	81.4906	10
13.	64.7906	5
14.	48.9340	9
15.	47.0366	6
16.	37.6284	16
17.	34.6187	15
18.	34.6187	2

Classification results obtained from k-NN classifier and leave-one-out method using only one feature are shown in Fig. 8.

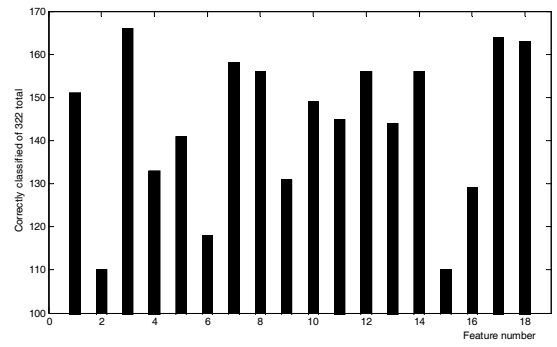


Figure 8. Results of k-NN classification using single features shown in Table II.

The results of correct classification after sorting according to the Chi-square statistic are shown in Fig. 9.

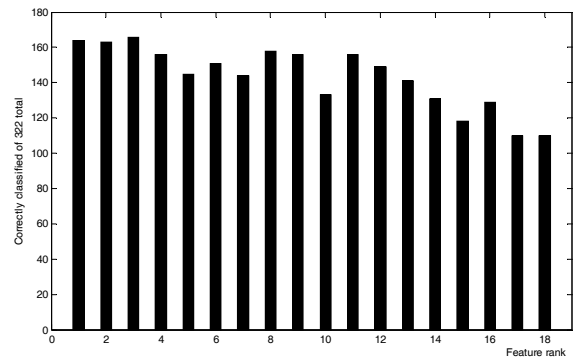


Figure 9. Results of k-NN classification with respect to the feature rank using Chi-square feature evaluator method.

After sorting the results shown in Fig. 8 according to the highest classification accuracy we get the results shown in Table III. With the Classifier feature evaluator, the results of feature selection are slightly different. In Table IV the results of the Classifier feature evaluator for k-NN classifier are shown. The features are sorted according to the rank although this sorting does not necessarily provide the best results. It is obvious that this Classifier feature evaluator ranking algorithm gives better results for the k-NN classifier than Chi-square method. These results are shown in Fig. 10 from which is obvious that columns representing the number of correctly classified instances are sorted almost monotonously decreasing.

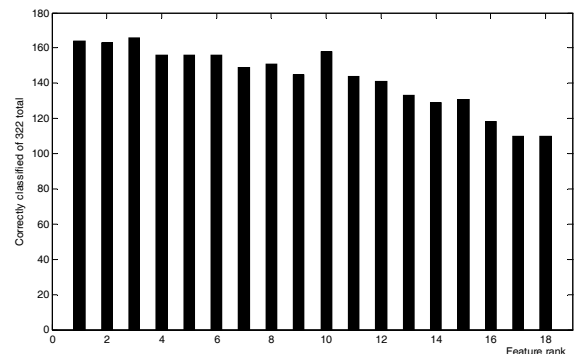


Figure 10. Results of k-NN classification with respect to the feature rank using Classifier feature evaluator method.

TABLE III. FEATURES WITH BEST INDIVIDUAL CLASSIFICATION RESULTS

Rank	Correct classification using k-NN	Feature number
1.	166	3
2.	164	17
3.	163	18
4.	158	7
5.	156	14
6.	156	12
7.	156	8
8.	151	1
9.	149	10
10.	145	11
11.	144	13
12.	141	5
13.	133	4
14.	131	9
15.	129	16
16.	118	6
17.	110	15
18.	110	2

TABLE IV. FEATURES ACCORDING TO THE CLASSIFIER FEATURE EVALUATOR

Rank	Correct classification using k-NN	Feature number
1.	164	17
2.	163	18
3.	166	3
4.	156	12
5.	156	8
6.	156	14
7.	149	10
8.	151	1
9.	145	11
10.	158	7
11.	144	13
12.	141	5
13.	133	4
14.	129	16
15.	131	9
16.	118	6
17.	110	15
18.	110	2

The next method we have tested is Linear Forward Selection (LFS) [13]. We have also used IB1 classifier with the Correlation-based feature subset evaluator algorithm [14] implemented in Weka. LFS should reduce the number of used features with increase in classification accuracy. Preselection of subsets can be done with and without previous feature ranking. There are two techniques of LFS: the fixed-set technique and the fixed-width technique. The fixed-set technique uses the pool of first k highest ranked features and creates subsets so the highest classification accuracy is achieved. The fixed-set technique uses also the first k highest ranked features but in each step when one feature is used in forming a new subset, $k+1st$ highest ranked feature is added to the pool. These two LFS techniques are shown in Fig. 11.

Classification with IB1 classifier using all 18 features described in Table I gave the overall accuracy of 214/322 which equals 66.5%. Using the LFS method number of features is reduced from 18 to 11 and correct classification rate is improved to 222/322 which equals 68.9%. To reduce number of used features it is possible to use backward selection method. Backward selection method starts with a subset which includes all features and reduces them to obtain maximal classification accuracy. This approach gave 223/322 or 69.3% correctly classified images. The last method we have tested is feature selection using wrappers [13].

Wrappers use a search algorithm, like LFS, to search through the features space and evaluate each subset by running a model on the subset. We have used 10-fold cross validation in our experiment for estimating the accuracy. The problem with wrappers is that they can be computationally expensive and have a risk of overfitting to the model. The overfitting problem can be avoided by having the different training and test sets. We have tested the wrapper feature selection scheme with different search methods and test results are shown in Table V.

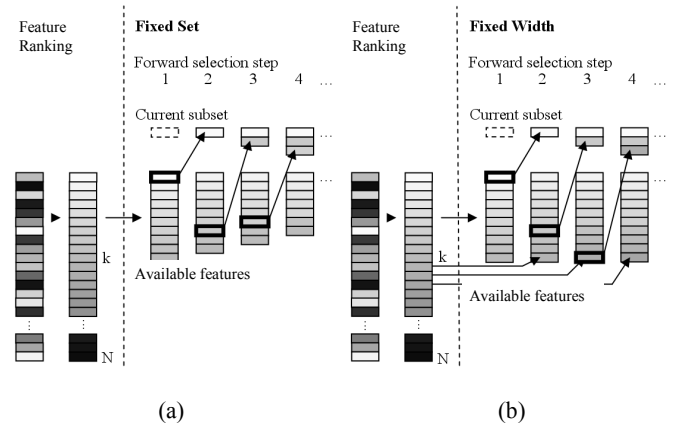


Figure 11. LFS techniques: (a) The fixed-set technique, (b) The fixed-width technique.

TABLE V. WRAPPER FEATURE SELECTION RESULTS WITH DIFFERENT SEARCH METHODS

Search method	Number of selected features	Correctly classified instances of 322	Correct classification (%)
Best first, forward search	13	229	71.1
Best first, backward search	8	232	72.0
Best first, bi-directional search	5	232	72.0
Linear forward selection	11	222	68.9
Greedy stepwise	6	229	71.1
Genetic search	9	236	73.3
Random search 25% of all subsets	8	232	72.0
Exhaustive search through all subsets	8	233	72.4

The graphical comparison of feature selection methods is shown in Fig. 12.

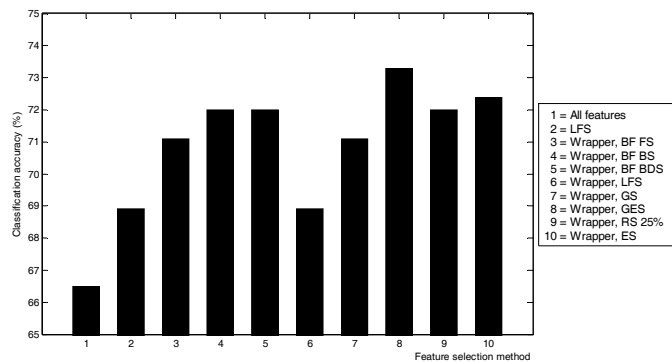


Figure 12. The graphical comparison of feature selection methods (BF=Best first, FS=Forward search, BD=Backward search, BDS=Bi-directional search, GS=Greedy stepwise, GES=Genetic search, RS=Random search, ES=Exhaustive search).

IV. CONCLUSION

In this paper we have investigated different feature selection algorithms. The idea was to get as good as possible classification results from the optimal feature set. The classification task was to classify mammograms from the mini-MIAS database into three categories, according to the description by the authors of the database. We have extracted

18 Haralick and Soh texture features from the ROI selected in a way that only breast tissue pixels are inside of it. This provides better results for some features, because the dynamic range of our texture is smaller. Feature selection has been done using three different approaches: individual ranking, linear forward selection and wrappers. The quality of feature selection has been proved by classification using IB1 classifier and leave-one-out method. The results show a significant improvement in classification accuracy even on a very small feature set like the one we used. The best classification result was obtained using wrapper method and genetic search method for the feature selection. The obtained improvement in classification accuracy is in 22 more (of 322) correctly classified images, or 6.8% more than without using feature selection. Our further work will be based on extracting more features that provide better classification results and on finding feature selection methods that provide better results while trying to avoid overfitting.

ACKNOWLEDGMENT

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] J.N. Wolfe, "Risk for Breast Cancer Development Determined by Mammographic Parenchymal Pattern", *Cancer*, vol. 37, issue 5, May 1976, pp. 2486-2492.
- [2] American College of Radiology, "Illustrated Breast Imaging Reporting and Data System (BI-RADS)", American College of Radiology, Third Edition, 1998.
- [3] J. Suckling, J. Parker, D.R. Dance, S. Astley, I.Hutt, C.R.M. Boggis, I. Ricketts, E. Stamatakis, N. Cernaes, 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, 10-12 July 1994, pp. 375-378.
- [4] 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, Medical Physics Publishing, 2001, pp. 212-218.
- [5] I. Muhimmah, R. Zwigelaar, "Mammographic Density Classification using Multiresolution Histogram Information", *Proceedings of the International Special Topic Conference on Information Technology in Biomedicine*, itab 2006, Ioannina - Epirus, Greece, 26-28 October 2006, 6 pages.
- [6] A. Oliver, J. Freixenet, R. Zwigelaar, "Automatic Classification of Breast Density", *Proceedings of the IEEE International Conference on Image Processing*, ICIP 2005, vol. 2, 11-14 September 2005, pp. 1258-1261.

- [7] A. Oliver, J. Freixenet, R. Martí, J. Pont, E. Pérez, E.R.E. Denton, R. Zwigelaar, "A Novel Breast Tissue Density Classification Methodology", *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, issue 1, January 2008, pp. 55-65.
- [8] R.M. Haralick, K.S. Shanmugan, I. Dunstein, "Textural features for image classification", *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, November 1973, pp. 610-621.
- [9] 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, January 2010, pp. 33-43.
- [10] 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, March 1999, pp. 780-795
- [11] 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, July 2009, pp. 10-18.
- [12] I. Guyon and A. Elissee, "An introduction to variable and feature selection", *Journal of Machine Learning Research*, vol. 3, March 2003, pp. 1157-1182.
- [13] M. Gütlein, E. Frank, M. Hall, A. Karwath, "Large-scale attribute selection using wrappers", *Proceedings IEEE Symposium on Computational Intelligence and Data Mining*, March 30 - April 2 2009, pp. 332-339.
- [14] M. Hall, "Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning", *Proceedings of the 17th International Conference on Machine Learning (ICML2000)*, Stanford, CA, 2000, pp. 359-366.