Deep learning and non-negative matrix factorization in recognition of mammograms

Bartosz Swiderski, Jaroslaw Kurek, Stanislaw Osowski, Michal Kruk, Walid Barhoumi

SPIE.

Deep Learning and Non-Negative Matrix Factorization in Recognition of Mammograms

Bartosz Swiderski [a], Jaroslaw Kurek [a], Stanislaw Osowski [b], Michal Kruk [a], Walid Barhoumi [c]

[a] Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences, Warsaw, Poland

[b] Institute of Theory of Electrical Engineering, Measurement and Information Systems, Warsaw University of Technology and Military University of Technology, Warsaw, Poland

[c] Research Team on Intelligent Systems in Imaging and Artificial Vision (SIIVA) – RIADI, Laboratory, ISI, Ariana, Tunisia

ABSTRACT

This paper presents novel approach to the recognition of mammograms. The analyzed mammograms represent the normal and breast cancer (benign and malignant) cases. The solution applies the deep learning technique in image recognition. To obtain increased accuracy of classification the nonnegative matrix factorization and statistical self-similarity of images are applied. The images reconstructed by using these two approaches enrich the data base and thanks to this improve of quality measures of mammogram recognition (increase of accuracy, sensitivity and specificity). The results of numerical experiments performed on large DDSM data base containing more than 10000 mammograms have confirmed good accuracy of class recognition, exceeding the best results reported in the actual publications for this data base.

Keywords: Mammograms, non-negative matrix factorization, deep learning, classification.

1. INTRODUCTION

Breast cancer is one of the most important causes of fatality among the cancers for women. Its detection in an early stage allows treatment before it is spread to the other parts of the body. The manual diagnostic recognition of breast cancer on the basis of mammograms is a difficult task and prone to error, since searching for abnormality of the image is very difficult, due to very small changes, which are difficult for interpretation. Therefore, the computer supported decision system is needed [1]-[3]. It will provide a fast, repeatable, and cost-effective "second opinion" to aid physicians with the diagnostic management of the breast cancer.

The aim of this work is to develop an automatic system of analysis of mammographic images leading to the recognition of the normal from abnormal (benign and malignant lesions) cases. The introductory step in this procedure is to detect the lesion and define the region of interest (ROI) in the analyzed area of the image. These ROI images will be used in further processing by applying deep learning in the form of convolutional neural networks (CNN). Direct application of these images to CNN is not fully successful since the accuracy and sensitivity are limited. Therefore, in this paper we propose to enrich the input data by the images created on the basis of the non-negative matrix factorization (NMF) and statistical self-similarity. They fulfill the significant role in the classification system and allow increasing the accuracy of the image recognition. The results of numerical experiments performed on very large DDSM data base [4] containing more than 10000 mammograms have confirmed good accuracy of class recognition, exceeding the best results reported in the actual publications for the investigated DDSM data base containing large number of mammograms.

2. DATA BASE

The numerical investigations have been carried out using the publically available database of mammographic images "Digital Database for Screening Mammography" (DDSM), the largest publicly available breast cancer database, created in University of South Florida [4]. The dataset contains 2604 cases, each composed of 4 mammograms (left and right breast from above representing Cranial-Caudal view and oblique representing Medio-

Lateral-Oblique view). In this paper the ROI images were used. Fig. 1 presents the examples of mammograms representing the normal (Fig. 1a, b) and abnormal cases: benign (Fig, 1c, d) and malignant (Fig. 1e, f). We can observe the significant differences among the images representing the same classes of data.

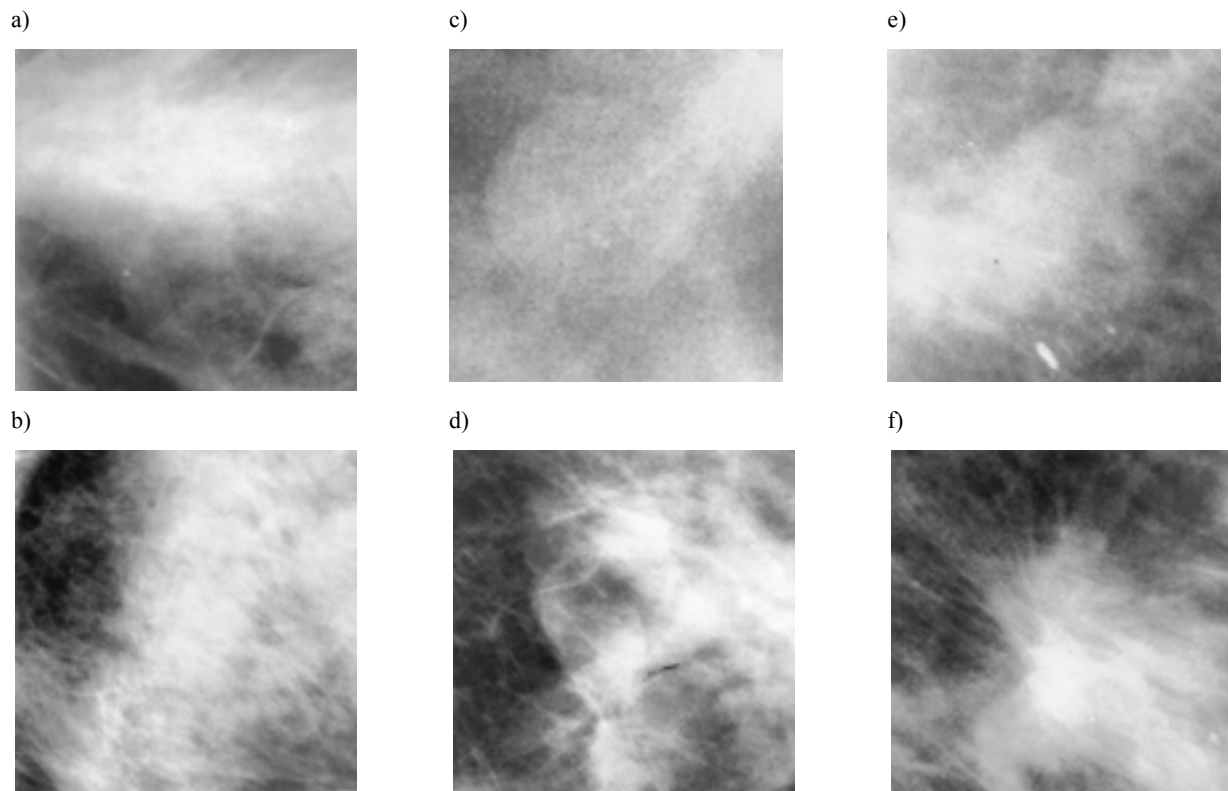a)   c)   e)

b)   d)   f)

Figure 1. The examples of mammograms (ROI) representing normal (a, b), benign (c, d) and malignant (e, f) breast cancer

From each mammographic image the region of interest (ROI) has been selected. The ROI for cancer changes has been defined as a rectangular area with the lesions observed in the center. For normal cases the ROI was the region extracted manually from normal tissues. The size of ROI images was 128×128 pixels. The total number of ROI images used in experiments was 11218.Their composition was as follows: normal tissues – 9215, benign lesions – 888 and malignant lesions – 1115.

Our work is directed to recognize the normal cases from abnormal. The developed system was tested in the recognition of three classes (normal, benign and malignant) and also on two classes, recognizing normal from abnormal (benign and malignant joined together). The recognition of mammograms will be done using the deep learning technique based on convolutional neural network. In this approach the extraction of features is made by CNN directly on the images delivered to its input. However, to increase the information of the class differences the additional preprocessing of images is done by applying NMF and so called statistical self-similarity approach. They deliver additional representations of the mammograms, enhancing the diagnostic information contained in the original data base.

3. IMAGE REPRESENTATION USING NON-NEGATIVE MATRIX FACTORIZATION

Non-negative matrix factorization decomposes the given matrix **P** is into two matrices **W** and **H**, i.e.,

$$\mathbf{P=WH} \tag{1}$$

With the property that all three matrices have no negative elements [5,6]. The columns of **W** are called basis vectors and the columns of **H** represent the encodings, which are in one-to-one correspondence with the basis vectors of **W**. Let us assume the size of the matrix **P** as $N{\times}M$, where N represents the size of original vectors forming matrix **P** and

M is the number of such vectors. The size of the factors **W** and **H** are $N \times r$ and $r \times M$, respectively. The factorization rank *r* is chosen by the user and in general it satisfies the condition $(N+M)r<NM$. We have applied the NMF algorithm of Cichocki [5] implemented in Matlab [7] to the analyzed images, represented in Hilbert vector form.

The NMF decomposition of mammograms is performed to enhance the representation of the images. Each image is represented in the form of Hilbert matrix depicted in a vector form. All these vectors corresponding to the mammograms in data base form the matrix **P**. The NMF operation is performed on the set of representative images of normal cases, since these images are more uniform. Half vectors belonging to the normal class has been used in this step of processing. The matrix **P** is decomposed into **W** and **H** non-negative matrices, where columns of **W** (the size $N \times r$) represent the basis vectors and **H** (the size $r \times M_1$) the encodings. The rank *r* is assumed by the user. It means, that *i*th vector $\mathbf{p}_i$ (the *i*th column of **P**) is represented as the weighted sum of basis vectors, which may be presented in Matlab notation [7] as following

$$\mathbf{p}_i = \sum_{j=1}^{r} \mathbf{W}(:,j)\mathbf{H}(j,i) \tag{2}$$

The set of original images is first converted to the NMF factors and then is reconstructed on the basis of limited number *r* of basis vectors (in our case *r*=10). The reconstruction of all analyzed images (normal and abnormal) is performed on the basis vectors formed in the NMF decomposition phase, which was performed using only normal cases. Such reconstruction will better represent the images representing normal cases. The abnormal images reconstructed on the basis vectors corresponding to normal cases only will show larger difference to the original ones. The main idea of reconstructing each image using only "healthy components" is to approximate an image as it is "healthy". This form of information enhances the differences between normal and abnormal details.

Fig. 2 presents some examples of transformation of original images by applying the NMF principle of reconstruction. Fig. 2a and b show the normal cases (original and reconstructed images), Fig. 2c and d – benign lesions and Fig. 2e and f – the malignant lesions (the originals and their reconstructions). The images after NMF are more smooth and uniform within the same class. This means easier classification of the testing samples not taking part in learning process (increase of the generalization ability of classifier).
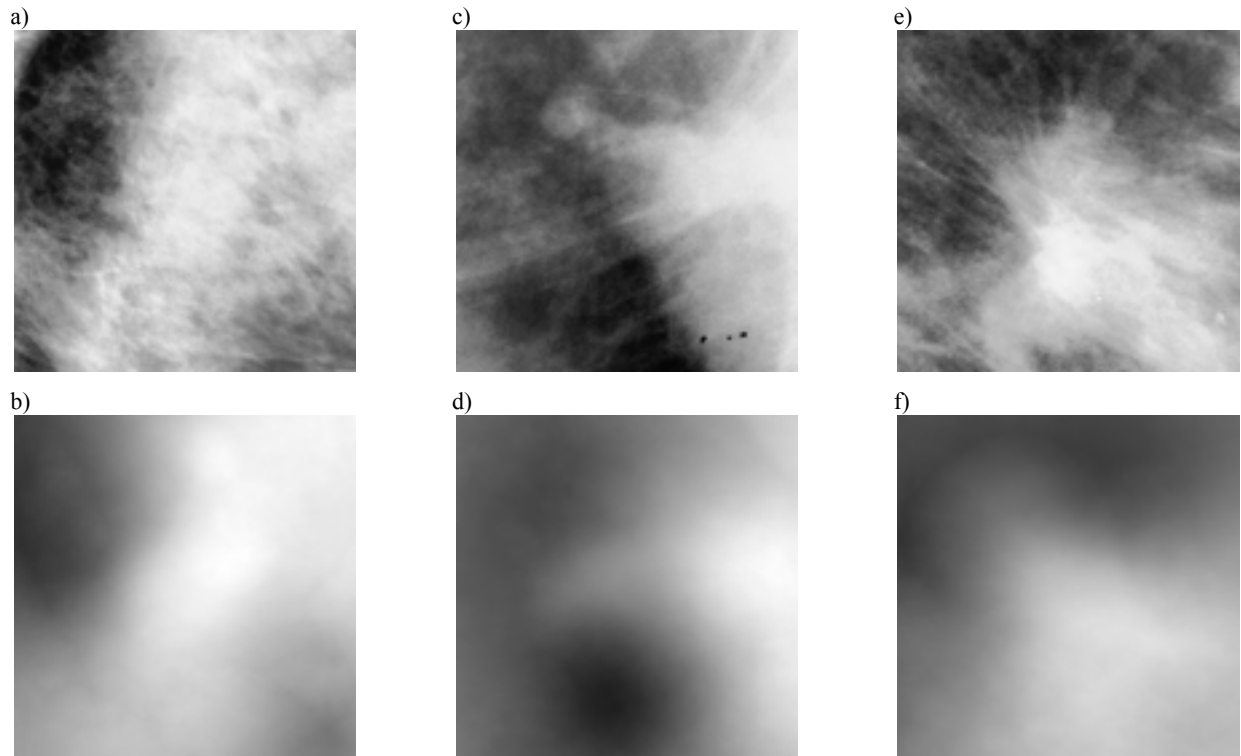

Figure 2. The illustration of the results of NMF reconstruction: a) original image of normal case, b) NMF reconstruction of image of Fig. a, c) original image of benign lesions, d) NMF reconstruction of image of Fig. c, e) original image of malignant lesions, f) NMF reconstruction of image of Fig. e.

4. THE IMAGE REPRESENTATION ON THE BASIS OF STATISTICAL SELF-SIMILARITY

The third type of images applied to the convolutional neural network is created using the so called statistical self-similarity of the image to itself. It is defined on the basis of pixel intensity distribution in small regions compared to the whole image. First, the image is resized to the dimension 1024×1024 and then split into 5×5 compact regions.

As a result we get the representation of the image by 256×256 small sub-images. In the next step the similarity of these sub-images to the whole image is determined using the Kolmogorov-Smirnov (KS) statistical distance [7]. Each sub-region is represented now by the appropriate value of 1-KS distance extending from 0 to 1. The lower value of KS, the more similar is the sub-image to the whole image. In final stage this 256×256 KS image is scaled back to the original dimension of 128×128.

Fig. 3 presents the examples of the normal and abnormal cases, representing the original ROI and their corresponding images obtained from self-similarity representation. The transformed images have higher contrast compared to the original images. Thanks to this property, the small details differing representatives of various classes are enriched, which helps in their recognition.

a)   c)   e)

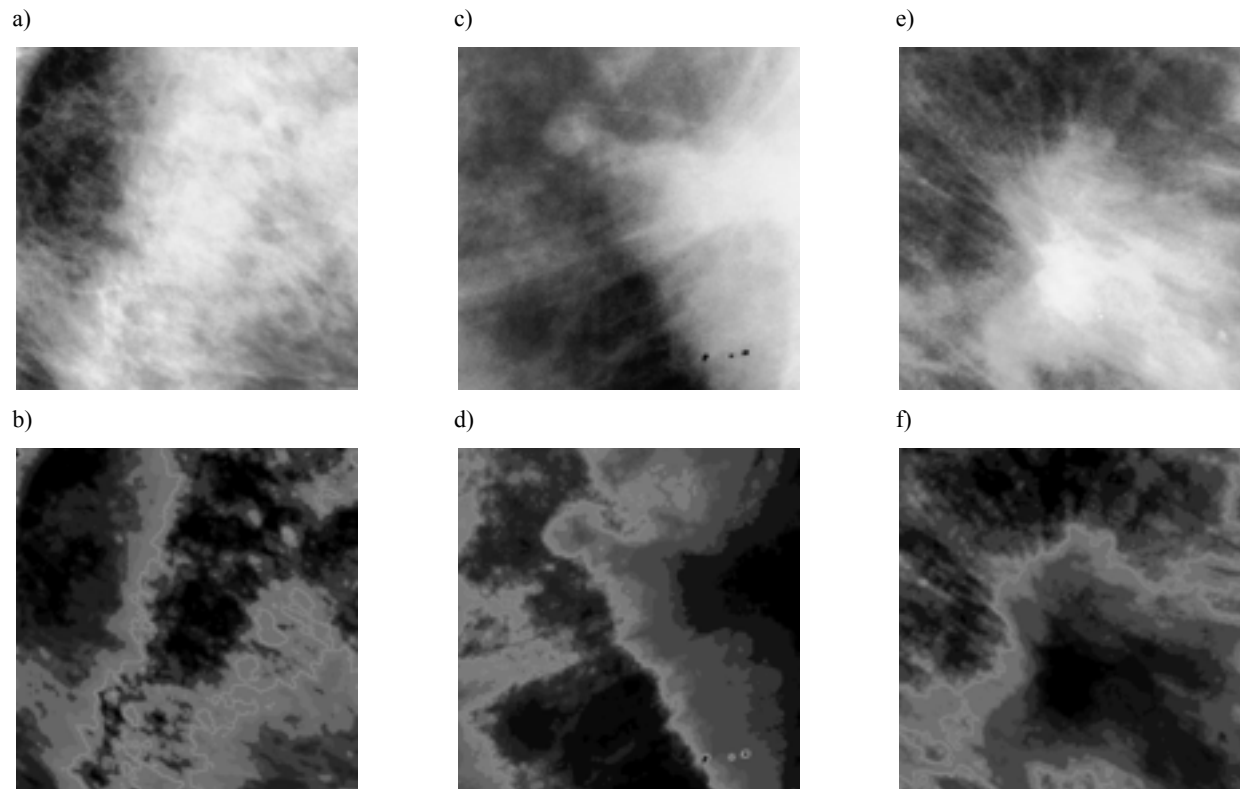b)   d)   f)

Figure 3. The illustration of the results of self-similarity reconstruction: a) original image of normal case, b) the reconstruction of image of Fig. a, c) original image of benign lesions, d) the reconstruction of image of Fig. c, e) original image of malignant lesions, f) the reconstruction of image of Fig. e.

5. RECOGNITION OF MAMMOGRAMS USING DEEP LEARNING

Deep learning is a very powerful tool in signal and image recognition [8]. It represents the group of algorithms applying the high-level abstraction models by using multiple hidden layers with very complex nonlinear structures.

Deep learning approach to image recognition is based on distributed representation of pixels and the interactions of factors organized in layers. These layers of factors correspond to different levels of abstraction or composition. Deep learning exploits the idea that higher level, more abstract concepts are learned from the lower level ones.

In this work the convolutional neural network (CNN) containing the multiple layers has been used [8]. Each neuron represents group of pixels. It receives inputs from a set of units located in a small neighborhood in the previous layer and is responsible for processing it. These neurons extract the elementary visual features, such as oriented edges, end points, corners, etc. The features are combined by the subsequent layers to detect features of the higher order, which form the input signals to the output classification layer. The output is made by the softmax, which produces the final class recognition unit. Softmax layer calculates the output value based on the multinomial logistic regression objective, maximizing the average across training cases of the log probability of the correct class label [8]. The class of the highest probability is regarded as the winner.The applied network architecture is as following: input in the form of three RGB representations of the images, three convolution layers and two Fully Connected layers on the output. The details of the solution are given in the Table 1.

Table 1. The details of the applied convolutional neural network.

no	Layer	Details
1	Image Input	128x128x3 images with zero-center normalization
2	First Convolution Layer	32 filters 5x5x3 convolutions with stride 1x1 and padding 2x2
3	Max Pooling	3x3 max pooling with stride 2x2 and padding 0x0
4	Activation	ReLU
5	Second Convolution Layer	32 filters 5x5 convolutions with stride 1x1 and padding 2x2
6	Activation	ReLU
7	Average Pooling	3x3 average pooling with stride 2x2 and padding 0x0
8	Third Convolution Layer	64 filters 5x5 convolutions with stride 1x1 and padding 2x2
9	Activation	ReLU
10	Average Pooling	3x3 average pooling with stride 2x2 and padding 0x0
11	First Fully Connected Layer	64 fully connected neurons
12	Activation	ReLU
13	Second Fully Connected Layer	2 fully connected neurons
14	Activation	Softmax
15	Classification Output cross-entropy	

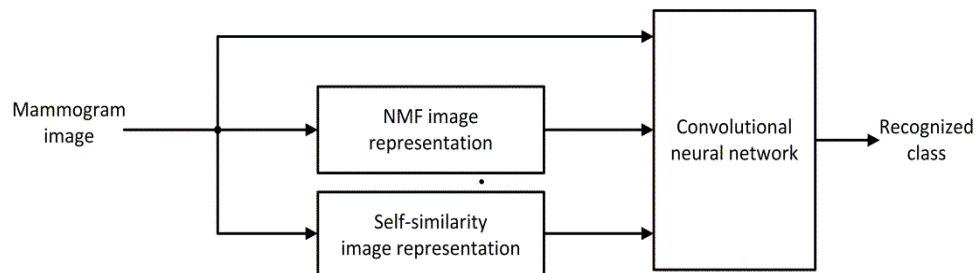The general view of the mammogram recognition system applied in experiments is presented in Fig. 4.

Figure 4. The applied system for mammogram recognition.

The convolutional neural network is supplied by the stream of three images: the original one, the image reconstructed on the basis of NMF and the image reconstructed from the self-similarity. In this way the number of

different representations of the images is triplicated. Thanks to this enhanced representation of the database the probability of correct classification is increased.

6. NUMERICAL RESULTS OF EXPERIMENTS

The numerical experiments of mammogram recognition have been performed using the DDSM, the largest publicly available breast cancer database of University of South Florida [4].Two types of experiments have been performed: recognition of three classes (class 1 – normal, class 2 – benign and class 3 – malignant) and recognition of normal cases (class 1) from abnormal (combined benign and malignant) representing class 2.

The experiments have been performed in 10-fold cross validation approach. The dataset was split into 10 subsets, each containing the same proportion of both classes, related to their populations in the database. Nine parts of samples are used in learning the CNN classifier and the last one used for testing the learned system. The same experiments were repeated ten times, exchanging the testing subset. In this way each fold served for testing only once. The training and testing sets included ROIs from completely different mammograms. Only half of the normal cases in the learning sets have been used in NMF decomposition to get the basis vectors of **W** used in reconstruction of all mammogram images.

The procedure of class recognition was performed in the following way. The learning set (90% of data used in 10-fold cross validation) was used to learn the classifier. The continuous output of the classifier in the learning mode (before binarization) was subject to the dynamic thresholding at different values of the threshold. The threshold generating the highest quality measure for learning data was fixed and used in the testing mode. Different quality measures might be taken into account: the sensitivity of recognizing the abnormal cases, total accuracy, the largest value of area under curve (AUC) of receiver operating characteristic (ROC) [9], etc. In this work the quality measure was defined on the basis of sensitivity and specificity (the inferior value of them on the learning data was maximized).

Table 2 presents the detailed results of recognition of three and two classes of mammograms obtained for testing data in 10-fold cross validation approach. The results representing the sensitivity, specificity, average accuracy and AUC are given. The sensitivity of recognizing the abnormal case from normal (2-class problem) was equal 82.28%. In three class problem (malignant, benign, normal) the sensitivity of recognition of malignant lesions has achieved the value of 73.36%.

Table 2. The results of numerical experiments of mammogram recognition using CNN deep learning.

	AUC	Sensitivity	Specificity	Accuracy
abnormal vs normal	0.919	82,28%	86,59%	85,82%
malignant vs non-malignant	0.909	73,36%	86,01%	84,75%

These results belong to the best for this large DDSM database. The same data base was used in 2015 by authors of [10]. The sensitivity of recognizing the malignant cases reported in this paper was 60.43%. Our results 73.36% is significantly better. To assess the importance of inclusion the NMF and self-similarity images in recognition process the additional experiments have been performed using only the original (raw) images of mammograms, without NMF and self-similarity representations. The obtained AUC results for normal vs abnormal and non-malignant vs malignant was 0,902 and 0,893 respectively. This difference is significant according to one-sided *signrank* test for crossvalidated pairs of AUC values (at 5% significance level).

7. CONCLUSIONS

The paper has presented an application of a deep learning for the recognition of three classes of mammogram images. Deep learning in the form of the convolutional neural network, applied in recognition of the images, was found very efficient tool in image analysis. This high quality of operation was possible thanks to application of

additional image processing using NMF and self-similarity principle. The obtained results belong to the best that were already reported for this large DDSM data base.

REFERENCES

[1] A. Jalalian, S.B. Mashohor, H.R. Mahmud, M.I. Saripan, A.R. Ramli and B. Karasfi, "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound," Clinical Imaging **37**, 420-426 (2013)

[2] M. Lobbes, M. Smidt, K. Keymeulen, R. Girometti, C. Zuiani, R. Beets-Tan, J. Wildberger and C. Boetes, "Malignant lesions on mammography: Accuracy of two different computer-aided detection systems," Clinical Imaging **37**, 283-288 (2013)

[3] B.K. Elfarra and I.S. Abuhaiba, "Mammogram computer aided diagnosis," International Journal of Signal Processing, Image Processing and Pattern Recognition **5**(4), 1-29 (2012)

[4] M. Heath, K. Bowyer, D. Kopans, R. Moore and P. Kegelmeyer, "The digital database for screening mammography," in Digital Mammography, Springer, Netherlands, 457-460 (1998)

[5] A Cichocki, R Zdunek, A.H. Phan and S Amari, Nonnegative Matrix And Tensor Factorizations: Applications To Exploratory Multi-Way Data Analysis And Blind Source Separation, Wiley, N.Y. (2009)

[6] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature **401**, 788-791 (1999)

[7] Matlab User Manual – Image Processing Toolbox, MathWorks, Natick (2015)

[8] A. Krizhevsky, I. Sutskever and G. Hinton, "Image net classification with deep convolutional neural networks," Advances in Neural Information Processing Systems **25**, 1-9 (2012)

[9] P.N. Tan, M. Steinbach and V. Kumar, Introduction to Data Mining, Pearson Education Inc., Boston (2006)

[10] S. Dhahbi, W. Barhoumi and E. Zagrouba, "Breast cancer diagnosis in digitized mammograms using curvelet moments," Computers in Biology and Medicine **64**(1), 79-90 (2015)