Non-negative Dimensionality Reduction for Mammogram Classification

I. Buciu and A. Gacsadi
Department of Electronics,
University of Oradea, Faculty of Electrical Engineering and Information Technology,
410087, Oradea, E-Mail: {ibuciu, agacsadi}@uoradea.ro

Abstract* – Directly classifying high dimensional data may exhibit the ``curse of dimensionality'' issue that would negatively influence the classification performance with an increase in the computational load, depending also on the classifier structure. When working with classifiers not affected by this issue (such as Support Vector Machines, for instance), the computational load still exists due to the required time in computing the kernel matrix. Moreover, the performance is affected when a few samples per class is available for the training procedure. One common solution is to carry out a feature extraction step for reducing the data dimension prior to classification. The paper describes the application of Nonnegative Matrix Factorization (NMF) for extracting features from mammogram medical images with different resolution, further used for recognizing breast tumors. For comparison, Principal Component Analysis (PCA) and Independent Component Analysis (ICA) were explored. Experiments show that NMF method outperforms PCA and ICA, leading to higher classification accuracy.

***Keywords:* non-negative matrix factorization, feature extraction, mammogram classification.**

I. INTRODUCTION

Dimensionality reduction implies either feature selection or feature extraction [1]. While feature selection refers to selecting subsets from initial data, feature extraction rather deals with data transformation.

Mammographic images are usually high dimension images. A mammogram is basically an X - ray capture of the breast region that displays points with bigger intensities that are suspected of being tumors. Artifacts appearing in the mammogram could indicate a potential presence of a benign or malignant tumor. Important visual clues of breast cancer include preliminary signs of masses and calcification clusters. Masses and calcium deposits can be easily identified by visual inspection in X-ray images, as they are much denser (highly attenuate X-ray) than all other types of soft tissues around. Unusual smaller and clustered calcifications are associated with malignancy while there are other calcifications (diffuse, regional, segmental and linear) that are typically benign. Such calcifications are termed as microcalcifications. Automatic tumor detection is extremely challenging as the suspicious calcifications or masses appear as free shape and no precise pattern can be associated to them. Moreover, the presence of more or less prominent blood vessels and muscle fibers makes the issue to become harder to deal with. Several techniques have been proposed to analyze or to extract features from mammogram images. Ferreira and Borges [2], [3], describe a method based on wavelet decomposition. The images are decomposed in wavelet basis. Using a minimum subset of representative wavelet features of image textures based on a specific threshold, the work investigated different wavelet bases, variation of the selection strategy for the coefficients, and different metrics. Haar wavelets and PCA were proposed by Swiniarski et al [4] to extract relevant features, and rough sets methods are employed for classifying those features. Recently, the authors extended the work by extracting ICA features [5]. To detect microcalcifications, Lemaur et. al. [6] developed a method based on wavelets derived from Matzinger polynomial with high Sobolev regularity index, suitable for detecting singularities in image. A multiresolution statistical model has been proposed by Strickland and Hahn [7]. It should be mentioned that, while wavelets based methods are rather used to enhance the image and analyze it from various frequency perspectives and do not actually reduce data dimension, PCA or ICA approach does. Sheshadri and Kandaswamy [8] attempted to classify breast tissue using simple image statistics such as its intensity level of histogram. Statistical features extracted such as mean, standard deviation, smoothness, third moment, uniformity and entropy are employed to classify breast tissue into four basic categories like fatty, uncompressed fatty, dense and high density. The extracted features are next classified using various classifiers. A binary tree classifier based on the use of a 2-D Quincux wavelet transform is developed by Sun et al. [9], while Wei et al. [10] employed Support Vector Machines (SVMs), Relevance Vector Machines (RVM) and Kernel Fisher Discriminant (KFD). A classification model based on association rules was developed by Zaiane et al in [11].

In this paper, we investigate the suitability of using Non-negative Matrix Factorization (NMF) method for extracting discriminant features from mammogram image samples and, in the same time, to reduce data dimension. Compared to PCA or ICA, NMF features seem to have more discriminant power in terms of classification accuracy.

II. NON-NEGATIVE MATRIX FACTORIZATION

NMF approach [12] is a method that leads to parts-based representation, and, unlike other decomposition methods (such as PCA, ICA, etc) allows only additive, not subtractive combinations of the original data. NMF is an unsupervised learning method that imposes non-negative factorization, i.e., both decomposition factors have only non-negative entries. Given n non-negative input m - dimensional vectors $\mathbf{a}_i$ ($i = 1, \dots, n$), each vector representing the pixel intensity values of the image lexicographically scanned and stored in the columns of a matrix $\mathbf{A}$, NMF decomposes this matrix into two matrices, $\mathbf{W}$ of dimension m x r and $\mathbf{P}$ of size r x n so that their product approximates, to some extent, the original matrix $\mathbf{A}$. Here, the r columns of $\mathbf{W}$ are named NMF bases images and the rows of $\mathbf{P}$ represents their encoding coefficients. The rank of the factorizations $\mathbf{W}$ and $\mathbf{P}$ is typically chosen such that $(n + m)$ r < nm. Hence, the compression of data is achieved and the compression ratio of NMF is provided by $nm/(n + m)\, r$. The low-dimensional representation of the m - dimensional original vector $\mathbf{a}_i$ is thus given by the r - dimensional vector $\mathbf{p}_i$. Each original image can be reconstructed as $\mathbf{a}_i = \mathbf{W}\,\mathbf{p}_i$.

The quality of reconstruction depends on the cost function associated to the decomposition process. Two cost functions are usually employed: Kullback-Leibler divergence

$$KL(a \| Wp) = \sum_j \left(a_j \log \frac{a_j}{W_{jk} p_k} - a_j + \sum_k W_{jk} p_k \right)$$

and squared Euclidean distance $D(a \| Wp) = \sum_j \left\| a_j - \sum_k W_{jk} p_k \right\|^2$ between $\mathbf{a}_i$ and its $\mathbf{W}\,\mathbf{p}_i$ for $j = 1 \dots m$ and $k = 1 \dots r$.

Since its first development, NMF knew a huge interest from the scientific community, due to its simplicity and intuitive decomposition. NMF applications include image processing (face and facial expression recognition, medical imaging, etc.), audio data processing or text mining and decomposition [13]. In the medical field, NMF was applied for discovering metagenes and molecular patterns in a genomic signal processing task [14], for decomposing multichannel EEG signals [15], or for source spectra separation from magnetic resonance chemical shift imaging of human [16].

III. DATA DESCRIPTION

The data samples used in our experiments were taken from the Mammographic Image Analysis Society (MIAS) [17]. The database contains 322 samples that belong to three categories: normal, benign and malign. There are 208 normal images, 63 benign and 51 malign cases, which are considered abnormal. Each image is centered and its size is of *1024* x *1024* pixels. The abnormal cases are further divided into six categories: microcalcification, circumscribed masses, spiculated masses, ill-defined masses, architectural distortion and asymmetry. However, we only considered the three classes above mentioned. For each abnormal case, the coordinates of center of abnormality are provided along with the approximate radius (in pixels) of a circle enclosing the abnormality. The widest identified abnormality corresponds to a radius of 197 pixels, while the tightest correspond to a radius of 3 pixels. In some cases calcifications are widely distributed throughout the image rather than concentrated at a single site. Here, the center locations and radii have been omitted. Knowing the location and the approximate size of abnormality allows us to extract subimages (patches) with proper dimension representing the tumor zone.

III. EXPERIMENTAL SETTING AND RESULTS

To discard irrelevant information like the breast contour, patches of *140* x *140* pixels surrounding the abnormality region were extracted from the original *1024* x *1024* pixels images. The patches size assures that, for most abnormal cases not only the abnormality region is captured but also the surrounding area, providing us information about the abnormality shape.

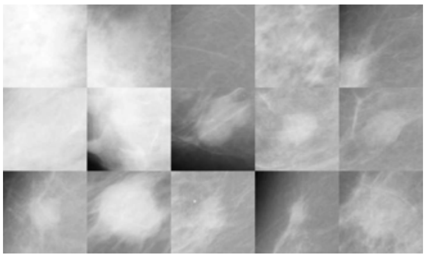

Fig.1 Patches of *140* x *140* pixels extracted from mammographic images. Top row represents 5 samples for the normal case, middle row represents 5 samples for the benign case, while the bottom row illustrates 5 samples for the malign case.

For the normal case, the patches were extracted from the middle of the breast images. Fig. 1 illustrates 5 samples per case. Finding the final NMF decomposition factors is a time consuming iterative process. Therefore, to reduce the computational load we downsized the patches to *60* x *60*, *40* x *40* and *20* x *20* pixels, respectively, prior to NMF. We further split the mammographic data into training and disjoint test set. We picked up 80 \% samples to form the training set and the remaining samples are included in the test set. Each image patch was reshaped into m = 3600 dimensional vector (for the patches of *60* x *60* size) and stored in the columns of matrix $\mathbf{A}$.

TABLE I. MAXIMUM CLASSIFICATION RATE EXPRESSED IN PERCENTAGE (%) FOR ALL FOUR METHODS INVESTIGATED IN THE PAPER AND CSM CLASSIFIER.

Patch size	MCC			
	NMF	PCA	ICA I	ICA II
60 x 60	**67.1**	57.8	57.8	59.3
60 x 60	**65.6**	57.8	57.8	60.9
60 x 60	**65.6**	57.8	57.8	59.3

TABLE II. MAXIMUM CLASSIFICATION RATE EXPRESSED IN PERCENTAGE (%) FOR ALL FOUR METHODS INVESTIGATED IN THE PAPER AND PSVM CLASSIFIER.

Patch size	PSVM			
	NMF	PCA	ICA I	ICA II
60 x 60	**70.8**	67.1	68.7	68.7
60 x 60	**68.7**	67.1	67.1	65.6
60 x 60	**68.7**	65.6	64	59.3

The dimensionality reduction algorithms (NMF, PCA and ICA) were applied for several decomposition ranks, more precisely, $r = \{5, 10, 20, 30, \dots, 100\}$. Ten NMF bases images are depicted in the top row of Fig 2. Assuming zero mean input (training or test) data, the NMF training feature vectors are comprised in the columns of $\mathbf{F}_{tr(NMF)} = \mathbf{W}^{-1}\mathbf{A}_{tr}$ where $\mathbf{A}_{tr}$ is the zero mean training set.

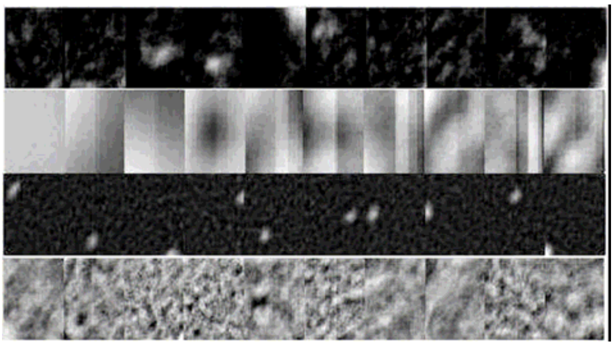

Fig.2 A number of 10 basis images out of 100, corresponding to NMF on the top row, PCA on the second row, and ICA architecture I and II, respectively, on the last two bottom rows.

Whenever a new zero mean unseen patch $\mathbf{a}_{test}$ comes, its corresponding feature vector is formed in a similar way, i.e., $\mathbf{f}_{test(NMF)} = \mathbf{W}^{-1}\mathbf{a}_{test}$. PCA feature vectors are formed by projecting the data into the PCA eigenvectors (PCA basis images illustrated in the second row of Fig. 2 i.e., $\mathbf{F}_{tr(PCA)} = \mathbf{V}^{T}_{r}\ \mathbf{A}_{tr}$ where $\mathbf{V}^{T}$ is the r - rank PCA projection matrix. For the ICA method two architectures exist, depending on the image storing type in $\mathbf{A}$ [18]. When images are stored in the rows of $\mathbf{A}$, we have the first ICA architecture, denoted here as "ICA I", while the second architecture, "ICA II", considers the images in the columns of $\mathbf{A}$. After ICA training, the corresponding basis images are retrieved, as depicted in the last two rows of Fig. 2 for the two architectures. ICA feature vectors are comprised in the rows of $\mathbf{F}_{tr(ICA)} = \mathbf{A}^{T}\ \mathbf{V}_{r}\ \mathbf{B}^{-1}$, where $\mathbf{B}$ is the unmixing matrix found by ICA algorithm. Next, two classifier types are employed: a simple Euclidean distance measure based classifier, named Maximum Correlation Classifier (MCC) and Proximal Support Vector Machines (PSVM) [19] with polynomial kernel of degree *1,2* and *3*. As PSVM is typically designed for two-class problem, a "one against all" strategy was applied for our three-class case.

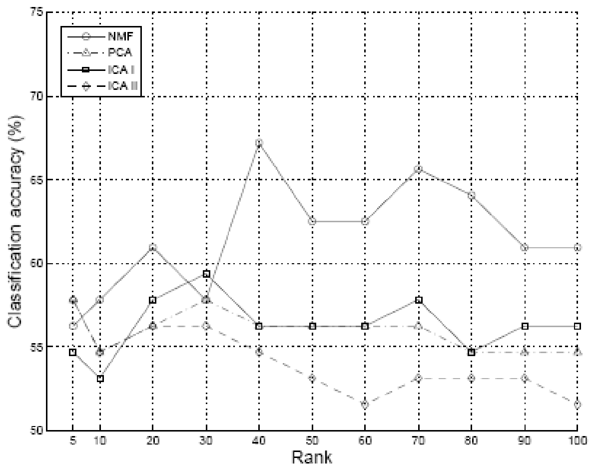

Fig.3 Classification accuracy attained with MCC for image patches of 60 x 60 pixels and different numbers of dimensions (ranks).

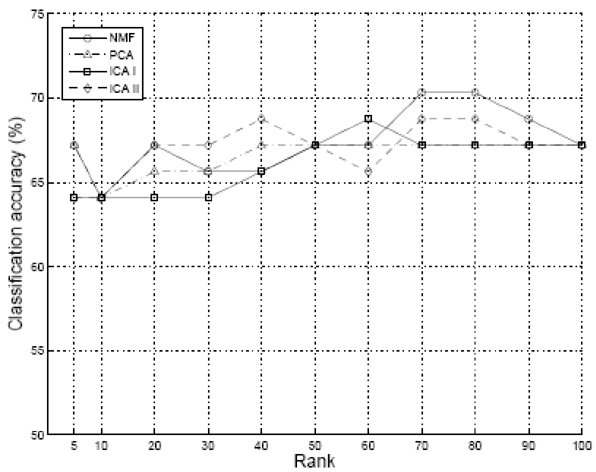

Fig.4 Classification accuracy attained with PSVM for image patches of 60 x 60 pixels and different numbers of dimensions (ranks).

Fig. 3 and 4 shows the variation of the classification accuracy corresponding to MCC and PSVM, respectively, for different low dimensions. As can be noticed for the MCC classifier, while all feature extraction methods lead to approximately the same classification accuracy up to rank r = 30, the PCA and ICA classification performance gets lower with an increase in the dimension rank. On the other side, an increase in the NMF rank conducts to higher classification accuracy. As expected, using more complex classifiers such as PSVMs leads to better classification results than the ones corresponding to MCC, for all feature extraction methods, but it requires more dimensions. Table I and II tabulate the maximum classification accuracy in percentage for different patch sizes, obtained by each feature extraction method coupled with MCC and PSVM, respectively. When MCC is employed, PCA and ICA conducted to the same accuracy value regardless of the patch size. NMF clearly outperforms both PCA and ICA in terms of classification accuracy with a maximum of 67.1 % compared to 57.8 % corresponding to PCA or ICA. NMF seems to be more sensitive to the dimension patch with a performance decrease from 67.1 % to 65.6 % associated to downsampling from *60 x 60* to *40 x 40* pixels. Overall, the best performance is obtained with linear PSVMs, where the classification accuracy gets higher. NMF features associated to PSVMs again yielded the highest accuracy.

III. CONCLUSIONS

Dimensionality reduction for high dimension data is a necessary step for any pattern classification system. Particularly, this paper dealt with reducing the dimension of mammographic images through applying the NMF feature extraction technique. The extracted features were further classified into three classes: normal, benign and malign. For comparison purpose, PCA and ICA features were also extracted as baseline in classifying the features. The preliminary experimental results indicate that NMF is able of retrieving more discriminant features than PCA or ICA, leading to better classification performance.

ACKNOWLEDGMENT

This work was supported by a grant from the Romanian National University Research Council, PNCDI Programe ID- 668/2009.

REFERENCES

[1] Jain, A. J., Duin, R. P. W, Mao, J.: Statistical Pattern Recognition: A Review. IEEE Trans. on Pattern Analysis and Machine Learning, 22(1), pp. 4–37 (2000).

[2] Ferreira, C. B. R., Borges, D. L.: Analysis of Mammogram Classification Using a Wavelet Transform Decomposition. Pattern Recognition Letters, 24, pp. 973–982 (2003)

[3] Ferreira, C. B. R., Borges, D. L.: An Evaluation of Wavelet Features Subsets for Mammogram Classification. Progress in Pattern Recognition, Image Analysis and Applications 3773, pp. 620–630 (2005)

[4] Swiniarski, R., Luu, T., Swiniarska, A., Tanto, H.: Data Mining and Online Recognition of Mammographic Images Based on Haar Wavelets, Principal Component Analysis and Rough Sets Methods. International SPIE Symposium Medical Imaging, pp. 17–23 (2001)

[5] Swiniarski, R., Lim, H. K., Shin, J. H., Skowron, A.: Independent Component Analysis, Principal Component Analysis and Rough Sets in Hybrid Mammogram Classification. Proc. of the 2006 International Conference on Image Processing, Computer Vision, Pattern Recognition II, pp. 640–645 (2006)

[6] Lemaur, G., Drouiche, K., DeConinck, J.: Highly Regular Wavelets for the Detection of Clustered Microcalcifications in Mammograms. IEEE Trans. on Medical Imaging, 22, pp. 393–401 (2003)

[7] Strickland, R. N., Hahn, H.: Wavelet Transforms for Detecting Microcalcifications in Mammograms. IEEE Trans. on Medical Imaging, 15, pp. 218–229 (1996)

[8] Sheshadri, H. S., Kandaswamy, A.: Breast Tissue Classification Using Statistical Feature Extraction Of Mammograms. Medical Imaging and Information Sciences, 23(3), pp. 105–107 (2006)

[9] Sun, Y., Babbs, C.F. Delp, E.J.: Normal Mammogram Classification based on Regional Analysis. The 2002 45th Midwest Symposium on Circuits and Systems, 2, pp. 375–378 (2002)

[10] Wei, L., Yang, Y, Nishikawa, R. M., Jiang, Y.: A Study on Several Machine-Learning Methods for Classification of Malignant and Benign Clustered Microcalcifications. IEEE Trans. on Medical Imaging, 24(3), pp. 371–380 (2005)

[11] Zaiane, O. R., Antonie, M.-L., Coman, A.: Mammography Classification by an Association Rule-based Classifier. In Third International ACM SIGKDD workshop on multimedia data mining (MDM/KDD'2002) in conjunction with eighth ACM SIGKDD, pp. 62–69 (2002)

[12] Lee, D. D., Seung, H. S.: Learning the Parts of the Objects by Nonnegative Matrix Factorization. Nature, 401, pp. 788–791 (1999).

[13] Buciu, I.: Non-negative Matrix Factorization, A New Tool for Feature Extraction: Theory and Applications. International Journal of Computers Communications and Control, 3, Suppl. S, pp. 67–74 (2008).

[14] Brunet, J. P., Tamayo, P., Golub, T. R., Mesirov, J. P.: Metagenes and Molecular Pattern Discovery using Matrix Factorization. Proc. Natl. Acad. Sci U S A, 101(12), pp. 4164–4169 (2004)

[15] Rutkowski, T. M., Zdunek, R., Cichocki, A.: Multichannel EEG Brain Activity Pattern Analysis in Time-Frequency Domain with Nonnegative Matrix Factorization Support. International Congress Series, 1301, pp. 266–269 (2007)

[16] Sajda, P., Du, S., Brown, T., Stoyanova, R., Shungu, D., Mao, X., Parra, L.: Non-negative Matrix Factorization for Rapid Recovery of Constituent Spectra in Magnetic Resonance Chemical Shift Imaging of the Brain. IEEE Trans. on Medical Imaging, 23(12), pp. 1453–1465 (2004)

[17] Mammographic Image Analysis Society, http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html

[18] Bartlett, M. S., Movellan, J. R., Sejnowski, T. K.: Face recognition by independent component analysis. IEEE Trans. on Neural Networks, 13(6), pp. 1450–1464 (2002)

[19] Fung, G., Mangasarian, O. L.: Proximal Support Vector Machine Classifiers. Proceedings KDD-2001: Knowledge Discovery and Data Mining, pp. 77–86 (2001)