Independent Component Analysis, Principal Component Analysis and Rough Sets in Hybrid Mammogram Classification

R. Swiniarski
Department of Computer Science, IPIPAN
San Diego State University
San Diego CA, USA

Hun Ki Lim
Department of Computer Science
San Diego State University
San Diego CA, USA

Joo Heon Shin, San Diego State University

A. Skowron, Warsaw University, Poland

Abstract- The paper describes the hybrid methods of mammogram recognition which are based on independent component analysis, principal component analysis and rough set theory.

Keywords: pattern recognition, principal component analysis, independent component analysis, rough sets, mammogram recognition

1 Introduction

Mammogram recognition is one of the most difficult pattern recognition tasks. Several techniques have been developed for mammogram recognition [12]. One of the most prominent methods is based on application of the Haar wavelets for feature extraction from mammographic images [9]. Recently Independent Component Analysis (ICA) [5] has gained popularity as a promising method for discovering the statistically independent variables (sources) for data, blind source separation, as well as for feature extraction from images [11].

We have tried to explore the potential of ICA, combined with the rough sets method, as one of the processing steps in creating a hybrid method of classifier design for cancer recognition based on mammograms. For comparison, we have also studied the application of classic Principal Component Analysis (PCA) combined with rough sets method for extraction and selection of mammographic image features as well as for classification. For feature extraction, reduction, and selection as well as for the classifier design for mammographic images, we have applied the following hybrid sequences for processing operations:

1. Independent Component Analysis for feature extraction, reduction, and pattern forming of mammographic images.
2. Rough sets method for feature selection and data reduction.
3. Rough set-based method for rule-based classifier design.

A similar processing sequence has been applied using Principal Component Analysis (PCA) for feature extraction, feature reduction, and pattern forming of mammographic images.

In addition, in all experiments we have tried, for comparison study to design other classifiers, namely: error-backpropagation and Learning Vector Quantization (LVQ) neural networks.

2 An Application of Rough Sets Theory to Feature Selection and Rule-based Classification

The rough sets theory proposed by Professor Pawlak [6] provides a mathematically rigorous data mining technique for knowledge discovery in databases and experimental data sets. In rough sets, an information system can be presented in the form of the decision table

$$DT = < U, C \cup D, V, f >, \quad (1)$$

where U is the *universe*, a finite set of N objects $\{x_1, x_2, \cdots, x_N\}$, $Q = C \cup D$ is a finite set of *attributes*, C is a set of *condition* attributes, D is a set of *decision* attributes, $V = \bigcup_{q \in C \cup D} V_q$, where V_q is the set of *domain* (*value*) of attribute $q \in Q$, f

: $U \times (C \cup D) \rightarrow V$ is a total *decision function* (information function, decision rule in DT) such that $f(x,\, q) \in V_q$ for every $q \in Q$ and $x \in V$.

For a given subset of attributes $A \subseteq Q$, the $IND(A)$ (denoted by $\tilde{A}$)

$$IND(A) = \{(x,y) \in U : \text{for all} a \in A,\ f(x,a) = f(y,a)\} \quad (2)$$

is an *equivalence relation* on universe U (called an *indiscernibility relation*).

For a given decision table S, a given subset of attributes $A \subseteq Q$ determines the approximation space $AS = (U, IND(A))$ in S. For a given $A \subseteq Q$ and $X \subseteq U$ (a concept X), the *A-lower approximation* $\underline{A}X$ of set X in AS and the *A-upper approximation* $\bar{A}X$ of set X in AS are defined as follows:

$$\underline{A}X = \{x \in U : [x]_A \subseteq X\} = \bigcup\{Y \in A^* : Y \subseteq X\} \quad (3)$$

$$\bar{A}X = \{x \in U : [x]_A \cap X \neq \emptyset\} = \bigcup\{Y \in A^* : Y \cap X \neq \emptyset\} \quad (4)$$

One of the most essential notions of rough sets, related to a decision table, is a *reduct*. A reduct relates to a subset of attributes from an information system, which can discern all objects discernible by the original information system with all attributes. We have applied the rough sets reduct in the technique for feature reduction/selection of mammographic images.

In addition, we have utilized rough sets as a method for designing a *rule-based classifier*. Let $C^* = \{X_1, X_2, \cdots, X_r\}$ be C-definable classification of U, and $D^* = \{Y_i, Y_2, \cdots, Y_l\}$ be D-definable classification of U. A class Y_i from a classification D^* can be identified with the decision i ($i = 1, 2, \cdots, l$). Then i^{th} *decision rule* is defined as follows:

$$Des_C(X_i) \Longrightarrow Des_D(Y_j) \text{ for } X_i \in A^* \text{ and } Y_j \in D^* \quad (5)$$

And these decision rules are logically described as follows

if (a set of conditions), *then* (a set of decisions)

The set of decision rules for all classes $Y_j \in D^*$ is denoted as follows:

$$\{\tau_{ij}\} = \{Des_C(X_i) \Longrightarrow Dec_D(Y_j) : X_i \cap Y_j \neq \emptyset \text{ for } X_i \in A^*, Y_j \in D^*\} \quad (6)$$

The set of decision rules for all classes $X_i \in D^*$, $(i = 1, 2, \cdots, r)$ generated by set of decision attributes D (D-definable classes in S) is called the *decision algorithm* resulting from the information system S.

2.1 Rough Sets for Feature Reduction / Selection

One of possibilities for selecting features from feature patterns is to apply rough sets theory [10]. Specifically, defined in rough sets, computation of a reduct can be used for selection of some of extracted features (constituting a reduct) [6,7,10,12] as reduced pattern attributes. These attributes will describe all concepts in a training data set. We have used the rough sets method to find reducts from the discretized feature patterns and to select features forming the reduced pattern based on chosen reduct.

3 Independent Component Analysis for Feature Extraction and Reduction

Independent component analysis (ICA) is an unsupervised, computational and statistical method for discovering intrinsic statistically independent variables from data sets. The ICA model assumes that the observed sensory signals x_i are given as the pattern vectors $\mathbf{x} = [x_1, x_2, \cdots, x_n]^T \in \mathbf{R}^n$. The sample of observed patterns are given as a set of N pattern vectors $T = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$, which can be represented as a $n \times N$ data set matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N] \in \mathbf{R}^{n \times N}$ which contains patterns as its columns. The ICA model for the element x_i is given as the linear mixtures $x_i = \sum_{j=1}^{m} h_{i,j} s_j, \quad (i = 1, 2, \cdots, n)$, of m independent variables s_j. Here, x_i is an observed variable, s_j is the independent component (source signals) and $h_{i,j}$ are the mixing coefficients. The independent variables constitute the independent variable vector (source pattern) $\mathbf{s} = [s_1, s_2, \cdots, s_m]^T \in \mathbf{R}^m$. The ICA model can be presented in the matrix form $\mathbf{x} = \mathbf{H}\mathbf{s}$, where $\mathbf{H} \in \mathbf{R}^{n \times m}$ is $n \times m$ unknown mixing matrix where row vector $\mathbf{h}_i = [h_{i,1}, h_{i,2}, \cdots, h_{i,m}]$ represents mixing coefficients for observed signal x_i. The purpose of ICA is to estimate both the mixing matrix $\mathbf{H}$ and the independent component vectors $\mathbf{s}$ based on sets of observed vectors $\mathbf{x}$.

The ICA model for the set of N patterns from $\mathbf{X}$ can be written as $\mathbf{X} = \mathbf{HS}$, where $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N]$ is the $m \times N$ matrix whose columns correspond to independent component vectors $\mathbf{s}_i = [s_{i,1}, s_{i,2}, \cdots, s_{i,m}]^T$ discovered from the observation vector $\mathbf{x}_i$. Once the mixing matrix $\mathbf{H}$ has been estimated, we can compute its inverse $\mathbf{B} = \mathbf{H}^{-1}$, and then the independent component for the observation vector $\mathbf{x}$ can be computed by $\mathbf{s} = \mathbf{B}\mathbf{x}$.

Usually, ICA is preceded by preprocessing which includes centering and whitening. The purpose of whitening is to transform the observed vector $\mathbf{x}$ linearly so that we obtain a new vector $\mathbf{y}$ (which is white) whose elements are uncorrelated and their variances equal unity

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \quad so \quad E\{\mathbf{y}\mathbf{y}^T\} = \mathbf{I}_l, \quad (7)$$

where $\mathbf{y} \in \mathbf{R}^l$ is the $l-dimensional$ ($l \leq n$) whitened vector, and $\mathbf{W}$ is $l \times n$ whitening matrix. Whitening also allows dimensionality reduction by projecting of $\mathbf{x}$ onto the first l eigenvectors of the covariance matrix of $\mathbf{x}$.

Whitening is usually realized using the eigenvalue decomposition (EVD) of the covariance matrix $E\{\mathbf{x}\mathbf{x}^T\} \in \mathbf{R}^{n \times n}$ of observed vector $\mathbf{x}$

$$\mathbf{R_{xx}} = E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{E_x} \mathbf{\Lambda_x^{\frac{1}{2}}} \mathbf{\Lambda_x^{\frac{1}{2}}} \mathbf{E_x^T} \quad (8)$$

Here, $\mathbf{E_x} \in \mathbf{R}^{n \times n}$ is the orthogonal matrix of eigenvectors of $\mathbf{R_{xx}} = E\{\mathbf{x}\mathbf{x}^T\}$ and $\mathbf{\Lambda}$ is the diagonal matrix of its eigenvalues

$$\mathbf{\Lambda_x} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \quad (9)$$

with positive eigenvalues $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n \geq 0$. The whitening matrix can be computed as

$$\mathbf{W} = \mathbf{\Lambda_x^{-1/2}} \mathbf{E_x^T} \quad (10)$$

Recalling that $\mathbf{x} = \mathbf{H}\,\mathbf{s}$, we can find that

$$\mathbf{y} = \mathbf{\Lambda_x^{-1/2}} \mathbf{E_x^T}\ \mathbf{H}\,\mathbf{s} = \mathbf{H}_w \mathbf{s} \quad (11)$$

We can see that whitening transforms the original mixing matrix $\mathbf{H}$ into a new one, $\mathbf{H}_w$

$$\mathbf{H}_w = \mathbf{\Lambda_x^{-1/2}} \mathbf{E_x} \mathbf{H} \quad (12)$$

Whitening allows us to reduce the dimensionality of the whitened vector by projecting the observed vector on the first l ($l \leq n$) eigenvectors corresponding to the first l eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_l$ of the covariance matrix $\mathbf{E_x}$. Then, the resulting dimension of the matrix $\mathbf{W}$ is $l \times n$, and there is reduction of the size of observed transformed vector $\mathbf{y}$ from n to l.

The output vector of the whitening process can be considered as an input to the ICA algorithm. The whitened observation vector $\mathbf{y}$ is an input to the unmixing (separation) operation

$$\mathbf{s} = \mathbf{B}\mathbf{y}, \quad (13)$$

where $\mathbf{B}$ is an original unmixing matrix.

An approximation (reconstruction) of the original observed vector $\mathbf{x}$ can be computed as

$$\tilde{\mathbf{x}} = \mathbf{B}\mathbf{s}, \quad (14)$$

where $\mathbf{B} = \mathbf{W}_w^{-1}$.

For the set of N patterns $\mathbf{x}$ forming as columns in the matrix $\mathbf{X}$, we can provide the following ICA model

$$\mathbf{X} = \mathbf{B}\,\mathbf{S}, \quad (15)$$

where $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N]$ is the $m \times N$ matrix whose columns correspond to independent component vectors $\mathbf{s}_i = [s_{i,1}, s_{i,2}, \cdots, s_{i,m}]^T$ discovered from the observation vector $\mathbf{x}_i$.

The estimation of the mixing matrix and independent components has been realized using Karhunen and Oja's FastICA algorithm [5]. In this method, the following maximization criterion has been exploited

$$J(\tilde{\mathbf{s}}) = \sum_{i=1}^{m} |E\{\tilde{s}_i^4\} - 3[E\{\tilde{s}_i^2\}]^2| \quad (16)$$

This equation corresponds to 4^{th} order cumulant kurtosis.

3.1 Feature Extraction using ICA

In feature extraction which is based on Independent Component Analysis [5,11], one can consider an independent component s_i as the i^{th} feature of the recognized object represented by the observed pattern vector $\mathbf{x}$. The feature pattern can be formed from m independent components of the observed data pattern.

In order to form the ICA patterns, we propose the following procedure:

1. Extraction of n_f element feature patterns $\mathbf{x}_f$ from the recognition objects. Composing the original data set T_f containing N cases $\{\mathbf{x}_{f,i}^T, c_i\}$. The feature patterns are represented as matrix $\mathbf{X}_f$ and corresponding categorical classes (represented as column $\mathbf{c}$).

2. Heuristic reduction of feature patterns from the matrix $\mathbf{X}_f$ into n_{fr} element reduced feature patterns $\mathbf{x}_{fr}$ (with resulting patterns $\mathbf{X}_{fr}$). This step could be directly possible, for example, for features computed as singular values of image matrices.

3. Pattern formation through Independent Component Analysis of reduced feature patterns $\mathbf{x}_{fr}$ from the data set $\mathbf{X}_{fr}$.

   (a) Whitening of the data set X_{fr} including reduced feature patterns of dimensionality n_{fr} into n_{frw} element whitened patterns $\mathbf{x}_{rfw}$ (projected reduced feature patterns into n_{frw} principal directions).

   (b) Reduction of the whitened patterns $\mathbf{x}_{frw}$ into the first n_{frwr} element reduced whitened patterns $\mathbf{x}_{frwr}$ through projection of reduced feature patterns into first principal directions of data.

4. Computing the unmixing matrix $\mathbf{W}$ and computing reduced number n_{icar} of independent components for each pattern $\mathbf{x}_{frwr}$ obtained from whitening using independent component analysis

(projection patterns $\mathbf{x}_{frwr}$ into independent component space).

5. Forming n_{icar} element reduced ICA patterns $\mathbf{x}_{icar}$ from corresponding independent components of whitened patterns, with the resulting data set $\mathbf{X}_{icar}$. Forming a data set T_{icar} containing pattern matrix $\mathbf{X}_{icar}$ and original class column $\mathbf{c}$

6. Providing rough sets based processing of the set T_{icar} containing ICA patterns $\mathbf{x}_{icar}$. Discretizing pattern elements and finding relative reducts from set T_{icar}. Choosing one relative reduct. Selecting the elements of patterns $\mathbf{x}_{icar}$ corresponding to the chosen reduct and forming the final pattern $\mathbf{x}_{fin}$. Composing the final data set $T_{final,d}$ containing discrete final patterns $\mathbf{x}_{fin,d}$ and class column. Composing the real-valued data set T_{fin} from the set T_{icar} choosing elements of real-valued pattern using the selected relative reduct.

3.2 ICA and Rough Sets

ICA does not guarantee that the selected first independent components, as a feature vector, will be the most relevant for classification. As opposed to PCA, ICA does not provide an intrinsic order for the representation features of a recognized object (for example an image). Thus, one cannot reduce an ICA pattern just by removing its trailing elements (which is possible for PCA patterns). Selecting features from independent components is possible through application of rough sets theory [5,10,11]. Specifically defined through rough sets computation, a reduct can be used for selecting some of the independent component-based attributes constituting a reduct. These reduct-based, independent component-based features will describe all concepts in a data set. The rough set method is used for finding the reducts from the discretized reduced ICA patterns. The final pattern is formed from reduced ICA patterns based on the selected reduct.

The results of the method of feature extraction/selection discussed in this paper depend on data set type and designer decisions: a) selection of dimension for the independent component space, b) discretization method applied, and (c) the selection of a reduct, etc.

4 Mammogram Recognition using ICA and Rough sets

We applied Independent Component Analysis and rough sets for the mammogram recognition system. We used the following methods for feature extraction and pattern formation from gray scale mammogram subimages. In order to provide a numerical experiment for normal, benign and malignant cases, we created 20×20, 40×40 and 60×60 pixel subimages from the original 1024×1024 pixel images and stored them into each class directory (i.e. Normal, Benign and Malignant). Then, from the subimages, we constructed a row vector by concatenating each row of an image. It means that the row vector can be considered as a raw pattern for a subimage. As a result, the final pattern set of those raw patterns for each class category was constructed. As an example using the 60×60 pixel subimages, a final pattern set can be represented with a 330×3601 matrix or a 300×3601 matrix for MIAS MiniMammographic database and MiniDB, respectively. Then we applied Independent Component Analysis (including the whitening phase) to the patterns in order to transform feature patterns into independent component space and to reduce pattern dimensionality.

Additionally, the rough sets method was applied for feature reduction and selection. The original feature patterns with 60×60 pixel subimages were projected into a 20-element independent component space (including a 40-element whitening phase). Then, we found a reduct from the ICA patterns which had 15 elements by using the rough sets method, and the final pattern was formed. Finally, the rough sets rule-based classifier provided 82.22% of classification accuracy for the test set.

5 Principal Component Analysis for Feature Extraction and Reduction

We have applied Principal Component Analysis (PCA) for the orthonormal projection (and reduction) of $N \times n$ data pattern matrix $\mathbf{X}$. Let us assume that the eigenvalues of the covariance matrix $\mathbf{R}_x$ of the data matrix $\mathbf{X}$ are arranged in the decreasing order $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0$ (with $\lambda_1 = \lambda_{max}$), with the corresponding orthonormal eigenvectors $\mathbf{e}^1, \mathbf{e}^2, \cdots, \mathbf{e}^n$ arranged correspondingly in decreasing order. The optimal linear transformation

$$\mathbf{y} = \hat{\mathbf{W}}\mathbf{x} \tag{17}$$

is provided using the $m \times n$ optimal Karhunen-Loéve transformation matrix $\hat{\mathbf{W}}$

$$\hat{\mathbf{W}} = \left[\mathbf{e}^1, \mathbf{e}^2, \cdots, \mathbf{e}^m\right]^T \tag{18}$$

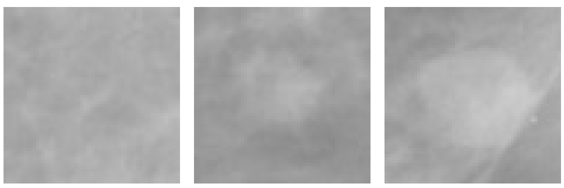
composed with m rows being the first m orthonormal eigenvectors of the original data covariance matrix $\mathbf{R}_x$. The optimal matrix $\hat{\mathbf{W}}$ transforms the original *n-dimensional* patterns $\mathbf{x}$ into *m-dimensional* ($m \leq n$) feature patterns $\mathbf{y}$

$$\mathbf{Y} = (\hat{\mathbf{W}}\mathbf{X}^T)^T = \mathbf{X}\hat{\mathbf{W}}^T \tag{19}$$

The projected, on the reduced *-dimensional* principal component space, patterns $\mathbf{y}$ can be considered as resulting decorrelated patterns, which elements represent principal components of the data matrix $\mathbf{X}$.

5.1 Numerical Experiment

We used two different mammogram databases for the numerical experiments: MIAS MiniMammographic database[3] and MiniDB. The MiniDB is a mammogram database built with 300 images from the DDSM(Digital Database for Screening Mammography)[4]. We have downloaded 300 original mammogram images containing: 100 normal, 100 benign, and 100 malignant cases. Each of the 300 original images has a file size exceeding 10M. Therefore, we have resized them to a file size less than 1M with 1024×1024 pixels. In addition to resizing the images, we have created an information file called "MiniDBInfo" which contains all the information about the database such as image name, class, cancer type, and cancer position and radius. In summary, the MIAS MiniMammographic database consists of 330 mammogram images: 207 normal cases, 69 benign cases and 54 malignant cases. The MiniDB is comprised of 300 images: 100 normal cases, 100 benign cases and 100 malignant cases. Figure 1 shows three examples of 60×60 pixel subimages of a normal, a benign, and a malignant case.



(a) Normal (b) Benign (c) Malignant

Figure 1: Database

As feature extraction techniques, ICA and PCA were applied in order to extract independent components and principal components from each raw pattern set. In the case of a subwindow with dimension 60×60 pixels, the final pattern sets can be represented with a 330×21 ICA pattern matrix and a 330×21 PCA pattern matrix using the MIAS MiniMammographic database. Next, we converted real data into discrete date because the rough sets rule-based classifier which we used for classification can handle only discrete data. The discrete pattern set has been used to form our final data pattern with a 90% training set and 10% testing set. Therefore, we had 297 training cases and 33 testing cases for the MIAS MiniMammographic database, and we had 270 training cases and 30 testing cases for the MiniDB. Next in order to reduce the dimensionality of the patterns, we applied rough sets for feature reduction/selection. As a result, the size of the 60×60 ICA patterns from the MIAS database were reduced from 330×21 to 330×16. In our next step, we applied the rough sets rule-based classifier to the reduced patterns for recognition of the three classes of images. Table 1, 2, 3 and 4 show the results of classification using the rough sets rule-based classifier.

Table 1: Classification of ICA patterns in MIAS database

Size (pixels)	Normal (%)	Benign (%)	Malignant (%)	Total (%)
20×20	100	57.14	40	65.71
40×40	95.23	42.85	40	59.36
60×60	95.23	71.42	80	82.22

Table 2: Classification of PCA patterns in MIAS database

Size (pixels)	Normal (%)	Benign (%)	Malignant (%)	Total (%)
20×20	100	85.71	60	81.90
40×40	100	85.71	80	88.57
60×60	90.47	57.14	60	69.27

Table 3: Classification of ICA patterns in MiniDB database

Size (pixels)	Normal (%)	Benign (%)	Malignant (%)	Total (%)
20×20	90	90	90	90
40×40	100	100	100	100
60×60	100	100	100	100

Table 4: Classification of PCA patterns in MiniDB database

Size (pixels)	Normal (%)	Benign (%)	Malignant (%)	Total (%)
20×20	90	100	100	96.66
40×40	100	90	80	90
60×60	100	100	90	96.66

We also applied two neural network classifiers: the error back propagation classifier and the Learning Vector Quantization classifier to the ICA extracted patterns. As shown in Table 5, the error back propagation classifier yields 51.5% of accuracy for the testing set from the

MIAS data, and the Learning Vector Quantization classifier provides 63.6% of accuracy for testing set from the same data.

Table 5: Results of the other classifiers.

DB	Class-ifier	Normal (%)	Benign (%)	Malign. (%)	Total (%)
MIAS	BP	71.42	28.57	0	51.51
MIAS	LVQ	100	0	0	63.63
MiniDB	BP	40	50	30	40
MiniDB	LVQ	70	30	30	43.33

6 Conclusion

We studied several hybrid methods for feature extraction/reduction, feature selection, and classifier design for breast cancer recognition in mammograms. The methods included ICA, PCA, and rough sets. Three classifiers were designed and tested: a rough sets rule-based classifier, an error back propagation neural network, and a Learning Vector Quantization neural network. We provided comparative study for two different data sets of mammograms. In both data sets, the rough sets rule-based classifier performed with a significantly better level of accuracy than the other classifiers. Therefore, the use of ICA or PCA as a feature extraction technique in combination with rough sets for feature selection and rule-based classification is an improved solution for mammogram recognition in the detection of breast cancer.

7 Reference

[1] A. Cichocki, R.E. Bogner, L. Moszczynski, Pope, A.: Modified Herault-Jutten algorithms for blind separation of sources. *Digital Signal Processing*, 7, (1997) 80 - 93

[2] Cios, Pedrycz, K., Swiniarski, R.: Data Mining Methods for Knowledge Discovery. Kluwer Acad. Publ., Boston (1998)

[3] Hadjiiski, L .M., & H.-P. Chan, M. N, Gurcan, B, Sahiner, N. Petrick, M. A. Helvie, M. A. Roubidoux. 2002. Computer-aided characterization of malignant and benign microcalcication clusters based on the analysis of temporal change of mammographic features. Proceedings of SPIE International Conference on Image Processing,0 25-28 February 2002, Vol. 4684, 4684-79.

[4] Heath, M., Bowyer, K., Kopans, D., Moore, R., and Kegelmeyer, P. (2000). The digital database for screening mammography. Int. Workshop on Dig. Mammography.

[5] Hyvrinen, A., Oja, E.: Independent component analysis by general nonlinear Hebbian-like learning rules. *Signal Processing*, 64(3) (1998) 301-313

[6] Pawlak, Z.: Rough sets, Theoretical aspects of reasoning about data, Kluwer, Dordrecht (1991)

[7] Skowron, A. 1990. The rough sets theory and evidence theory. Fundamenta Informaticae vol. 13, 245-262.

[8] Skowron, A. & R. Swiniarski. 2001. Rough Sets in Pattern Recognition. In S. K. Pal, and Pal. Pattern Recognition. From Classical to Modern Approaches. World Scientific, Singapore,. 385-426.

[9] Swiniarski, R.: An Application of Rough Sets and Haar Wavelets to Face Recognition. In Proceedings of RSCTC'2000 In the Proceedings of The International Workshop on Rough Sets and Current Trends in Computing. Banff Canada, October 16-19 (2000) 523-530

[10] Swiniarski, R. Skowron, A.: Rough Sets Methods in Feature Selection and Recognition. *Pattern Recognition Letters.* 24(6). (2003) 833-849.

[11] R. W. Swiniarski, A. Skowron. "Independent Component Analysis and Rough Sets in Face Recognition". In J. Grzymala-Busse, B. Kostek, R. Swiniarski, M. Szczuka (Editors-in Chief of a special I volume) Transaction on Rough Sets I. In (Edito rs- in-Chief J. Peters, A. Skowron), Lecture Notes in Computer Sciences on Rough Sets, 3100, Springer, Berlin, New York, 2004, pp. 392-404.

[12] R. Swiniarski, Thuc Luu, A. Swiniarska, Hendrik Tanto. "Data Mining and On-line Recognition of Mammographic Images Based on Haar Wav elets, Principal Component Analysis and Rough Sets Methods". International SPIE Symposium Medical Imaging 2001. San Diego, Febr. 17-23, 2001