

Advanced texture analysis techniques for building textural models, with applications in the study of the pathology evolution stages, based on ultrasound images

Delia Mitrea¹, Sergiu Nedevschi¹, Monica Platon-Lupsor², Radu Badea²

¹ *Computer Science Department, Technical University of Cluj-Napoca, Baritiu Street, No. 26-28, Cluj-Napoca, Romania*

{Delia.Mitrea, Sergiu.Nedevschi}@cs.utcluj.ro

² *Department of Medical Imaging, I. Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania*

{rbadea}@umfcluj.ro

Abstract. This chapter describes specific, texture-based methods for the detection, characterization and recognition of some severe affections and of their evolution phases, using only information from ultrasound images. We perform the recognition of the considered affections in supervised manner, and we also discover the disease evolution phases in unsupervised manner. In both cases, the imagistic textural model is defined, consisting of: the relevant features for the characterization of the disease, respectively of its evolution phase; the specific values of the relevant textural features: arithmetic mean, standard deviation, probability distribution. Advanced texture analysis techniques, consisting of textural microstructure co-occurrence matrices based on Laws' features, are involved in this process. At the end, the imagistic textural model is validated through powerful, supervised classifiers, the resulting accuracy being around 90%.

1 Introduction

The research described in this paper was performed in the context of a post-doctoral scholarship, part of the "PARTING – Inter-university partnership for excellence in engineering" project, co-financed from the European Social Funds through the Sectorial Operational Program for Human Resource Development (2007-2013). Our entire research is performed in the context of the collaboration with the medical imaging specialists from Iuliu Hatieganu University of Medicine and Pharmacy of Cluj-Napoca, the purpose being to develop non-invasive, computerized methods for the prevention and automatic diagnosis of some severe abdominal diseases such as cirrhosis or malignant tumors.

Cancer is a lethal diseases, one of the most frequent affections nowadays. The abdominal malignant affections, such as the hepatocellular carcinoma [1], [2] and the colorectal tumors [3], occur very often. Also, other diseases that precede these affections, such as cirrhosis, present a major importance. The detection of the cirrhosis severity grades and of the HCC development phases is also important, the objective being the early detection and prevention of the malignancy evolution. The golden standard for the diagnosis of these affections is the biopsy, but it is invasive, dangerous. In order to detect the severity grades of the diseases various clinical methods exist, but none of them is scientifically objective [1]. We aim to develop non-invasive, computerized methods for the automatic diagnosis and grading of the above mentioned, severe abdominal affections, based only on the information from ultrasound images. The ultrasonography is a safe method for patient examination, easy to apply, inexpensive and repeatable. Other equivalent methods, such as the computer tomography (CT), the magnetic resonance imaging (MRI), the endoscopy or the contrast enhanced ultrasonography are irradiating or expensive [2], [3]. We defined the imagistic textural model of the considered abdominal affections and of their evolution phases, consisting of the most relevant textural features for their characterization and of the specific values of the relevant textural features (arithmetic mean, standard deviation, probability distribution). The texture is an important property of the body tissues, able to provide subtle information upon the pathology and to overpass the sensitivity of the medical experts' eyes. Besides the classical methods for texture analysis, we developed and experimented, in the above mentioned context, superior order generalized co-occurrence matrices [4], [5], [6]. In this work, we synthesize the role that the co-occurrence matrices of textural microstructures, computed after applying the Laws convolution filters [7], [8], [9], have in the previously described applicative domain. The texture analysis methods, in combination with classifiers, have been widely implemented in the nowadays research, in order to perform the recognition, in supervised or unsupervised manner, of some severe diseases, based on the information obtained from medical images [10], [11], [12], [13]. The generalized co-occurrence matrices, were also defined and experimented in [14], [15], in the form of the texton co-occurrence matrices [15] and LBP co-occurrence matrices of order two [14]. However, no significant study exists, which performs a systematic study of the considered affections based on ultrasound images and defines a unified, imagistic textural model of these affections. Also, the textural microstructure co-occurrence matrices, based on Laws' features, were not approached before.

2. Medical considerations. The aspect of the considered affections in ultrasound images

Cirrhosis is a lethal diffuse liver diseases which constitutes the first phase of liver parenchyma restructuring. The main characteristics of this affection are the association of fibrosis, regeneration nodules and necrosis, leading to important changes in the structure of the hepatic parenchyma. Cirrhosis constitutes the end stage of multiple liver diseases and sometimes it leads to death by itself. The regeneration nodules, specific to cirrhosis, can transform into dysplastic nodules, evolving towards hepato-

cellular carcinoma (HCC). Within ultrasound images, cirrhosis is featured mainly by the tissue homogeneity decrease, due to the fact that the nodules continuously appear and evolve. These nodules could be hypo-echogenic, or even unapparent. Other changes that could occur are: increased volume, in the case of toxic cirrhosis; decreased volume, in the case of viral cirrhosis; shape and contour modification; vessel structure modifications. As cirrhosis is a very serious disease, monitoring its severity presents a great importance. The Metavir classification system assimilates cirrhosis with the fourth evolution phase of liver fibrosis, a preliminary medical assessment revealing the next evolution stages for cirrhosis: Stage A – incipient, "compensated"; Stage B – intermediate phase, beginning to decompensate; Stage C – final stage, "decompensated". However, no objective study exists, for establishing the cirrhosis severity grades, using the information derived from ultrasound images [1]. *The hepatocellular carcinoma (HCC)* is the most frequent malignant liver tumor, occurring in 70% of the liver cancer cases. It evolves from cirrhosis, after a phase of liver parenchyma restructuring, at the end of which sometimes dysplastic nodules result, which evolve into HCC [1], [2].

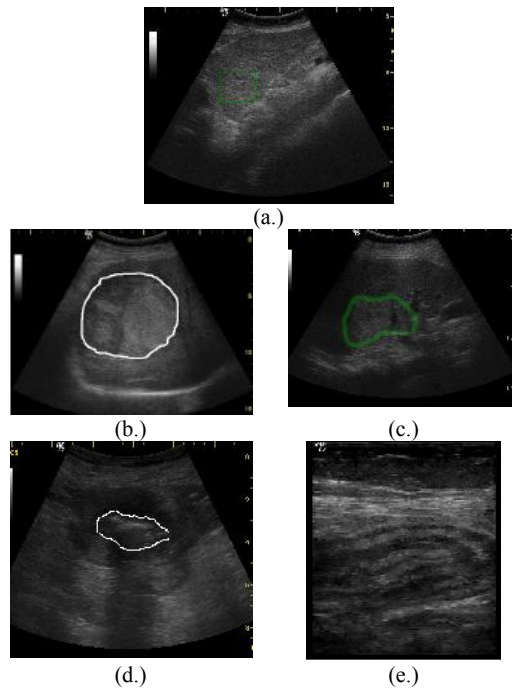


Fig. 1. Ultrasound images illustrating the considered affections: (a) – cirrhosis; (b) – HCC (marked contour); (c) – Hemangioma (marked contour); (d.)- Colorectal tumor (marked contour); (e.) - IBD example (Crohn disease)

The colorectal tumors are abdominal malignant structures that frequently occur in the population in the modern era. Concerning the aspect of the considered tumors in

ultrasound images, they usually have a hyperechogenic and heterogeneous aspect, due to the presence of multiple tissue types (fibrosis, necrosis, fat cells and active growth tissues), as well as to the complex structure of vessels. However, it is difficult to distinguish these tumors within the ultrasound images, as the HCC tissue resembles the surrounding cirrhotic parenchyma and also the hemangioma benign tumor, which is a collection of blood vessels and usually has a homogeneous, hyperechogenic visual appearance. The colorectal tumors are similar in aspect with the Inflammatory Bowel Diseases (IBD) as both classes of disease lead to bowel layer inflammation and sometimes to the loss of the stratification [3]. Some eloquent examples of B-mode ultrasound images representing these affections are provided in *Fig. 1*.

3. The state of the art

The texture-based methods were widely used in nowadays' research, in combination with classifiers, in order to perform the automatic recognition of various diseases and of their evolution phases from medical images [10], [11], [12], [13]. Concerning the supervised recognition of the malignant tumors, the run-length matrix parameters, in combination with the Haralick features derived from the GLCM matrix, were employed together with ANN classifiers, Support Vector Machines and Fisher Linear Discriminants, for the automatic diagnosis of the liver lesions from ultrasound images [10]. Methods like the Wavelet was used in conjunction with Artificial Neural Networks (ANN) for the recognition of the liver tumors from ultrasound images [11]. The fractal-based methods were adopted for the recognition of the salivary gland tumors in the research described in [12]. In [13] the authors implemented the Gray Level Co-occurrence Matrix (GLCM) in combination with morphologic features referring to shape and orientation, in order to separate the malignant and benign tissues for patients affected by colorectal cancer. These features were derived from biopsy slides. Concerning the automatic recognition of the disease evolution phases in a supervised manner, an important approach regarding the detection of the diffuse liver diseases and of their severity grades from ultrasound images is described in [16]. For this purpose, the authors employed a hierarchical tree, structured in the following manner: at the first level, the differentiation between the normal and the unhealthy liver was performed, using textural features such as the GLCM mean and energy; at the second level, the difference between steatosis and cirrhosis was highlighted, by using features such as the GLCM mean and variance; at the last level, the authors differentiated between multiple severity grades of steatosis and cirrhosis, using GLCM features such as the variance, entropy, sum entropy, difference entropy. For supervised classification, a Multilayer Perceptron (MLP) classifier was employed. An important work is presented in [17], where the authors determined the evolution stages of HCC from histological images, using newly defined textural features, based on multifractal analysis. A supervised, bag-of-feature classifier was used in order to identify one of the five evolution stages of HCC (four taken from the *Edmondson and Steiner* grading system, and also an additional stage, preceding the malignity). The final resulted accuracy was 95%. Another significant approach refers to the combination between the supervised and unsupervised classification methods, the purpose being to detect the HCC tumor

in incipient phase, based on histological features [18]. Regarding the unsupervised classification of the disease evolution stages, the authors performed in [19] the detection of the cirrhosis severity grades based on Magnetic Resonance Images (MRI). For this purpose, they employed textural features derived using the finite differences of the image intensity function and also an improved k-means clustering method. The authors determined, in this manner, the existence of four cirrhosis severity grades, the achieved performance, assessed through the area under the ROC curve, being 0.704. Concerning the activity of grading other affections, in unsupervised manner, the most important approach involves detecting the evolution stages of the glioma tumors, by combining hierarchical classification methods and dimensionality reduction techniques [20]. The method of Laplacian Eigenmaps, applied before some unsupervised classification methods such as k-means clustering or hierarchical agglomerative clustering, provided better results than the Independent Component Analysis (ICA) method applied in the same manner. The finally obtained classification accuracy was situated above 91%. Concerning the generalized co-occurrence matrices, they were recently employed, in the form of the Local Binary Pattern (LBP) Co-occurrence Matrix [14], respectively of the texton and texture orientation co-occurrence matrix [15]. These methods were experimented on the *Vistex* and *Gabor* standard textures and provided a recognition accuracy situated between 89% - 98.8%. From the above described approaches, it results that it doesn't exist any relevant work referring to a co-occurrence matrix based on extended textural microstructures determined after applying the Laws' convolution filters. Also, there doesn't exist any significant study that defines an imagistic textural model of the previously mentioned diseases and of their evolution phases, putting into evidence the relevant textural features and their specific values. We performed these studies in our research, the corresponding techniques being detailed below.

4. The Adopted Solutions

4.1. The imagistic textural model

4.1.1. The definition of the imagistic textural model

The imagistic textural model of the considered affections and of their evolution phases consists of: the complete set of the relevant textural features which characterize each disease or disease evolution stage and thus can distinguish the corresponding class from the similar existing classes; the specific values of the relevant textural features: arithmetic mean, standard deviation, probability distribution [4].

In order to define the imagistic textural model, we start from an initial set of potentially relevant textural features, as indicated in (1):

$$F = \{f_i\}_{i=1,\dots,n} \quad (1)$$

The set of the relevant textural features, F_R , constitute a sub-set of F , as indicated in (2), derived by using the most appropriate dimensionality reduction methods, the final purpose being to obtain the optimum classification performance, aiming first an increased accuracy. The dimensionality reduction techniques consist mainly of feature selection methods, which separate the relevant textural features from the non-relevant ones, while preserving the original features unchanged [21].

$$F_R \subseteq F \quad (2)$$

The final form of the imagistic textural model is depicted in (3). Thus, the textural model (TM), consists of the vectors V_{f_r} composed by the specific values associated to the relevant textural features (the value of the relevance index, the arithmetic mean, the standard deviation and the probability distribution).

$$TM = \{V_{f_r} \mid V_{f_r} = [\text{Relevance}, \text{Mean}, \text{St_dev}, \text{Probability_distribution}]\} \quad (3)$$

4.1.2. The phases due in order to build the imagistic textural model

In order to build the imagistic textural model, a *preliminary phase* of image gathering for training set building is due firstly. Then, *the image analysis phase* is performed, assuming the application of the texture analysis methods in order to obtain the potentially relevant textural features. Both classical and newly defined, advanced texture analysis methods are considered in this phase. The following classical texture analysis techniques were taken into account: first order statistics of the gray levels (arithmetic mean, maximum and minimum value); second order statistics such as the autocorrelation index, the Gray Level Co-occurrence Matrix (GLCM) and the associated Haralick features, edge based statistics and gradient features, the Hurst fractal index, the density and frequency of the textural microstructures obtained after applying the Laws' convolution filters, the Shannon entropy computed after applying the Wavelet transform recursively, twice [22]. Generalized, superior order co-occurrence matrices were also developed and analyzed in our previous research, in the form of the superior order GLCM [4], of the third order Edge Orientation Co-occurrence Matrix (EOCM) [4], respectively of the multiresolution versions of these matrices [5], [6]. The newly defined texture analysis methods, described in this work, were based on textural microstructure co-occurrence matrices, of order two and three and also of their multiresolution versions, as detailed in the next sections. *The learning phase* follows next. In the case of the unsupervised approach, automatic class discovery is performed during this phase, through clustering techniques, as described within the next sub-sections. In both unsupervised and supervised approaches, the feature selec-

tion methods are also applied during the learning phase. *The validation phase* is employed at the end in order to assess the classification performance due to the previously built imagistic textural model. During this phase, the relevant textural features are provided as inputs to powerful, supervised classifiers and meta-classifiers.

4.2. Advanced texture analysis techniques employed during the image analysis phase

4.2.1. The superior order, generalized co-occurrence matrices

The second order, generalized co-occurrence matrix was firstly defined by Davis [23]. In our work, we defined the superior order, generalized co-occurrence matrix. The definition for the elements of this matrix is provided in (4):

$$\begin{aligned}
C_D(\epsilon_1, \epsilon_2, \epsilon_3, \dots, \epsilon_n) &= \#\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)\} : \\
A(x_1, y_1) &= \epsilon_1, A(x_2, y_2) = \epsilon_2, \dots, A(x_n, y_n) = \epsilon_n, \\
|x_2 - x_1| &= |\vec{d}x_1|, |x_3 - x_2| = |\vec{d}x_2|, \dots, |x_n - x_1| = |\vec{d}x_{n-1}|, \\
|y_2 - y_1| &= |\vec{d}y_1|, |y_3 - y_2| = |\vec{d}y_2|, \dots, |y_n - y_1| = |\vec{d}y_{n-1}|, \\
\text{sgn}((x_2 - x_1)(y_2 - y_1)) &= \text{sgn}(\vec{d}x_1 \cdot \vec{d}y_1), \dots \\
, \text{sgn}((x_n - x_1)(y_n - y_1)) &= \text{sgn}(\vec{d}x_{n-1} \cdot \vec{d}y_{n-1})
\end{aligned} \tag{4}$$

In the formula (4), $\#S$ represents the size of the set S . Each element of this matrix is equal with the number of n -tuples of pixels, having the spatial coordinates (x_i, y_i) , respectively the value v_i for the attribute A . The spatial relation between the pixels (x_i, y_i) was defined by the set of the displacement vectors, depicted in (5):

$$\vec{d} = ((\vec{d}x_1, \vec{d}y_1), (\vec{d}x_2, \vec{d}y_2), \dots, (\vec{d}x_{n-1}, \vec{d}y_{n-1})) \tag{5}$$

The attribute A represents the image features taken into consideration in order to define the co-occurrence matrix in a particular case: the gray level of the pixel in the case of the superior order GLCM matrix [7], the edge orientation in the case of the Edge Orientation Co-occurrence Matrix (EOCM) [7], the value obtained after applying a certain Laws' convolution filter in the case of the Textural Microstructure Co-occurrence matrix, and the label associated to the complex and complex extended textural microstructures, in the case of the Complex Textural Microstructure Co-occurrence Matrix (CTMCM), respectively of the Complex Extended Textural Microstructure Co-occurrence Matrix (CETMCM). The superior order GLCM and EOCM matrices were defined and detailed in [4].

The TMCM, CTMCM and CETMCM matrices will be described in the next paragraphs. The second order co-occurrence matrices were computed for the following directions of the displacement vectors: 0° , 90° , 180° , and 270° . For the third order co-occurrence matrices, the following direction combinations were taken into account: $(0^\circ, 180^\circ)$, $(90^\circ, 270^\circ)$, $(45^\circ, 225^\circ)$, $(135^\circ, 315^\circ)$ in this case the pixels being collinear.

ar, the current pixel being situated in the central position; respectively $(0^\circ, 90^\circ)$, $(90^\circ, 180^\circ)$, $(180^\circ, 270^\circ)$, $(0^\circ, 270^\circ)$, $(45^\circ, 135^\circ)$, $(135^\circ, 225^\circ)$, $(225^\circ, 315^\circ)$, $(45^\circ, 315^\circ)$, in this case the pixels forming a right angle triangle, the current pixel being situated in the position of the right angle vertex. The direction combinations considered in the case of the third order matrices are illustrated in Fig. 2. For these matrices, the extended Haralick features were computed as described in [4], for each direction or direction combination, the final resulting values being averaged. Also, features referring to the shape of the 2D and 3D histograms associated to the co-occurrence matrices, were derived. These features were: cluster shade, standing for histogram skewness; cluster prominence, standing to histogram kurtosis; the maximum area corresponding to the intersection of the histogram with a horizontal plane in the bi-dimensional case and its extension for superior order cases. These features were computed as described in [9].

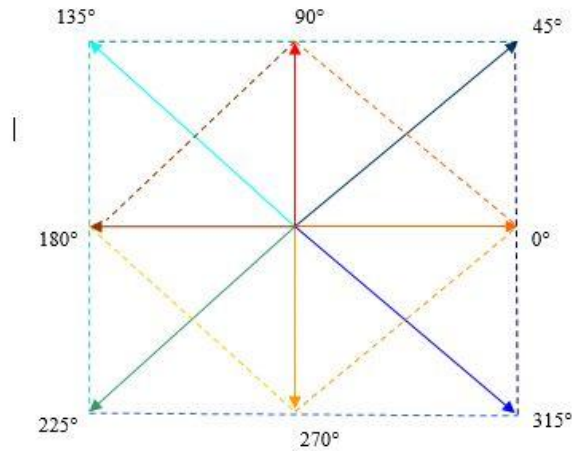


Fig. 2 The combination of the displacement vector directions in the case of the third order generalized co-occurrence matrices

The Textural Microstructure Co-occurrence Matrix (TMCM)

In the case of the Textural Microstructure Co-occurrence Matrix (TMCM), the attribute A represents the value obtained after applying a Laws' convolution filter, which puts into evidence a certain textural microstructure: level, edge, spot, wave or ripple [24]. In our work, the bi-dimensional, 5×5 versions of these convolution filters (L_5L_5 , E_5E_5 , S_5S_5 , W_5W_5 , R_5R_5), were considered. Also, the S_5R_5 and R_5S_5 combined convolution filters were taken into account, as they provided good results in our former experiments [4], [5], [6].

The Complex Textural Microstructure Co-occurrence Matrix (CTMCM)

Concerning the Complex Textural Microstructure Co-occurrence Matrix (CTMCM), the values of the attributes A are the labels, assigned to each pixel, associated to the clusters formed by the vectors of textural microstructures, which could also be assimilated with the concept of textons [15]. The feature vector associated to each

pixel consisted of the values resulted after the application of the considered Laws' convolution filters in the neighborhood of this pixel. The clusters resulted after the application of an improved k-means clustering method, performed in the following manner: (1) We started from a minimum number of cluster centers ($k=50$); (2) We applied the classical k-means clustering algorithm and increased the number of centers by splitting the corresponding centers in two other centers, if the standard deviation of the items within the corresponding class (cluster) was greater than $\frac{3}{4}$ of the average standard deviation of all the existing classes. The values of the first new center were determined as being $\frac{1}{2}$ of the values associated to the old center; the second new center was $\frac{3}{4}$ of the old center. (3) All the labels of the pixels from the ROI were re-assigned after splitting the old centers. The condition for the algorithm to finish was the convergence, the maximum number of centers being also established to 200.

The Complex Extended Textural Microstructure Co-occurrence Matrix (CETMCM)

The Complex Extended Textural Microstructure Co-occurrence Matrix (CETMCM) was computed in the same manner as the CTMCM matrix, but the initial feature vector associated to each pixel resulted as the union between the results of the Laws' convolution filters and the results of representative edge detection convolution filters. In order to perform edge point detection, we considered the Sobel filters for identifying horizontal and vertical edges, the Kirsch Compass filters for finding edges with different orientations, as well as the Laplacian convolution filter, in order to compute the second order derivative of the image intensity function [22].

4.2.2. The multi-resolution, generalized co-occurrence matrices

The multi-resolution, generalized co-occurrence matrices are computed at two resolution levels, after applying the Wavelet transform, in the following manner: (1) First, the Haar Wavelet transform was applied on the original image; (2) then, the desired co-occurrence matrix was computed on each resulted component: low-low (smoothed), low-high (vertical edges), high-low (horizontal edges), and high-high (diagonal edges); (3) The Haar Wavelet transform was applied on each resulted component and the co-occurrence matrices of the desired type were computed on the new resulted components. In our research, the multiresolution GLCM and EOCM were previously approached [5], [6]. In this work, we highlighted the role of the multiresolution, third order CTMCM concerning the improvement of the abdominal tumor recognition accuracy. In order to determine the multiresolution CTMCM (MCTMCM), the following steps were performed (1) The improved k-means clustering method was applied on the original image, considering that each pixel had associated a feature vector formed by the values resulted of the application of the considered Laws' convolution filters. A clustered (texton) image. (2) The Haar Wavelet transform was applied on the clustered image recursively, twice. (3) The co-occurrence matrix, CTMCM, was computed on each component.

4.3. The techniques applied during the learning phase

4.3.1. Relevant feature selection in the case of the supervised classification

Specific feature selection methods of both filter and wrapper type [21] were experimented in our research in order to separate the relevant textural features from the non-relevant ones. The methods from the category of *filters* provided the best results. Thus, the techniques of Correlation-based Feature Selection (CFS), Consistency based Feature Subset Evaluation, Information Gain Attribute Evaluation and Gain Ratio Attribute Evaluation provided the best results in most of the cases. The first two techniques, which performed the assessment of feature subsets, were usually combined with *genetic search*, while the last two techniques, which performed the evaluation of individual attributes, were combined with the *Ranker* method. [21] The *final relevance score* for a certain attribute was computed as the arithmetic mean between the individual relevance indexes provided by each on the employed method.

4.3.2. The case of the unsupervised classification

Class discovery within the data derived from ultrasound images

In order to perform automatic discovery of the classes within the data derived from ultrasound images, aiming to detect the evolution stages of the considered diseases, we applied clustering (grouping) methods. The following specific techniques were implemented individually: Expectation Maximization (EM), k-means clustering, X-means clustering, Particle Swarm Optimization (PSO) in conjunction with k-means clustering. The final results of these methods were combined, for establishing the number of classes. In order to assess the performance in each case, the metrics specific to each method were taken into account, and also other parameters were considered, such as the average difference between the cluster proportions and the number of insignificant clusters (clusters of small size, containing less than 10% of the data). We provide below a brief description of each adopted technique.

The *Expectation Maximization (EM)* technique is a well-known, powerful method that iteratively estimates the desired parameters, by maximizing the log-likelihood value. The likelihood of the model is determined by formula (10), while the log-likelihood is the natural logarithm of the likelihood [25]:

$$Likelihood = \prod_{i=1}^n \sum_{j=1}^m p_j P_r[x_i | j] \quad (6)$$

Within the formula (6), n stands for the number of instances, m represents the number of clusters, p_j stands for the cluster priors and $Pr[x_i|j]$ is the conditional probability for the instance x_i to be belong to the cluster j . The parameters estimated in our research were the number of clusters and the sample distributions within the clusters. In the case of the *EM* method, the performance was estimated using a score defined as a weighted mean that considered the log-likelihood, the number of small (insignificant) clusters, respectively the maximum difference between the cluster proportions:

$Score = 0.5 * \log_likelihood + 0.3 * (1 - no_insignifiant_clust) + 0.2 * (1 - max_dif_clust_prop)$. All the terms were normalized between 0 and 1.

The *k-means clustering technique* [25] is often employed for unsupervised classification, due to its reduced computational complexity and accelerated convergence. The *k-means clustering algorithm* consists of the following steps: (1) randomly initialize the cluster centers $\mu_1, \mu_2, \dots, \mu_c$, where c is the number of clusters, considering the available data, containing n instances; (2) assign each instance to the cluster that corresponds to the nearest center, μ_i ; (3) compute again the cluster centers, by employing the arithmetic mean for this purpose; (4) if the position of the cluster centers has remarkably changed, go to step 2, otherwise, return $\mu_1, \mu_2, \dots, \mu_c$. Thus, given a set of observations $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, each observation being a d -dimensional vector, the *k-means clustering algorithm* partitions the n observations into k sub-sets ($k \leq n$) $\mathcal{S} = \{S_1, S_2, \dots, S_k\}$, using the algorithm described above. In our study, the performance of this method was measured by a weighted mean that considers the within-cluster sum of squared errors (WCSS) of the resulted model, besides the maximum difference between the cluster proportions and the number of insignificant clusters, all the corresponding values being normalized in the interval $[0,1]$: $Index = 0.5 * WCSS + 0.2 * max_dif_clust_prop + 0.3 * no_insignifiant_clust$.

The *X-means clustering technique* represents an improved version of the *k-means clustering method*. The classical *k-means clustering technique* [25] has some drawbacks, the most important being the execution speed, and also the fact that a fixed value k has to be provided a-priori. The *X-means clustering technique* expects a maximum and a minimum value for the k parameter and it consists of the following steps: (1) Run conventional *k-means clustering* to convergence, for a fixed value k ; (2) Decide whether new cluster centroids should appear, by splitting the old centroids into two; (3) If $k > k_{max}$, then stop and report the best model found by the algorithm, identified using the Bayesian Information Criterion – BIC [25]. The BIC criterion is used both in order to decide which centroids have to be split, and also for identifying the best resulted model. The final algorithm performance is estimated by the distortion measure, expressing the average squared distance from the points to their centroids, for the best resulted model (Pelleg, 2000). In the case of the *X-means clustering technique*, the *distortion* measure was estimated and an index was computed as follows: $Index = 0.5 * distortion + 0.3 * no_insignifiant_clust + 0.2 * max_dif_clus_prop$. All the terms were normalized between 0 and 1. The smallest index value indicated the best solution in this situation.

Being inspired from the bird flocks behavior, the *Particle Swarm Optimization (PSO)* method aims to optimize the solution of a problem by simulating the movement of a particle swarm and by determining then the best position for each particle (Das, 2008). Each particle has assigned a position and a velocity. The velocity (speed) is influenced by a cognitive component, referring to the distance from the personal best position, as well as by a social component, referring to the distance from the best global position. The best particle positions are determined through an evaluation function, which is defined according to the specific of each problem [26]. In our case of unsupervised classification through clustering (grouping), a particle is represented by a certain cluster configuration, respectively by the way the cluster labels are assigned to the input data, for a given number of clusters. We combined the PSO technique with *k-means clustering*. Thus, the initial configuration of the swarm resulted after the

application of the k-means method upon the initial data. We defined the evaluation function using the same metric employed in the case of the k-means clustering.

Relevant feature selection in this case

The methods for relevant feature selection, in the case of unsupervised classification, aim to provide best separation among the resulted clusters. The overlapping area between two neighboring clusters must be as small as possible. For each textural feature f , a relevance score was defined, as described in (7):

$$Relevance(f) = \sum_{i,j} (1 - Overlapping_reg_size_{i,j}) \quad (7)$$

In (7), i and j represent neighboring clusters. The relevance of each feature f depends on the sum of the overlapping region sizes that exist between each pair of neighboring clusters. The overlapping region size was determined according to (8); we assumed a Gaussian distribution for the textural features:

$$Overlap_reg_size = \begin{cases} S, & \text{if } S > 0 \text{ and } \mu_i \neq \mu_j \text{ (overlap.clust.)} \\ 0, & \text{if } S \leq 0 \text{ and } \mu_i \neq \mu_j \text{ (non-overlap.clust.)} \\ 2 \cdot \frac{min_f}{Max_f - Min_f}, & \text{if } \mu_i = \mu_j \end{cases} \quad (8)$$

In (8)

$$S = \frac{(\mu_{min_f} + 2\sigma_{min_f}) - (\mu_{max_f} - 2\sigma_{max_f})}{Max_f - Min_f} \quad (9)$$

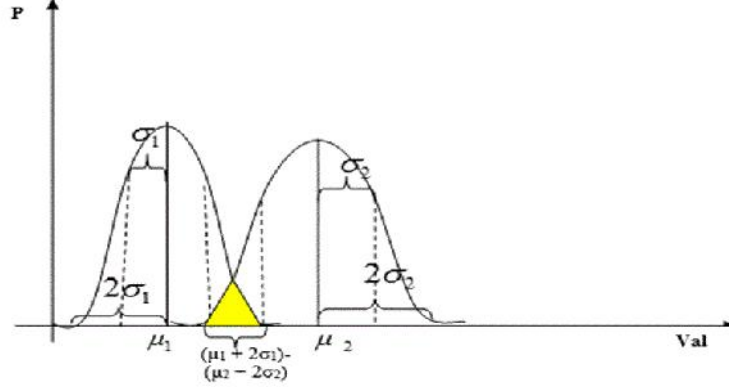


Fig. 3. The overlapping region size between two neighboring clusters ($S > 0$)

In (9), μ_{min_f} represents the minimum value for the arithmetic mean of f , considering the clusters i and j ; σ_{min_f} is the standard deviation of the textural feature f for the cluster where μ has minimum value, μ_{max_f} stands for the maximum arithmetic mean of the feature f , considering the clusters i and j , and σ_{max_f} is the standard devi-

ation of f within the cluster where μ is maximum [9]. If $S \leq 0$, no overlapping region exists between the two clusters. If the two centers of the clusters i and j are the same with respect to f (the arithmetic means, μ_i and μ_j are equal), then the overlapping region size will be the double of the minimum standard deviation. The graphical representation for the classical case, $S > 0$, is provided in *Fig. 3*.

4.3.3. Computing the specific values of the relevant textural features

The arithmetic mean and the standard deviation of each relevant textural feature were computed for each class. In the case of disease evolution phase analysis, the graphical representation of the relevant textural feature arithmetic mean as a function of the evolution stage was performed. In order to determine the probability distribution for each relevant textural feature, with respect to each considered class, the Bayesian Belief Networks technique [25] was adopted. The method of Bayesian Belief Networks detects influences among the features, by generating a dependency network, represented as a directed, acyclic graph (DAG). In this graph, the nodes represent the features, while the edges stand for the causal influences between these features, having associated the values of the corresponding conditional probabilities. Each node X in this graph has a set of parents, P and a set of children, C . The probabilities of the nodes are computed using a complex inference mechanism. Within a Bayesian Belief Network, each node has assigned a probability distribution table, indicating the specific intervals of values for that node, given the values of its parents. Considering the textural features that influence the class, the technique of Bayesian Belief Networks provides the most probable variation intervals for them.

4.4. Validation of the imagistic textural model

In order to evaluate the imagistic textural model, the values of the relevant textural features were provided at the supervised classifier inputs and the resulting classification performance was assessed. The following supervised classifiers, well known for their efficiency, were employed for this purpose: the Multilayer Perceptron (MLP), the Support Vector Machines (SVM), the Random Forest (RF) method, the C4.5 method of decision trees and also the AdaBoost meta-classifier combined with the C4.5 classification technique, the latter being validated in the literature as a powerful classifier combination [25]. The case of binary classification was taken into account for the supervised approach. In the case of the unsupervised approach, when multiple disease evolution phases were analyzed at once, the multiclass classification methods were taken into account. Both the natural extensions of the usual classification methods to the case of multiple classes, as well as specific combination strategies of the binary classification results were employed for this purpose [25]. The Exhaustive Correction Code strategy [27] was considered in the latter case. For classification performance assessment, the following metrics were adopted: the recognition rate (accuracy), sensitivity (TP rate), specificity (TN rate) and the area under ROC (AUC). The method of Self Organizing Maps (SOM) [28] was also employed, for performing data representation, in order to visualize the existing classes (clusters).

5 Experiments and discussions

5.1. The experimental dataset and the experimental environment

The experimental dataset consisted of 300 cases of HCC, 100 cases of hemangioma, 65 cases of colorectal tumors, 65 cases of IBD, respectively 75 cases of cirrhosis. Three ultrasound images, corresponding to various orientation of the transducer, were included for each case (patient). All the images were acquired by a Logiq 7 ultrasound machine using the same settings: frequency of 5.5 MHz, gain of 78, depth of 16 cm. The following pairs of classes were considered for the case of supervised classification: HCC/cirrhotic parenchyma on which HCC had evolved; HCC/hemangioma; colorectal tumors/IBD. The classes of a certain pair contained the same number of items. On each image, regions of interest (ROI) were selected on the desired tissue. The textural features were computed on each ROI, using our own modules, implemented in Visual C++. The textural features were derived after the application of a median filter for noise reduction, independently on orientation, illumination and region of interest size. An item in the data set consisted of the textural feature values computed on the corresponding ROI, having also associated the class parameter in the case of supervised classification. *In the case of supervised classification*, the methods for relevant feature selection, respectively the classifiers, were implemented using the Weka 3.6 library [27]. The *CfsSubsetEval*, the Weka 3.6 correspondent of the CFS technique and the *ConsistencySubsetEval* method of the same library, both described in *Section 4*, were employed in conjunction with Genetic Search. Also, the *InfoGainAttributeEval* and the *GainRatioAttributeEval* techniques of Weka 3.6, in combination with the Ranker method were implemented, according to the previous descriptions. In order to compute the probability distributions for the relevant textural features, the Bayesian Belief Networks (*BayesNet*) method of Weka 3.6 was adopted, with Bayesian Model Averaging (BMA) estimation and K_2 search. Regarding the classification methods, the *John's Platt Sequential Minimal Optimization Algorithm (SMO)* of Weka 3.6, was applied for implementing the SVM method, with polynomial kernels of second and third degree, which provided the best results in our former experiments; the input data was normalized in this case. For the classifier of Multilayer Perceptron (MLP), the corresponding method of Weka 3.6 (*Multi-layerPerceptron*) was considered, containing, within the single hidden layer, a number of nodes equal with $a = (\text{number_of_features} + \text{number_of_classes})/2$, the learning rate being tuned to 0.2, respectively the momentum α being 0.8. The *RandomForest* classification method of Weka 3.6, standing for the RF classifier, with 10 trees, was also adopted. The *AdaBoostM1* meta-classifier of Weka was employed as well, using the *J48* method (equivalent of the C4.5 technique) as a basic learner. For classification performance assessment, the strategy of cross-validation with 5 folds was applied. For the case of multiclass classification, the *MultiClass* meta-classifier of Weka 3.6 was employed, together with the Exhaustive Correction Code strategy. *In the case of unsupervised classification*, the feature selection method, *described in sub-section 4.3.2*, and the Particle Swarm Optimization (PSO) technique combined with k-means clus-

tering were implemented in Matlab, using a specific framework in the latter situation [29]. The other unsupervised classification methods (classical k-means clustering, X-means clustering, Expectation Maximization) were employed using the Weka 3.6 library [27]. The SOM method was implemented in the Matlab environment using a specific library [28].

5.2. The role of the textural microstructure co-occurrence matrices in the supervised classification of the abdominal tumors

5.2.1. The Textural Microstructure Co-occurrence Matrix (TMCM) in the context of abdominal tumor classification

The following pairs of classes were considered in this situation: HCC/cirrhotic parenchyma on which HCC had evolved, colorectal tumors/IBD. The second and third order TMCM matrices were determined and the resulted Haralick features were combined with the previously computed textural features. Concerning the most *relevant textural features* obtained in the case of the comparison between HCC and the cirrhotic parenchyma on which HCC had evolved, we noticed the importance of the second order TMCM based features, as well as of the third order TMCM based features, computed after the application of the Laws' ripple and combined filters, respectively after the application of the Laws' filter for spot detection, in the case of the third order TMCM entropy. All the detected relevant textural features denote the complex, inhomogeneous, chaotic character of the HCC tumor, as well as differences in granularity between HCC and the cirrhotic parenchyma, through the second and third order GLCM and TMCM correlation. In the second case, when the colorectal tumors and the IBD classes were compared, the following relevant textural features resulted: $\{Autocorrelation_index, EOCM3_Homogeneity, EOCM3_Contrast, GLCM3_Homogeneity, GLCM3_Correlation, TMCM3_Homogeneity_ripples, TMCM3_S5R5_Entropy, TMCM_Spots_Entropy, TMCM_R5S5_Contrast, TMCM3_Spots_Energy, TMCM3_S5R5_Correlation\}$. We can notice the presence of the second and third order TMCM features: homogeneity, contrast, entropy, energy and correlation, referring to all the considered microstructures (spots, ripples and combined). These textural features denote the heterogeneous, complex, chaotic structure of the colorectal tumors, concerning the gray levels, the edge orientation variability, respectively the textural microstructures. At the end, the classification accuracy due to the set formed only by the old textural features was compared, in both cases, with the accuracy provided by the set formed by the old textural features and by the newly defined, TMCM features. The experiments were performed after feature selection (FS). The TMCM textural features led, in most of the considered situations, to a classification accuracy improvement, in comparison with the old textural features. In the case of HCC/cirrhotic parenchyma comparison, the maximum accuracy, of 82.14%, the maximum sensitivity, of 82.12% and the maximum specificity, of 82.35%, resulted for the combination between AdaBoost and the J48 basic classifier, while the maximum AUC, of 85.35%, resulted for the J48 classifier. Concerning the comparison between the colorectal tumors and the IBD class, the maximum accuracy, of 94.78%, the maximum sensitivi-

ty, of 95.90% and the maximum specificity, of 96.99%, resulted in the case of the combination between the AdaBoost meta-classifier and the J48 technique, while the maximum AUC, of 96%, resulted for the RF classifier.

5.2.2. The role of the CTMCM matrix in the recognition of the abdominal tumors

In the case of the differentiation between HCC and the cirrhotic parenchyma on which HCC had evolved, regarding the set of the relevant textural features, selected by the CFS method, combined with genetic search, we noticed the presence of the *contrast* and *variance* derived from the third order CTMCM matrix, denoting the complex character of the malignant tumor tissue. The comparison between the classification accuracy obtained when using only the old textural feature set, respectively that obtained by using the combination between the old textural features and the CTMCM features, in the case of *the differentiation between HCC and the cirrhotic parenchyma on which HCC had evolved*, was also performed. An increase in accuracy was noticed in most of the cases. The maximum recognition rate, of 84.09%, was obtained for the AdaBoost meta-classifier in combination with the J48 method of decision trees, being higher than that obtained when employing TCMC matrix. In the case of *the differentiation between HCC and the benign liver tumors (hemangioma)*, for the same textural features, we noticed an increase in accuracy, due to the newly defined textural features, for all the classifiers. The comparison was performed after the selection of the relevant textural features. The maximum recognition rate, of 98.141%, was obtained in the case of the SVM classifier. In the case of the *differentiation between the colorectal tumors and the IBD class*, concerning the set of the relevant textural features, we noticed the *homogeneity* and *contrast* features based on the second order CTMCM matrix, respectively the *contrast* derived from the third order CTMCM matrix. These features denoted the complex, inhomogeneous character of the colorectal tumors. After providing the set of the relevant textural features at the classifier inputs, the following values were obtained for the classification performance parameters: the maximum accuracy, of 98.33%, resulted in the case of the SVM classifier, all the values of the classification performance parameters overpassing those achieved in the case of the TCMC matrix. Concerning the comparison between the classification accuracy due to the relevant CTMCM features combined with the old textural features and that due only to the set of the old relevant textural features, in the case of differentiation between the colorectal tumors and IBD, we noticed, always, an improvement due to the newly defined textural features [7]. These results can be visualized in Fig. 6, Fig.7 and Fig.8.

5.2.3. The role of the MCTMCM in the classification of the abdominal tumors

In the case of the differentiation between HCC and the cirrhotic parenchyma on which HCC had evolved, within the set of the relevant textural features that perform the differentiation between these classes, we noticed the presence of the following textural features, derived from the third order MCTMCM matrix: the *contrast*, computed at the second level, on the components resulted after the second application of the Wavelet transform, on the image of the vertical edges; the *homogeneity*, computed on the fourth component, after the second application of the Wavelet transform, on the

image of the diagonal edges. These features expressed the difference in homogeneity between the cirrhotic parenchyma and the tissue affected by malignity, as well as the complex character of the HCC tissue. The improvement of the classification accuracy, due to the relevant textural features derived from the MCTMCM matrix, in comparison with that due to the old features, was also analyzed. An obvious classification accuracy increase was noticed in the cases of the MLP and RF classifiers, a slight increase in recognition rate was remarked for the AdaBoost meta-classifier employed in conjunction with the J48 technique, respectively a decrease of the recognition rate was noticed concerning the SVM classifier. The maximum recognition rate obtained in this case was 83.84%, for the AdaBoost meta-classifier combined with the J48 technique. This result was inferior to that obtained in the case of the CTMCM matrix. Thus, the textural information lost in relevance as the result of compressing and filtering, due to the application of the clustering method followed by the Wavelet transform. *Regarding the comparison between the colorectal tumors and the IBD class, the following third order MCTMCM features resulted as being relevant: the contrast, the variance, the correlation determined on the sub-image of vertical edges, the homogeneity determined at the second level on the sub-image of the horizontal edges, as well as on that of the diagonal edges.* These features highlighted the complex, heterogeneous character of the malignant tumor tissue, as well as differences in granularity between the malignant tumor tissue and the other class (through the third order MCTMCM correlation).

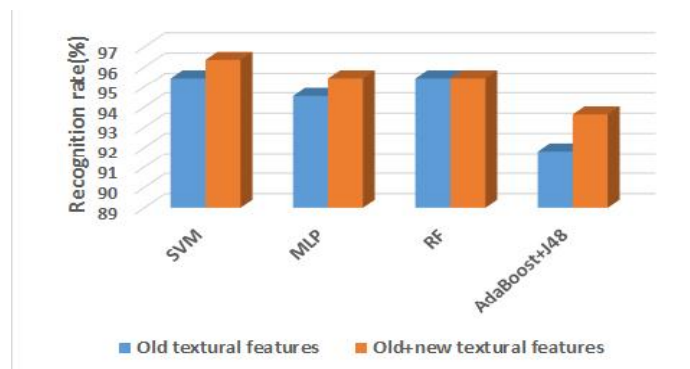


Fig 4. The increase in accuracy due to the MCTMCM features, in the case of colorectal tumors/IBD differentiation

Fig.4 illustrates the comparison between the recognition rates resulted before and after taking into account the MCTMCM features. An obvious accuracy increase, due to the newly defined, MCTMCM features, resulted in most of the cases. The maximum recognition rate resulted in this case was 96.36%, corresponding to the SVM classifier. This result was inferior to that obtained when considering the CTMCM textural features, for the same pair of classes [7].

5.2.4. The role of the textural Complex Extended Textural Microstructure Co-occurrence Matrix (CETMCM) in the recognition of the abdominal tumors

In the case of the comparison between HCC and the cirrhotic parenchyma on which HCC had evolved, the ranking, according to the relevance parameter, of the most significant textural features which were part of the imagistic textural model is provided in Fig. 5. We notice the presence of the *contrast* derived from the third order CETMCM among the first three most relevant textural features.

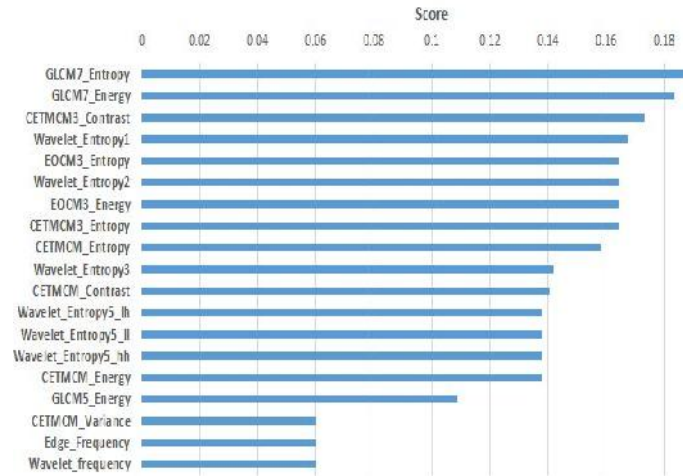


Fig. 5. The ranking of the relevant textural features in the case of HCC/cirrhotic parenchyma differentiation

The comparison of the classification accuracies resulted when considering only the old textural features, respectively after adding the newly defined textural features, is illustrated in Fig.6, Fig.7 and Fig.8, for all the considered class pairs. These figures also illustrate the classification accuracy obtained when considering the set formed by the old textural features and the CTMCM features. It results that the both the CTMCM and the CETMCM features led to an obvious classification accuracy increase in comparison with the old textural feature set, while the CETMCM feature led to a slight improvement in comparison with the set containing the CTMCM features. The comparison between the classification accuracies obtained when using the old textural features, the combination between the old textural features and the CTMCM features, respectively the combination between the old textural features and the CETMCM features, assessed in the case of the HCC/cirrhotic parenchyma differentiation, is depicted in Fig.6. The maximum recognition rate, of 84.33%, resulted in the case of the AdaBoost meta-classifier combined with the J48 method, when taking into account the CETMCM features together with the old textural features [8].

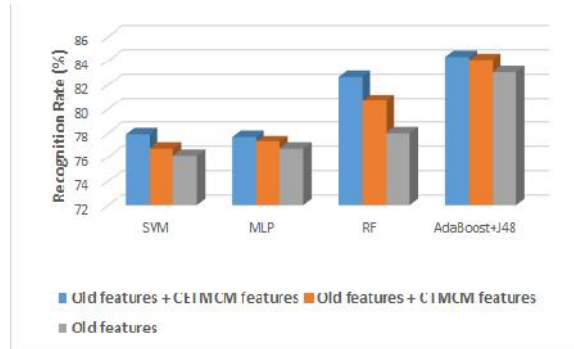


Fig. 6. The classification accuracy improvement, due to the CTMCM and CETMCM textural features, in the case of HCC/cirrhotic parenchyma comparison

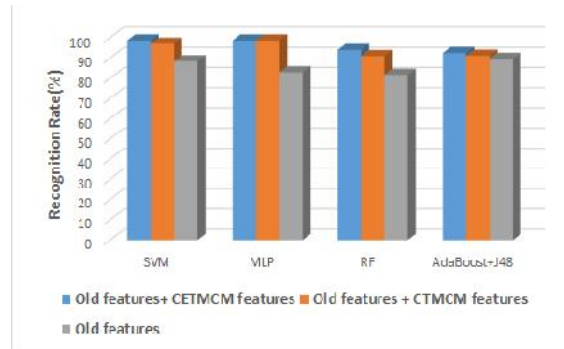


Fig. 7. The classification accuracy improvement, due to the CTMCM and CETMCM textural features, in the case of HCC/hemangioma

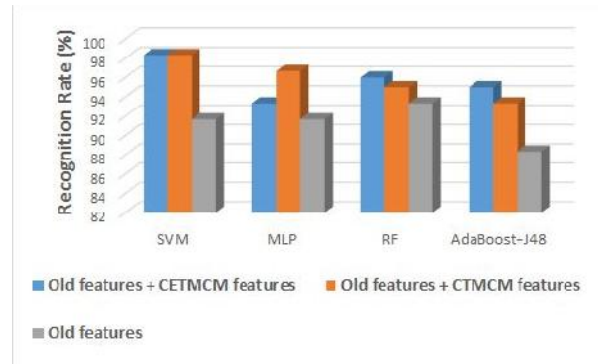


Fig. 8. The classification accuracy improvement, due to the CTMCM and CETMCM textural features, in the case colorectal tumors/IBD comparison

As it results from *Fig. 7*, in the case of HCC/hemangioma differentiation, the maximum recognition rate, of 98.8%, resulted in the case of the SVM classifier, when taking into account the newly defined, CETMCM textural features.

In the case of the comparison between the colorectal tumors and the IBD class, illustrated in *Fig. 8*, the maximum recognition rate, of 98.33%, resulted in the case of the SVM classifier, for the feature set containing the CETMCM textural features [8].

5.3. The role of the textural microstructure co-occurrence matrices in the unsupervised detection of the abdominal disease evolution phases

5.3.1. The role of the textural microstructure co-occurrence matrices in the discovery of the cirrhosis severity grades

First, the textural features derived from the *TMCM matrix* were taken into account. After the application of the EM and X-means clustering techniques and performing the evaluation as shown in sub-section 4.3.2, four clusters were discovered within the data. These clusters corresponded to the cirrhosis severity grades: an incipient evolution phase, an advanced evolution phase and two intermediate stages. In this case, the most relevant parameters were the *frequency of the textural microstructures detected by using the Laws' convolution filters*, together with the *Haralick features* derived from the second and third order TMCM matrix. These features highlighted the importance of the first, second and third order statistics based on textural microstructures in characterizing the evolution of cirrhosis.

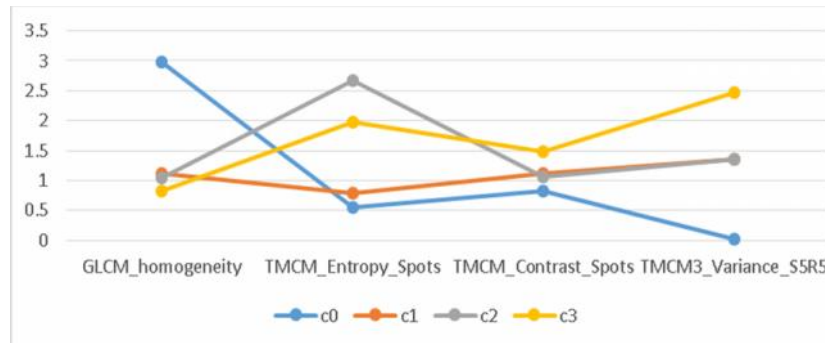


Fig. 9. The specific values of the most relevant textural features, for each cluster (cirrhosis severity grade)

The arithmetic means of the most relevant textural features, in correlation with the evolution phase, are depicted in *Fig. 9*. As it results from *Fig. 9*, the *cluster c0* corresponded, most likely, to the incipient phase of cirrhosis, as the GLCM homogeneity had maximum value, while the entropy, contrast and variance derived from the second and third order TMCM matrices had minimum values. The *cluster c3* corresponded, most likely, to the most advanced stage of this affection (when the tumors begin to

appear), as the GLCM homogeneity had a minimum value, the entropy computed from the TCMC matrix based on spot microstructures had high values, while the contrast and variance based on the second and third order TCMC had maximum values, denoting an increased structural complexity, due to the advanced restructuring process.

The clusters c_1 and c_2 corresponded to intermediate evolution stages. In the case of cluster c_2 , the entropy derived from the TCMC matrix based on spot microstructures had an increased value, so, class c_2 probably corresponded to a more advanced evolution stage, preceding the final stage.

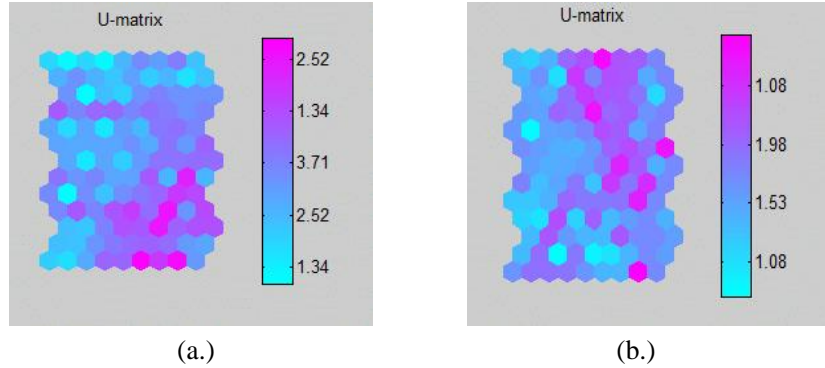


Fig.10. The U-Matrices corresponding to the Self Organizing Maps (SOM) obtained before (a.) and after (b.) feature selection.

The SOM technique was also applied for class visualization within the data, before and after relevant feature selection. According to *Fig. 10*, a better cluster differentiation can be observed after the feature selection operation. At the end, for the validation of the results, supervised classification methods were applied, the values of the relevant textural features being given at their inputs. The class labels of the instances, as resulted after the application of the EM method, were also provided. The most increased accuracy, of 93.75%, resulted in the case of the MLP method. We also compared the classification accuracy results obtained before and after adding the TCMC features to the old feature set. An accuracy increase, due to the newly defined textural features, resulted in all the considered cases.

Then, the *CTMCM* and *CETMCM* textural features were added to the old textural feature set, as well, and the number of clusters within the data was assessed again. Four clusters, corresponding to four cirrhosis evolution phases, resulted also in this case. The newly considered textural features were part of the relevant textural feature set and implicitly of the imagistic textural model of the cirrhosis evolution stages. The supervised classifiers were applied again in order to assess the resulted model. The maximum recognition rate, of 97.87%, resulted for the J48 classifier, when considering the *CETMCM* features, overpassing the maximum accuracy, of 97.5%, resulted in the case when the *CTMCM* features were taken into account. We can notice that these values of the classification accuracy also overpassed the recognition rate resulted when considering the TCMC features.

5.3.2. The role of the textural microstructure co-occurrence matrices in the discovery of the HCC evolution phases

In order to perform the discovery of the HCC evolution phases, when taking into account *the CTMCM, respectively the CETMCM* textural features, the results of the following methods were combined, in order to determine the number of clusters: *EM, X-means clustering and PSO combined with k-means clustering*.

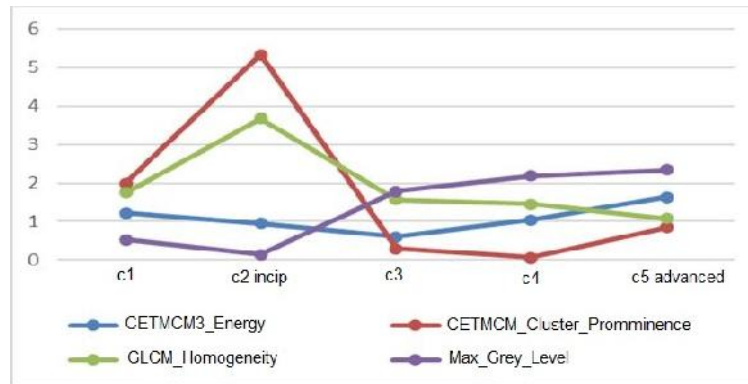


Fig. 11. The mean values of the relevant textural features for the HCC evolution stages

Thus, five clusters, corresponding to five HCC evolution stages (an incipient phase, an advanced phase and three intermediate phases), resulted after the data analysis. The CTMCM and the CETMCM textural features were part of the relevant feature set derived using the technique described in sub-section 4.3.2. The graphical representation for the values of the arithmetic means of some of the most relevant textural features, which corresponded to the cluster centers, is provided in *Fig. 11*. According to the a-priori knowledge, which reveals a decrease in homogeneity, respectively an increase in echogenicity during the evolution of HCC, the second cluster (c_2) corresponded to the incipient phase, having the most decreased value for the gray levels, respectively the maximum GLCM homogeneity, while cluster c_5 corresponded to the most advanced evolution phase for the HCC tumor, presenting the most increased value of the gray levels, respectively the minimum value of the GLCM homogeneity. The other clusters, c_1 , c_3 and c_4 corresponded to intermediate evolution phases: c_1 was closer to the incipient evolution phase, while c_3 and c_4 were closer to the advanced evolution phases. The arithmetic mean values of other features, such as the energy derived from the third order CETMCM and the cluster prominence derived from the second order CETMCM, were also analyzed. We notice, from *Fig. 11*, that the third order CETMCM energy had low values for the incipient phases and higher values for more advanced evolution phases, denoting an increase in the density of the complex extended textural microstructures during the HCC evolution. The cluster prominence derived from the CETMCM matrix, corresponding to the 3D histogram kurtosis, had maximum values during the incipient evolution phase and small values during the advanced evolution phases of HCC. The probability distributions generated using the Bayesian Belief Networks confirmed these results.

After instance labeling using the PSO method combined with k-means clustering, the model assessment through supervised classification was performed. The maximum recognition rate, of 93.35%, together with the maximum sensitivity (average TP rate), of 93.14%, respectively the maximum value of the AUC, of 98.3%, were obtained in the case of the MLP classifier, when considering the set formed by the old textural features and the CETMCM features, for the hypothesis of five clusters within the data. The accuracy results obtained when taking into account the combination between the old textural features and the CETMCM features in the five clusters hypothesis were compared with those obtained for the same feature set in the six cluster hypothesis, and also with the results obtained when considering the feature set formed by the old textural features combined with the CTMCM features, in the five cluster hypothesis. These results are depicted in *Fig. 12*. As it results from *Fig.12*, the newly defined CTMCM and CETMCM textural features, in combination with the old textural features, always led to a classification accuracy increase in comparison with the formerly defined textural features. Also, in the case when taking into account the CETMCM features, the classification accuracy was superior, in most of the cases, to that obtained in the case when taking into account the CTMCM features. The accuracy results were also superior when considering five evolution phases, in comparison to those obtained when considering six stages of HCC evolution [9].

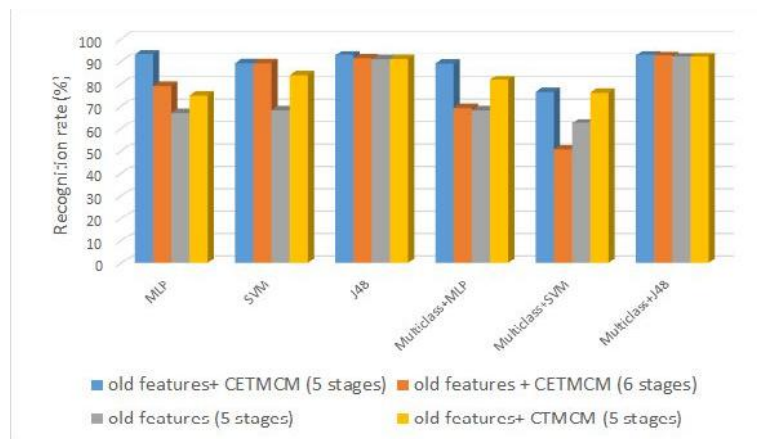


Fig. 12. Evaluation of the CTMCM and CETMCM textural features through supervised classification methods [9]

6. Conclusions and future work

The newly defined textural features, derived from the textural microstructure co-occurrence matrices, based on Laws' features, proved to be useful for the characterization and automatic diagnosis of the considered abdominal diseases and of their

evolution phases. The classification accuracy was about 85% in the case of HCC/cirrhotic parenchyma differentiation, respectively it was above 90% when distinguishing between HCC and the benign tumors, between the colorectal tumors and the IBD class, and also when performing supervised recognition of the cirrhosis and HCC evolution stages. Concerning our future research, we intend to validate our methods on larger datasets. We also aim to analyze the role of the textural microstructure co-occurrence matrices in the recognition of the real-world textures and to compare the efficiency of these techniques with co-occurrence matrices based on other features, such as the Local Binary Pattern (LBP).

Acknowledgement

The research presented in this chapter was supported by the Post-Doctoral Program POSDRU/159/1.5/S/137516, project co-funded from European Social Fund through the Human Resources Sectorial Operational Program 2007-2013.

References

1. American liver foundation (2016). Online: <http://www.liverfoundation.org/>
2. Sherman, M.: Approaches to the Diagnosis of Hepatocellular Carcinoma, *Current Gastroenterology Reports*, Vol. 7, No. 1. (2005) 11-18.
3. Martínez-Ares, D., Martín-Granizo Barrenechea, I, et al.: The value of abdominal ultrasound in the diagnosis of colon cancer, *Revista Espanola des Enfermedades Digestiva*, Vol. 27, No. 12 (2005) 877-886.
4. Mitrea, D., Mitrea, P., Nedevschi, S., et al.: Abdominal tumor characterization and recognition using superior order co-occurrence matrices, based on ultrasound image, *Computational and Mathematical Methods in Medicine* (2012). Online: <http://www.hindawi.com/journals/cmmm/2012/34813/>
5. Mitrea D., Nedevschi S., Badea R.: The role of the multiresolution textural features in improving the characterization and recognition of the liver tumors, based on ultrasound images, in *14th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC) proceedings* (2012) 192-199.
6. Mitrea D., Nedevschi S., Badea R.: Classification of the liver tumors using multiresolution, superior order EOCM textural features, *Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods – ICPRAM* (2014) 799-804.
7. Mitrea D., Nedevschi S., Badea R.: Colorectal cancer recognition from ultrasound images, using complex textural microstructure co-occurrence matrices, based on Laws' features, *Proceedings of the 38th International Conference on Telecommunications and Signal Processing – TSP* (2015) 458-462
8. Mitrea D., Nedevschi S., Abrudean M., Badea, R.: Abdominal tumor recognition from ultrasound images using Complex Extended Textural Microstructure Co-occurrence Matrices, *Automation, Computers and Applied Mathematics*, Vol. 23, No.1 (2014) 9-17.
9. Mitrea D., Nedevschi S., Badea R.: The Role of the Complex Extended Textural Microstructure Co-occurrence Matrix in the Unsupervised Detection of the HCC Evolution Phases

- es, based on Ultrasound Images, Proceedings of the International Conference on Pattern Recognition Applications and Methods – ICPRAM (2016) 698-705.
10. Sujana, H., Swarnamani, S.: Application of Artificial Neural Networks for the classification of liver lesions by texture parameters, *Ultrasound in Medicine & Biology*, Vol. 22, No. 9 (1996) 1177 –1181.
 11. Yoshida, H., Casalino, D.: Wavelet packet based texture analysis for differentiation between benign and malignant liver tumors in ultrasound images, *Physics in Med & Biol*, Vol. 48 (2003) 3735-3753.
 12. Chikui, T.: Sonographic texture characterization of salivary gland tumors by fractal analysis, *Ultrasound in Med & Biol*, Vol. 31, No. 10 (2005) 1297- 1304.
 13. Masood, K.: Co-occurrence and morphological analysis for colon tissue biopsy classification, Proceedings of the 4th International Workshop on Information Technology, Pakistan (2006) 211-216
 14. Sujatha, B., Kumar, V., Harini, P.: A new logical compact LBP Co-occurrence Matrix for Texture Analysis, *Int. J. of Scientific & Engineering Research*, Vol. 3, No. 2 (2012) 1-5.
 15. Sujatha, B., Sekhar Reddi, C.: Texture classification using texton co-occurrence matrix derived from texture orientation, *Int. J. Soft Computing&Eng.*, Vol. 2, No. 6 (2013) 18-23.
 16. Cavouras, D., Kandarakis, I., et al: Computer Image Analysis of Ultrasound Images for Discriminating and Grading Liver Parenchyma Disease Employing a Hierarchical Decision Tree Scheme and the Multilayer Perceptron Neural Network Classifier, *Medical Informatics Europe* (1997) 522-528.
 17. Atupelage, C., Nagahashi, H., Computational grading of hepatocellular carcinoma using multifractal feature description, *Computers in Med. Images and Graphics*,37, (2013) 61-71.
 18. Ciocchetta, F., et al.: Combining Supervised and Unsupervised Methods to Support Early Diagnosis of Hepatocellular Carcinoma, *Artificial Intelligence in Medicine*, 2780 (2000) 239-243.
 19. Lee, G., Kanematsu, M., Kato, H, Unsupervised classification of cirrhotic livers using MRI data, *Procs of SPIE*, 6915 (2008) 69514-1 – 69514-8.
 20. Yang, G., Raschke, F. et al.: Manifold Learning in MR spectroscopy using nonlinear dimensionality reduction and unsupervised clustering, *Magnetic Resonance in Medicine* (2014) Online: <http://onlinelibrary.wiley.com/doi/10.1002/mrm.25447/abstract>
 21. Hall, M.: Benchmarking attribute selection techniques for discrete class data mining”, *IEEE Trans on Knowledge and Data Eng.*, Vol. 15, No. 3 (2003) 1-16.
 22. Toennies, K.D.: Guide to medical image analysis. Methods and algorithms, Springer Verlag, London (2012)
 23. Davis, L.S. et al.: A Comparative Texture Classification Study Based on Generalized Cooccurrence Matrix, Proceedings of the 18-th IEEE Conference on Decision and Control (1979) 71-78.
 24. Laws, K.I.: Rapid texture identification, in *SPIE vol. 238 Image Processing for Missile Guidance* (1980) 76-380.
 25. Witten, I., Frank, E.: *Data Mining – Practical Machine Learning Tools and Techniques* (3rd edition), Morgan Kaufmann (2011)
 26. Das, S.: Particle Swarm Optimization and Differential Evolution Algorithms: Technical Analysis, Applications and Hybridization Perspectives. In *Studies in Comput Intelligence*, 116, (2008) 1–38.
 27. Weka3, Data Mining Software in Java, 2014, Online: <http://www.cs.waikato.ac.nz/ml/weka>
 28. SOM 2.0 toolbox for Matlab 2015. Online: <http://www.cis.hut.fi/somtoolbox/>
 29. Biswas, P.: Particle Swarm Optimization for Matlab, (2013), Online: <http://www.mathworks.com/matlabcentral/fileexchange/43541-particle-swarm-ptimization--pso>.

CONSENT TO PUBLISH and COPYRIGHT TRANSFER

For the mutual benefit and protection of Authors and Publishers, it is necessary that Authors provide formal written Consent to Publish and Transfer of Copyright before publication of the Book. The signed Consent ensures that the publisher has the Author's authorization to publish the Contribution.

Place/Date: **Rome, Italy; February, 2016**

Book Title: **European Project Space on Intelligent Technologies, Software engineering, Computer Vision, Graphics, Optics and Photonics**

Edited by: **Paolo Dell'Olmo, Marco Brambilla and Maria Raposo**

Publisher: **SCITEPRESS**

Title of the contribution: Advanced texture analysis techniques for building textural models, with applications in the study of the pathology evolution stages, based on ultrasound images

Author: (please print the **name** and **address** of corresponding **first** author): Delia Mitrea, Computer Science Department, Technical University of Cluj-Napoca, Baritiu Str., No. 26-28, 400027, Cluj-Napoca, Romania, e-mail: Delia.Mitrea@cs.utcluj.ro

It is herein agreed that:

The copyright to the contribution identified above is transferred from the Author to Science and Technology Publications, Lda (SCITEPRESS). The copyright transfer covers the exclusive, sole, permanent, world-wide, transferable, sub licensable and unlimited right to reproduce, publish, transmit and distribute the contribution, including reprints, translations, photographic reproductions, microform, electronic form (offline, online), or any other reproductions of similar nature, including publication in the aforementioned book or any other book. SCITEPRESS is also entitled to carry out editorial changes in the contribution with the sole purpose of enhancing the overall organization and form of the contribution. The Author retains the rights to publish the contribution in his/her own web site or in his/her employer's web site, as long as it is clearly stated that the contribution was presented at **European Project Space - Rome, 2016 (EPS)** and a link to the EPS web site is made available there.


The Author warrants that his/her contribution is original, except for such excerpts from copyrighted works as may be included with the permission of the copyright holder and author thereof, that it contains no libelous statements, and does not infringe on any copyright, trademark, patent, statutory right, or propriety right of others. The Author signs for and accepts responsibility for releasing this material on behalf of any and all co-authors.

In return for these rights:

The publisher agrees to have the identified contribution published, at its own cost and expense.

The undersigned hereby gives permission to SCITEPRESS to have the above contribution published.

Date: 24.06.2016

Author's Signature: 

Please return to: EPS Secretariat (Fax: +351 265 520 186; e-mail: eps@insticc.org) Address:
Avenida D. Manuel I, 27A - 2 Esquerdo 2910-595 Setúbal, Portugal