Building the imagistic textural model of the liver pathological stages for the early detection of hepatocellular carcinoma based on ultrasound images

Delia MITREA[1], Sergiu NEDEVSCHI[1], Monica LUPSOR[2], Radu BADEA[2]
[1] *Technical University of Cluj-Napoca, Computer-Science Department, Baritiu str, no. 26-28, Cluj-Napoca ROMANIA,*
[2] *3rd Medical Clinic, University of Medicine and Pharmacy, Department of Ultrasonography, Croitorilor str, no 19-21, Cluj-Napoca, ROMANIA*

Abstract: - The purpose of this study is to find the relevant features in order to detect the forerunner and early stages of the hepatocellular carcinoma (HCC); thus, we aim to characterize the evolution of the liver diseases towards hepatocellular carcinoma (HCC), by detecting the relevant textural features obtained from ultrasound images, which accurately surprise the changes of the liver tissue in the context of this evolution. For the computation of the textural features the following methods are chosen in order to emphasize the chaotic nature of the malignant tumor: first and second order statistics, edge-based statistics, fractal-based methods and multiresolution methods; specific methods for feature-selection are applied in order to determine the exhaustive set of independent and relevant features for each evolution phase. We focus on modeling the cirrhosis and hepatocellular carcinoma, but the differentiation of HCC and normal state is also taken into consideration. The final purpose is that of providing a reliable method for non-invasive characterization of the evolution towards HCC, in order to prevent this malignant tumor.

Key-words: - texture, relevant features, imagistic model, ultrasound images, hepatocellular carcinoma, computer-aided diagnosis

1 Introduction

Liver chronic diseases constitute an important public health issue, due to their frequency and to the danger of their evolution towards cancer. The evolution of diffuse liver diseases is variable, but it usually has generally long term. Whatever the nature of the liver aggression, it seems to follow a pattern characterized by the successive stages: inflammation, necrosis, fibrosis, regeneration (cirrhosis), dysplasia, and, at the end, the hepatocellular carcinoma. The hepatocellular carcinoma (HCC) is the most frequent malignant liver tumor (representing 75% of the liver cancer cases), besides hepatoblastoma (7%), cholangiocarcinoma and cystadenocarcinoma (6%). The visual aspect of the liver tissue, during the above-mentioned pathological stages, evolves from normal to inhomogeneous, due to the cirrhotic nodules, then the nodules increase giving birth to the hepatocellular carcinoma. The fundamental visual properties of the malignant liver tumors, and also of the hepatocellular carcinoma, which can be noticed by human eye from ultrasound images, are: the irregular-shaped, often vague contours, the complex structure of vessels and the heterogeneity of the tissue. However, the human observations are not enough in order to give a reliable diagnosis and the biopsy is an invasive method, dangerous for the patient. Thus, a more subtle analysis is due, and we aim to perform this by using computerized methods in order to extract the necessary information from ultrasound images. Texture is a very important visual feature in this context, as it provides a lot of information concerning the pathological state of the tissue. Although a commonly accepted definition does not exists, texture describes the regular arrangement of the grey levels in the region of interest, giving also information about the frequency and regularity of the local features, about the fundamental microstructures, being also able to provide multi-resolution information, appropriate for an accurate, subtle characterization of the tissue illness. Texture-based methods, in combination with classification methods have been widely used for the automatic diagnosis of various kinds of tumors. The Grey Levels Co-Occurrence Matrix, the Wavelet and Gabor transforms and fractal-based methods are only a few eloquent examples used in the characterization of the breast tumors, of the prostate adenocarcinoma, of the salivary gland tumors, of liver lesions, respectively of malignant and benign tumors. Also, the same kinds of methods (first and second order statistics, fractal-based methods, and Wavelet Transform) were successfully applied for diffuse liver diseases differentiation based on ultrasound images. However, a systematic study concerning the relevant features and their specific values for the characterization of HCC and its forerunner

stages based only on information extracted from ultrasound images, is not done yet. We aim to do this in our work, which will consist in modeling the pathological stages preceding HCC through textural parameters. The following phases will be due: (1) the *image analysis phase*, implying the computation of the textural features and (2) the selection of the exhaustive and non-redundant set of relevant features. First and second order statistics, edge-based statistics, fractal-based methods and the Wavelet Transform will be used for feature computation. For feature selection, specific methods and various classifiers will be compared and the most efficient one will be chosen. Then, a *learning step is due*, in which we analyze the specific values of the textural features for each class, we identify the relevant textural features and we obtain some specific features like the mean, the variance, the maximum and the minimum value and their probability distribution. At the end, there is a validation phase, in which we evaluate the values obtained for the specific features. The possibility of automatic diagnosis is also tested, by applying the methods of Bayesian Belief Networks and the Multilayer Perceptron.

2 State of the Art

For diffuse liver diseases differentiation (steatosis, hepatitis and cirrhosis from normal liver), as well as for the automatic detection of the malignant tumors, the most frequently used methods are the Grey Levels Co-occurrence Matrix (GLCM) and its second order parameters, the Run-Length Matrix parameters, fractal-based methods, and the Wavelet Transform, in combination with classifiers like k-nn, Artificial Neural Networks, Fisher Linear Discriminants, Support Vector Machines. C. Wu et al. proposed in [1] the implementation of a new fractal-based textural feature called Multiresolution Fractal Feature, MF, for the differentiation of the normal liver from the cirrhotic liver based on ultrasound images. This method proved to be efficient, the obtained accuracy being about 90%. In the case of the recognition of malignant tumors, the fractal-based methods and the multiresolution methods are much more emphasized, because of the chaotic, complex structure of tumor tissue. In [2] the authors compute the first order statistics, the Grey Level Co-occurrence Matrix second order parameters and the Run-Length Matrix parameters, which are used in association with an Artificial Neural Networks based classifier, as well as with a classifier based on Linear Discriminants, for the classification of the liver lesions. Fractal-based methods are used in [3] in order to distinguish the salivary gland tumors from ultrasound images. The Wavelet transform was also implemented in order to analyze the values of the textural parameters at multiple resolutions, in order to distinguish malignant and benign liver tumors in ultrasound images [4]. From the multiresolution methods class, the Gabor transform was also implemented in [5] and provided good results in combination with the GLCM second order parameters and the Bayesian classifier for prostate malignant tumors recognition. For feature relevance study, in the case of malignant tumors differentiation, methods like Decision Trees [6], Support Vector Machines [6], Independent Component Analysis [6], were used, but a systematic study of the most relevant textural features, extracted from ultrasound images, that characterize the evolution towards HCC, of their correlations with the visual features and of their specific values, is not done.

3 Visual Aspect of the Liver Tissue in each Pathological Stage

The studies demonstrate that the most relevant oncogenic agent for HCC development is chronic viral infection with the HBV or HCV, the areas with higher prevalence of viral infection presenting the highest HCC rates [7]. The next evolution phase, preceding HCC, is cirrhosis. Cirrhosis is a diffuse liver disease characterized through the association of fibrosis, regeneration nodules and hepatocitar necrosis, with hepatic structure alteration. The tissue homogeneity decreases due to the nodules. The nodules have low echogenicity, or they could be unapparent.

Other changes of the liver could be: increased volume (in the case of the toxic cirrhosis) or decreased volume (viral cirrhosis); shape and contour modification, due to the nodules; vessels modifications. HCC evolves from cirrhosis, after a restructuring phase at the end of which dysplasic nodules (future malignant tumors) result. In incipient phase, HCC appears like a small region having a different texture than the other parts of the tissue and a diameter of about 1.5 cm to 2 cm. In the case of an evolved HCC, the essential textural attribute is that of heterogeneity, due to the co-existence of regions with necrosis and fibrosis, and of regions with active growths. HCC is also characterized through a very complex structure of vessels. It can present one of the following forms: a clearly delimited, encephalic form, a nodular multicentric form, or a diffuse form [7]. It results the necessity of distinguishing between the early stages of cirrhosis, that not always evolves towards HCC and the advanced stages of this disease consisting in tissue restructuring when malignant nodules (future hepatocellular carcinoma) result. Thus, we chose to distinguish between

usual cirrhosis tissue and cirrhotic parenchyma that surround the HCC tumor.

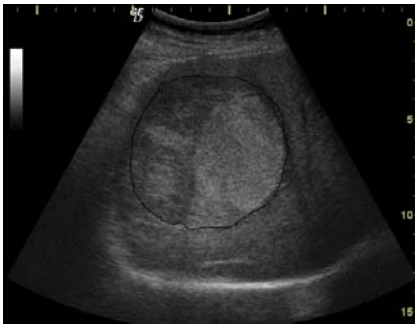

Fig.1. Example of HCC focal form (marked contour)

4 The Proposed Solution

We aim to build, for the liver pathological stages preceding HCC, an imagistic textural model, consisting in the exhaustive set of relevant, independent textural features, able to distinguish the considered stage from the other stages; the associated specific, statistical values of the textural features (maximum, minimum and mean value, and the most probable interval of variation) will also be a part of the model. The mathematical description of the imagistic textural model is given bellow:

Let F be the space of textural features, containing a number of n such features.

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

Then, the textural imagistic model of the tumor, TM, consists in a collection of vectors V_{fi}, associated with each textural feature f_i, containing the specific values that characterize each analyzed class.

$$TM_{stage} = \{V_{f_i}\}_{i=1,..,n} \quad (2)$$

In our case, we start with an initial feature space consisting in a set of textural features, like the *GLCM* second order parameters, edge-based statistics, the fractal Hurst coefficient and the entropy computed on four signal components obtained by applying the wavelet transform at two levels of resolution. The vectors of the imagistic textural model consist in the specific parameters previously mentioned and given bellow, where *Min* (the minimum value), *Max* (the maximum value), *Avg* (the average value) and the *Standard Deviation* are real numbers, the *Relevance*, consisting in an integer, represents the level of the corresponding feature inside the Decision Tree. The P*robability Distribution* is in our case a vector of intervals, each interval having associated a probability value, indicating the likelihood that the parameter has its values inside the interval, for the class of interest.

$$V_f = [Min, Max, Avg, St.Deviation, \mathrm{Re}levance, \mathrm{Pr}obability\ Distrib.] \quad (3)$$

The phases due in order to build these models, are:

1. First, the image selection for the training set building is due. For each considered stage (normal liver, cirrhosis, cirrhotic parenchyma around HCC, and HCC) a corresponding class is built.
2. Then, an image analysis phase is due: textural features computation, using specific methods for texture analysis, relevant features selection and elimination of the redundant features are involved in this process.
3. The learning phase is essential in order to determine the specific, statistical values and the corresponding intervals of variation.
4. At the end, a validation phase is necessary, involving the evaluation of the generated model by providing the selected features at the classifiers inputs and estimating the accuracy of each classifier. A new test set of images, different from the training set, will be used in this phase.
5. Conclusions will be drawn upon the relevant textural features, their specific values and their correlations with the visual features and the possibility of building a general model, characteristic for the evolution towards HCC, will be also studied. The detailed description of the methods used in each of these phases is given in the next paragraphs.

4.1. The Image Analysis Phase

4.1.1. Computation of the textural features

First order statistics, as well as second order statistics, edge-based statistics, fractal based methods and multiresolution methods will be considered here, as we intend to analyze all the imagistic features that could be appropriate for our purpose. These methods are descrybed bellow:

- *First order statistics*

The first order statistics of grey levels considered appropriate for liver tissue characterization in various pathological stages were the mean grey level, the maximum and minimum of the grey levels and the autocorrelation index. The last feature is obtained by applying the autocorrelation function, defined on a region of size N x N through the following expression [8]:

$$\rho(x,y)=\frac{\sum_{u=0}^{N}\sum_{v=0}^{N}I(u,v)I(u+x,v+y)}{\sum_{u=0}^{N}\sum_{v=0}^{N}I^2(u,v)} \quad (4),$$

where I is the intensity function of the ultrasound image. The spatial autocorrelation represents, actually, the correlation of a variable with the spatial localization of that variable. It measures the interdependence level between the values of that variable in different points in space and the strength of this interdependence. It is also characteristic for the texture granularity.

- *The Grey Levels Co-occurrence Matrix (GLCM)*

The Grey Levels Co-occurrence Matrix (GLCM) computes, for each possible pair of grey levels ($g1$, $g2$), the number of pairs of pixels, having intensities $g1$ and $g2$, being in a spatial relationship given by a specified displacement vector $(\vec{dx},\vec{dy})$. The scalar value of $\vec{dx}$, denoted by *dx*, is the horizontal distance between the two pixels of the considered pair belonging to the 2D image, while the scalar value of $\vec{dy}$, denoted by *dy*, is the vertical distance between the same two pixels [8], [9]. The mathematical form of the definition is illustrated bellow:

$$\begin{aligned}&C_D(g_1,g_2)=\#\{((x,y),(x',y')):\\&f(x,y)=g_1,f(x',y')=g_2,|x-x'|=dx,\\&|y-y'|=dy,\operatorname{sgn}(dx\cdot dy)=\operatorname{sgn}((x-x')\cdot(y-y'))\}\end{aligned} \quad (5)$$

where #*S* is the size of the set *S*.

In practice, the GLCM probability is used. In our implementation, we computed the GLCM probability matrix by considering *dx* = 2 and *dy* = 1, because, due to the nature of the ultrasound image, the grey level intensity decreases with the deepness, so we wanted to surprise very accurately the grey level differences in the vertical direction. The second order parameters that we compute from GLCM, considered as being able to best characterize the tumor tissue, are: contrast, variance, local homogeneity, also known as Inverse Difference Moment (IDM), correlation, energy and entropy.

- *Edge-based statistics*

The edge-based statistics, edge-frequency and edge-contrast, also provide useful information, as they compute the relative number of separations between regions with different intensity values (edge frequency) and also the relative amount of the difference between these regions (edge contrast) [10]. Edge contrast measures the differences in intensity level between neighboring pixels. Edge frequency refers to the amount of edge pixels in the area of interest. *The variability in edges orientation* is also taken into consideration, as it is considered appropriate to surprise the chaotic structure of the HCC tissue. In order to compute the value of this feature on the region of interest, the horizontal and vertical gradient is first applied, and then the orientations of the edges are computed in each edge point using the following formula:

$$edge\_orientation=\arctan\left(\frac{G_x}{G_y}\right) \quad (6),$$

In the above formula, G_x is the horizontal gradient and G_y is the vertical gradient obtained for the current image pixel[10]. Thus, the matrix of edge orientations is obtained and each element of this matrix is analyzed in relation with its' neighbors: if there exists an important difference between the value of the current element and the value of one of its' neighbors, the edge orientation variability parameter is increased.

- *The Hurst fractal index*

Fractals provide a measure of the complexity of the grey level structure in a certain region of interest, having the property of self-similarity at different scales. Every texture, characterized through the intensity *I*, can be represented as a reproduction of the copies of *N* basic elements:

$$I=N^D \quad (7),$$

where *D* is the fractal dimension of the texture.
One of the ways to express the fractal dimension is the Hurst coefficient, described in [11].

- *The Wavelet Transform*

The Wavelet Transform performs the decomposition of the signal spectrum in components having various frequencies, thus giving the possibility of analyzing the signal at various resolutions. The decomposition is made around the basic frequency of the spectrum. The expression of the decomposition through the Wavelet Transform is the following [11]:

$$W_{f(x)}(m,n)=\frac{1}{\sqrt{s}}\int_{-\infty}^{\infty}f(x)w_{m,n}(x)dx \quad (8),$$

where $w_{m,n}(x)$ has the following expression:

$$w_{m,n}(x)=g\left(\frac{x-n}{m}\right)=\frac{1}{\sqrt{|m|}}\psi\left(\frac{x-n}{m}\right) \quad (9)$$

Here, ψ is the wavelet basic function, (mother wavelet); it usually takes the form of a sinusoid, giving the name of the transform. In practice the Haar function is used more often [13]. We also consider here the Haar function as the basis for the Wavelet transform. For implementation, we use the discrete form of the Wavelet

decomposition formula. The decomposition is performed on two levels: on the first level, the signal components are computed on the original image; on the second level, the decomposition is performed for each component obtained on the first level: low-low, low-high, high-low and high-high sub-images. Then we compute the entropy value for each component (sub-image), using the expression [12]:

$$Entropy = -\sum_{x=1}^{N}\sum_{y=1}^{M} |I(x,y)| \log_2 |I(x,y)| \quad (10)$$

In our work, all the features were computed on regions of interest of rectangular shape, having 50x50 pixels in size.

4.1.2. Searching for the set of relevant and non-redundant features

Although it seems that the measurement of a big amount of features could lead to increased classification accuracy, the scientific results demonstrate the contrary [13], [14]. Thus, the presence of the irrelevant features could generate noise and reducing the number of features implies also the possibility of reducing the training set size without any loss of accuracy. In our case, we also need to find the most relevant features for building the imagistic textural model that is essential in the computer-aided diagnosis. We aim both for selecting the most relevant features, as well as for eliminating the redundant, dependent features. It is known that there exist dedicated methods (filters and wrappers) for features selection, which evaluates subsets of features using a relevance function; the same thing can be done also during the learning phase of some classifiers. From the class of dedicated methods, we chose the method of Correlation Based Feature Selection (CFS) [15], combined with exhaustive search, in order to analyze all the possible feature subsets. We must mention that this method eliminates the redundant, dependent features, choosing, in the same time, those that are mostly correlated with the class parameter. For each analyzed subset s, a merit is computed:

$$Merit_s = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \quad (11)$$

In the formula above, $\overline{r_{cf}}$ represents the average correlation of the features with the class parameter, while $\overline{r_{ff}}$ is the average correlation between the features. These correlations are established using the symmetrical uncertainty formula [15].

Concerning the classifiers, we will choose the Decision Trees [6] method, this being able to provide a clear set of relevant textural features, also ranked by their individual importance. In order to find the place of each attribute in the structure of the decision tree, the information impurity is computed. The most frequently used mathematical expression is the entropy impurity (information impurity):

$$i(N) = -\sum_{j} P(\omega_j) \log_2 P(\omega_j) \quad (12)$$

In our work, we will use the C4.5 algorithm, described in [6]. In order to eliminate the redundant features and also to improve the accuracy of the entire result, we will look to combine the Decision Trees with the CFS method [15], or with the method of mathematical regression [16]. The performances of these two combinations will be analyzed separately and the best combination will be adopted.

4.2. The Learning Phase

During the learning phase, a representative sample containing a big enough number of features will be used in order to estimate the maximum and minimum value and statistical parameters like the mean and standard deviation [16]. The Bayesian Belief Networks will be also used in order to estimate the specific intervals of values of the relevant features for each class, based on the class-conditional probabilities [17].

4.3. The Validation Phase

During the validation phase, different classifiers, from distinct categories of pattern classification methods, will be used in order to make an objective evaluation concerning the set of relevant features established during the image analysis phase. Thus, from the category of Bayesian Decision Theory, the Bayesian Belief Networks will be used, and from the Artificial Neural Networks category, the Multilayer Perceptron will be chosen. The obtained accuracy will be evaluated using specific accuracy parameters like the recognition rate, the true positive rate (TP rate) or sensitivity, the true negative rate (TN rate) or specificity and the area under the ROC curve. A new distinct image set, having the same structure as the training set, will be used in this phase.

5 Experiments

5.1. Building the training set

First, a training set containing 100 images for each class was built. The considered classes were the following: normal liver, cirrhosis, cirrhotic liver parenchyma on which the hepatocellular carcinoma is already developed and hepatocellular carcinoma. The images were taken under the same settings of the ultrasound machine (gain: 78, frequency: 5.5 MHz, deepness: 16 cm). Thus, the cir-

rhotic parenchyma on which the hepatocellular carcinoma is not evolved and the cirrhotic liver on which the hepatocellular carcinoma is already evolved are both considered, as we aim to establish the distinctions between these two classes in order to enable the detection of the forerunner phases of the hepatocellular carcinoma.

5.2. Experiments done during the image analysis phase

The textural features are computed for each class and their values are stored into the database. The following features are considered: the maximum, minimum and mean value of the grey levels, the autocorrelation index, the GLCM second order parameters (contrast, variance, local homogeneity, correlation, energy and entropy), the edge based statistics, the Hurst fractal index, the entropy computed on the region of interest after applying the wavelet transform at two levels of resolution. Then, all the pairs of classes, corresponding to the successive stages of evolution are considered and the relevant features are determined in each case. The method chosen for feature selection (Decision Trees) is applied individually or in combinations with specific methods in order to assure the elimination of the redundant features; the best method combination, from the accuracy point of view, is chosen for further use. The accuracy results obtained using each feature selection method and the Multilayer Perceptron (MLP) method for validation are illustrated in the table bellow, considering the experimental case that is the most important for us – that of differentiating HCC from the surrounding liver parenchyma, of cirrhotic nature.

Table 1. Comparison of the feature selection methods

MLP Method	**Correctly Classified instances**	**Sensi-ivity**	**Speci-ficity**	**Area under ROC**	**Time** (model building)
Original set	85.83%	0.833	0.883	0.92	4.02s
Decision Trees (J48)	90.66%	0,867	0.967	0.934	0.7s
Decision Trees(J48)+ CFS	89.166%	0.833	0.95	0.911	0.59s
CFS+ Decision Trees (J48)	89.166%	0.85	0.933	0.915	0.89s
Decision Trees(J48) + Regresion	89.166%	0.817	0.967	0.936	0.44s
Regression + Decision Trees	90%	0.85	0.95	0.938	0.7s

From the table above, it results that the method of Decision Trees provided the best values for the accuracy parameters, followed by the combinations between Decision Trees and mathematical regression. Thus, for further use, we have chosen the last combination (Regression + Decision Trees), as it guarantees a set of independent features with maximum accuracy. From our results previously obtained by applying regression [18], we know that the maximum of the grey levels, the minimum of the grey levels and the grey levels mean are linearly correlated, respectively GLCM energy and GLCM entropy are logarithmically correlated; just one feature from each pair is considered in the final feature set. The results that concern the sets of relevant features obtained in each case by applying the method of mathematical regression combined with Decision Trees are illustrated in the table bellow. The ranking of each feature inside each Decision Tree is also provided.

Table 2. Comparing the relevant features that separate between the evolution stages

Compared classes	**The set of relevant features**
HCC *and* surrounding cirrhotic parenchyma	1.*autocorrelation index*, 2.entropy computed after applying the Wavelet Transform, on the first sub-image from the first level, 3. edge contrast, 4. Hurst fractal index, 5. *entropy computed after applying the Wavelet Transform* on the fourth sub-image from the first level (high-high component of the original image), 6. *entropy computed after applying the Wavelet Transform* on the third sub-image from the second level (high-low component on the low-high component from the first level), 7.*Grey levels mean*, 8. Variability in edge orientations
cirrhotic parenchyma surrounding HCC *and* cirrhosis without HCC	*1.Grey levels mean*, 2. *autocorrelation index*, 3. GLCM energy, *4. entropy computed after applying the Wavelet Transform*, on the second sub-image from the first level (low-high component of the original image) 5. GLCM correlation, 6.edge-contrast
HCC *and* normal liver	1.Edge frequency, 2. homogeneity, 3. *autocorrelation index 4.Grey levels mean*

From the table above, we can notice that the mean of the grey levels, the entropy computed after applying the Wavelet Transform, as well as the autocorrelation index are always important for the separation between the successive stages (they are formatted in italic). The mean grey levels, denoting the echogenicity, increases from normal liver to cirrhosis, due to the evolution of the fibrosis grades. As the sub-images obtained after applying the Wavelet Transform result form convolving the image with combinations of high-pass and low-pass filters at various resolutions, we can deduce that the complexity in local features (i.e. edges) increases from one stage to another, due to fibrosis and to the cirrhotic nodules, then to the complexity of the malignant tumor (HCC). The auto-

correlation index is very important in the context of evolution from cirrhosis to HCC, as it characterizes the tissue granularity, which suffers changes due to the nodules. In the case of the differentiation between HCC, the cirrhotic parenchyma surrounding HCC and the simple cirrhosis (cirrhotic parenchyma without HCC), also GLCM correlation and edge contrast become important.

5.3. Experiments done during the learning phase

From the most relevant parameters determined during the image analysis phase, we will detail the evolution of edge contrast, which is relevant for the differentiation of the classes of interest – cirrhosis, cirrhotic parenchyma around HCC and HCC, and also of the autocorrelation index, which is the most important parameter in the evolution of cirrhosis towards HCC.

Table 3. The specific values of the edge contrast in each case

Class	Mean Value	Most probable interval
Cirrhosis	8. 61	(2.708, 13.803]
Cirrhotic parenchyma around HCC	16. 656	(13.803, ∞)
HCC	1. 243	(0, 2.708]

From the table above, it results that the edge contrast parameter takes lower values in the case of HCC and higher values in the case of cirrhosis. In the last case, the increased grey level variability in neighboring locations is due to the small, hyperechogenic nodules. The intervals of values are obtained using the method of Bayesian Belief Networks, and are taken from the probability distribution tables generated in Weka 3.5 [19].

Table 4. The specific values of the autocorrelation index in each case

Class	Mean Value	Most probable interval
Cirrhosis	0.0005	(0.000305,0.485649]
Cirrhotic parenchyma around HCC	0.00017	(0, 0.000305]
HCC	0.0007	(0.485649,∞)

From the table above, it results that the autocorrelation of the grey levels decreases from the incipient stages of cirrhosis towards the most advanced stages of this disease that precedes HCC. Thus, the homogeneity and the granularity decrease accordingly. In the case of HCC, the autocorrelation index is measured inside of the tumor, this is the reason why it has more increased values.

5.4. Experiments done during the validation phase

During this phase, a new test set was built having the same structure as the training set and a number of 70 items/class. The cross validation method with 100 folds was used for each classifier. The Multilayer Perceptron method, with a learning rate of 0.1, a momentum of 0.9 and a number of *a* hidden layers, where a = (number of features + number of classes)/2 was used in Weka 3.5 environment [19]. In the case of separation between the HCC and the surrounding cirrhotic parenchyma, the accuracy obtained by using the relevant features as inputs was 90%, being better than the one obtained using the original set of features (around 85%). The time is significantly improved compared to the original case (0.7s compared to 4.02s). In the other cases, the accuracy obtained using the corresponding relevant features was as follows: for the differentiation between cirrhosis and the cirrhotic parenchyma surrounding HCC, it was 70%; when distinguishing HCC from normal liver, the accuracy was 83.63%. Also, the Bayesian Belief Networks method with K2 search and BMA Estimator [17] was used in Weka 3.5 environment. While for the original feature set it provided an accuracy of 90%, in the case of the selected feature set, it provided an accuracy of 91%, when separating between HCC and surrounding cirrhotic parenchyma. In the other cases, the accuracy obtained using the relevant features with Bayesian Belief Networks was as follows: for the differentiation between cirrhosis and the cirrhotic parenchyma surrounding HCC, 74%; when distinguishing HCC from normal liver, it was 86.36%.

6 Conclusions and Future Work

From the experimental illustrations, it results that our proposed method, for building the imagistic textural models of the liver pathological stages that precede HCC, is appropriate for characterizing and monitoring the evolution towards HCC, being of a great importance in the context of computer aided, non-invasive diagnosis. The relevant textural features are correlated with the expected visual and physical properties of the liver and HCC tissue: the complexity in local features increases towards HCC, denoting the increasing complexity of the liver tissue structure. In order to improve the accuracy, the Gabor filter will be also implemented for texture analysis. In order to improve the imagistic textural model, an error coefficient will be computed during the evaluation phase, for determining the distance between the specific values obtained on the training set, during the learning phase, and the specific values recomputed on the test set. We also aim to build a general model of the evolution towards HCC, using the Markovian model, as each evolution stage depends only on its precedent stage. In this

way, we will be able to predict the probability of the next stage and to prevent the hepatocellular carcinoma.

References:

[1] C. Wu, C. Yung-Chang, H. Kai-Sheng , Texture features for classification of ultrasonic liver images, *IEEE Transactions on Medical Imaging*, Vol. 11, June 1992, pp. 141-152

[2] H. Sujana, S. Swarnamani, S. Suresh, Application of Artificial Neural Networks for the classification of liver lesions by texture parameters , *Ultrasound in Med. & Biol.*, Vol. 22, No. 9, 1996, pp. I177- 1181

[3] T. Chikui, K. Tokumori, K.Yoshiura, K.Oobu, S.Nakamura, K. Nakamura, Sonographic texture characterization of salivary gland tumors by fractal analysis, *Ultrasound in Med. & Biol.*, Vol. 31, No. 10, 2005, pp. 1297–1304

[4] H. Yoshida, D.Casalino, B. Keserci, A. Coskun, O. Ozturk, A. Savranlar, Wavelet-packet-based texture analysis for diferentiation between benign and malignant liver tumours in ultrasound images, *Physics in Medicine and Biology* Vol.48, 2003, pp.3735-3753

[5] A. Madabhushi., M.D. Felman, D.N. Metaxas, J. Tomaszeweski, D. Chte, Automated Detection of Prostatic Adenocarcinoma From High-Resolution Ex Vivo MRI, IE*EE Transactions on Medical Imaging*, December 2005, pp. 1611-1626

[6] R. Duda, P. E. Hart , D. G. Stork, *Pattern Classification (2nd ed)***,** Wiley Interscience, 2003

[7] J. Bruix, J.M. Llovet, Hepatitis B Virus and Hepatocellular Carcinoma, *EASL International Consonsus Conference on Hepatitis B*, 2002.

[8] D.A. Clausi, An analysis of cooccurrence texture statistics as a function of grey level quantization, *Canadian Journal of Remote Sensing*, Vol. 28, No. 1, 2002, pp. 45-62

[9]F. Valckx, J. M. Thijssen, Characterization of echographic image texture by cooccurrence matrix parameters**,** *Ultrasound in Medicine & Biology*,Vol.23, No.4, 1997, pp.559-571

[10] A. K. Jain, *Fundamentals of Digital Image Processing* Englewood Cliffs, NJ: Prentice Hall, Published 1989

[11] J. R. Parker, *Algorithms of Image Processing and Computer Vision***,** Wiley Computer Publishing, 1996, Ch. 4, pp. 171-175

[12]E. Stollnitz, T. DeRose, D. Salesin, Wavelets for computer graphics, *IEEE Computer Graphics and Applications*, Vol. 15, No.3, May 1995, pp.76–84

[13] A. Jain, D. Zonker, Feature Selection: Evaluation, Application, and Small Sample Performance, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, 1997**,** pp**.**153-158

[14] W. Siedlecki, J. Sklansky, On automatic feature selection*", International Journal of Pattern Recognition and Arti cial Intelligence*, Vol. 2, No. 2, 1988, pp. 197 -220

[15] M. Hall, G. Holmes, Benchmarking attribute selection techniques for discrete class data mining, *IEEE Transactions on Knowledge and Data Engineering* , vol. 15, no. 3, May/June 2003, pp. 1-16.

[16]C. Jalobeanu, I. Rasa, *Incertitudine si decizie. Statistica si probabilitati aplicate in mangement***.,** Editura UT Press, Cluj-Napoca, 2001

[17]R. Boukaert, Bayesian Network classifiers in Weka, *http://weka.sourceforge.net/manuals/weka.bn.pdf*,, 2004

[18]D. Mitrea, S. Nedevschi, M. Lupsor, R. Badea, I. Coman, Exploring the textural parameters of ultrasound images to build an imagistic model for prostatic adenocarcinoma (ADKP), *Journal of Automation, Computers and Applied Mathematics* (ACAM), vol. 16 , no. 3, 2007, pp.11-19

[19]Weka 3, Data Mining Software in Java *http://www.cs.waikato.ac.nz/ml/weka/,* 2004