

A Computer-Aided Diagnosis of Liver Tumors Based on Multi-Image Texture Analysis of Contrast-Enhanced CT. Selection of the Most Appropriate Texture Features

Dorota Duda¹, Marek Krętowski¹, Johanne Bézy-Wendling^{2,3}

¹ Faculty of Computer Science, Bialystok University of Technology, Poland

² Signal and Image Processing Laboratory (LTSI), University of Rennes 1, France

³ National Institute of Health and Medical Research (INSERM), University of Rennes 1, France

Abstract. In this work, a system for the classification of liver dynamic contrast-enhanced CT images is presented. The system simultaneously analyzes the images with the same slice location, corresponding to three typical acquisition moments (without contrast, arterial- and portal phase of contrast propagation). At first, the texture features are extracted separately for each acquisition moment. Afterwards, they are united in one “multiphase” vector, characterizing a triplet of textures. The work focuses on finding the most appropriate features that characterize a multi-image texture. At the beginning, the features which are unstable and dependent on ROI size are eliminated. Then, a small subset of remaining features is selected in order to guarantee the best possible classification accuracy. In total, 9 extraction methods were used, and 61 features were calculated for each of three acquisition moments. 1511 texture triplets, corresponding to 4 hepatic tissue classes were recognized (hepatocellular carcinoma, cholangiocarcinoma, cirrhotic, and normal). As a classifier, an adaptive boosting algorithm with a C4.5 tree was used. Experiments show that a small set of 12 features is able to ensure classification accuracy exceeding 90%, while all of the 183 features provide an accuracy rate of 88.94%.

Introduction

In *Global Cancer Statistics*, Jemal et al. (2011) reported that “Liver cancer in men is the fifth most frequently diagnosed cancer worldwide but the second most frequent cause of cancer death. In women, it is the seventh most commonly diagnosed cancer and the sixth leading cause of cancer death”. According to the mentioned research, in 2008, there were estimated to be 748 300 new liver cancer cases in the world, and 695 900 people died from liver cancer. Moreover, Jemal et al. (2011) revealed that the incidence and mortality rate of primary liver cancers were increasing across many

parts of the world. Most patients diagnosed with primary liver cancer die within six months of diagnosis. In this context, the earliest possible detection of such a disease becomes critical to successful treatment.

In clinical practice, a preliminary diagnosis of liver disorders is usually based on several contrast-enhanced Computed Tomography (CT) scans. The first series of images are acquired without contrast media injection. The next two series concern two scenarios (*i*) when injected contrast media reaches the liver through the hepatic artery (*hepatic phase*) and (*ii*) when it reaches the liver through the portal vein (*portal phase*). Thus, an enhancement of two vascular trees (branching from the hepatic artery and from the portal vein) is possible. An evolution of the liver tissue region appearance, over the contrast media propagation, could be a discriminating factor in tumor diagnosis.

Visual analysis of liver CT scans, performed by an experienced radiologist, often is not sufficient to correctly recognize the type of pathology. Due to the fact that those performing the analysis are able to identify only a small part of information stored in images, invasive techniques (such as a needle biopsy) still remain a gold standard for a definitive diagnosis of hepatic disorders. The use of invasive procedures could be avoided if doctors had the appropriate tools to interpret the image content. The solution could be the image-based Computer Aided/Assisted Diagnosis/Detection (CAD) systems, which have recently and rapidly become of growing interest. They include many techniques of image analysis, such as organ segmentation, lesion extraction, and tissue characterization, often based on texture analysis (Bruno et al., 1997), combined with classification algorithms. A large number of publications on the subject proves that the (semi)automatic CAD systems appear to be a powerful tool for supporting medical decisions.

Two main stages of work of a typical image-based CAD system exist, regardless of the imaging technique, the organ analyzed, or the possible diseases to diagnose. The first stage is a system preparation for recognition of a certain number of tissue classes. This stage, called *learning* (or *training*), consists in the classifiers' induction from a set of labeled vectors of features. The learning set is created on the basis of the tissue regions, traced on the images, for which a diagnosis has been verified. The second stage is an application of the classifiers to aid a diagnosis.

Our goal is to develop a (semi)automatic CAD system for aiding a diagnosis of hepatic diseases from dynamic contrast-enhanced CT images. The system we are working on simultaneously analyzes the triplets of images of the same slice of the liver, corresponding to the three different moments of contrast media propagation. Several CAD systems based on liver CT im-

ages have already been investigated. Nevertheless, systems taking texture evolution into account over contrast media propagation in hepatic vessels are still rare. In this study, we focus on the choice of texture features that best describe the “triphase” liver texture.

The rest of the paper is organized as follows: at first, we give a short overview of the proposed CAD systems concerning the hepatic diseases diagnosis, based on CT images. Afterwards, we detail the two stages of work of our system. We also discuss some strategies for choosing the most appropriate texture features for triphase texture characterization: (i) an assessment of a feature stability and of its dependency on the size of the analyzed image region, (ii) feature selection with two searching directions – *Forward* and *Backward*. Then, the classification results obtained with selected features are discussed. Finally, the conclusions and future works are outlined.

Review of Image-Based CAD Systems Based on Liver CT Images

One of the earliest studies concerning a computer-assisted diagnosis based on liver CT images was undertaken by Chen et al. (1998). Their system was able to automatically find the liver boundaries and to recognize two types of liver tumors: hepatoma and hemangioma. In this system, the image texture was characterized by its fractal dimension and the features obtained from the co-occurrence matrices. A probabilistic neural network was used as a classifier. A similar system, but one able to distinguish a healthy liver and liver disease, was later presented by Husain et al. (2000).

In 2003, Gletsos et al. described a system adapted for recognizing the four types of liver tissue: healthy, hepatic cysts, hemangioma, and hepatocellular carcinoma (HCC). They characterized the textures with the features calculated from the co-occurrence matrices. The classifier was composed of three sequentially placed feed-forward neural networks, trained with a back-propagation algorithm. The same tissue types were recognized in the study presented by Stoitsis et al. (2006). Their system tested several sets of texture features, derived from: the gray-level histogram, the co-occurrence matrices, the run-length matrices, the Laws’ texture energy method, and the fractal models. A feature selection, based on genetic algorithms, was performed in order to find the most useful features. Classification was carried out by neural networks and statistical methods. Further continuation of this research has resulted in the creation of a telematics-enabled system for image archiving, management, and diagnosis support (Mougiakakou et al., 2009).

This integrated system performed an image preprocessing, a semi-automatic image segmentation, an extraction of texture features, and a classification.

All of the aforementioned systems processed only one image at a time, acquired without contrast media. Later systems used contrast-enhanced images, but they were still adapted for the analysis of a single image. For example, in 2004, Bilello et al. presented a system working on portal-phase images. It combined the methods for detection, characterization and classification of liver hypodense hepatic tissue (cysts, hemangiomas, and metastases). The texture analysis was performed with frequency methods. As classifiers, the support vector machines were used.

The system described by Smutek et al. (2006) focused on the analysis of focal liver lesions. It used the first- and second order texture features. The analyzed images corresponded to the late portal phase. Another system, developed by Lambrou et al. (2006), differentiated healthy and tumor tissue. To extract texture features, it used a wavelet transform method, in combination with three statistical methods (based on the gray level histogram, the co-occurrence matrices, and the run length matrices). In both systems, an ensemble of Bayesian classifiers was applied.

Still, in 2006, Mala et al. described a system adapted for a recognition of HCC, cholangiocarcinoma, hemangioma, and hepatic adenoma. Their system was able to automatically detect regions affected by a disease, characterize a tissue (using methods based on wavelet transform), select the best texture features, and finally classify tissues, using a probabilistic neural network.

In 2009, Wang et al. tested yet another diagnostic system, which worked with the three types of liver tissue: HCC, hemangioma, and normal tissue. This system used four texture analysis methods (based on the gray level histogram, the co-occurrence matrices, the gray level difference matrices, and the run length matrices), and the support vector machines as classifiers.

Finally, in the work presented by Duda et al. (2004), a simultaneous analysis of the three images, corresponding to typical moments of contrast propagation (without contrast, arterial phase, portal phase) was proposed. At first, the three corresponding textures were characterized separately by features obtained from: the gray level histogram, Laws' texture energy method, the co-occurrence matrices, and the run length matrices. Then, the features corresponding to the three related textures were placed together in a one feature vector, characterizing a triphase texture. Three types of liver tissue were recognized: healthy, HCC, and cholangiocarcinoma. The classification results obtained with the triphase textures were significantly better than the results corresponding to each acquisition moment separately. Fur-

ther work on similar data has confirmed that considering texture evolution over contrast media, propagation could considerably improve classification accuracy (Duda et al., 2006).

The idea of a multiphase texture characterization was also exploited by Quatrehomme et al. (2013). In this case, four acquisition moments were considered: the first one was a pre-injection phase, the next three corresponded to arterial, portal, and late phase of contrast media propagation. Five types of hepatic lesions were classified: cysts, adenomas, hemangiomas, HCC, and metastasis. The multiphase vectors were composed of features calculated separately for four acquisition moments and united in a multiphase vector. The results obtained for the “four-phase” textures were significantly better than the single-phase ones. However, this work, like the two previous ones, did not select the most relevant features for each of the considered acquisition moments.

In the work described by Ye et al. (2009) the quadruples of images were analyzed in order to differentiate four tissue classes: normal, cyst, heman-gioma, and HCC. Only some combinations of the mean pixel values were considered as temporal features. Nevertheless, they did not outperform the texture features (based on the gray level histogram and the co-occurrence matrices) considered separately for each of four phases.

Methods

The two previously described stages of work can also be distinguished in the system that we are developing. The first stage – the construction of the classifiers from the preprocessed database of image triplets – is depicted in Figure 1. After a database creation, triplets of images are formed. An (ordered) triplet contains the images acquired at the same slice location. Each of the images in the triplet corresponds to a different moment of contrast media propagation in the hepatic vessels. The first of them is acquired without contrast, the second and the third – after its injection, in the two typical phases of its propagation, arterial and portal, respectively. The next step, the preprocessing of images, could be optional. It aims at improving the contrast, eliminating the noise or the artifacts. Then, a Region of Interest (ROI) is drawn on each of the three images. The three corresponding ROIs are of the same size and of the same anatomical position. Afterwards, a label is attributed to each triplet of ROIs. It refers to a tissue class, determined on the basis of a verified diagnosis, for example, confirmed by a histopathological study. Then follows a tissue characterization. It is based on the texture

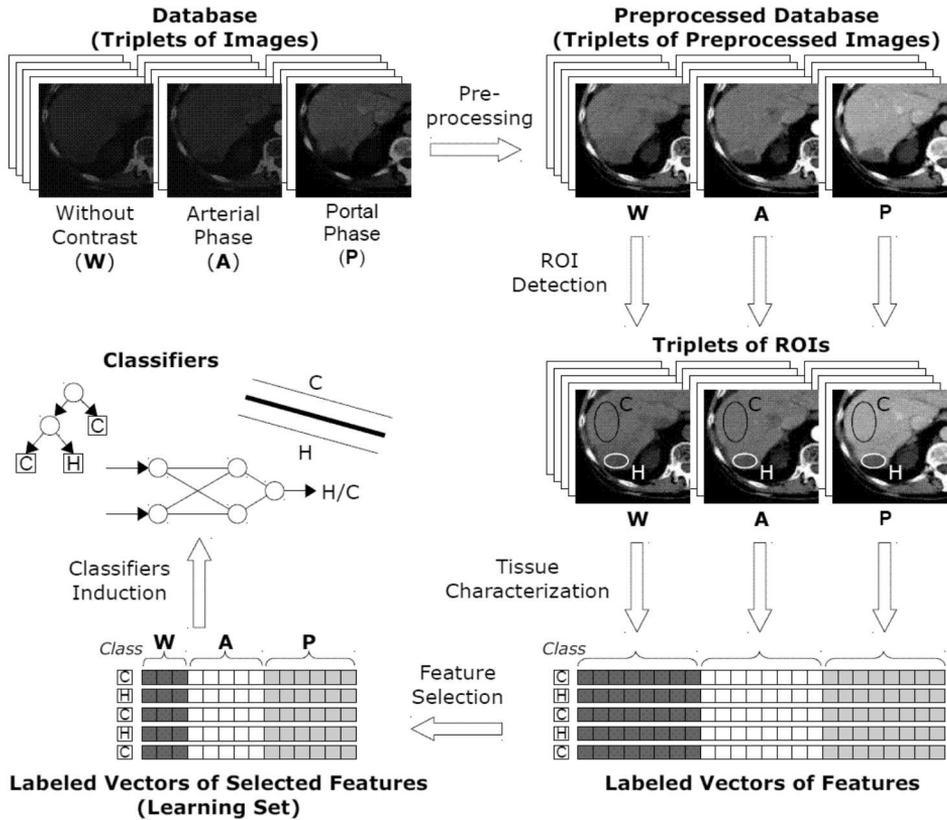


Figure 1. The system for texture-based classification of liver tissues. First stage of work: a construction of the classifiers from the preprocessed database of triplets of images. C and H stand for liver tissue classes, cirrhosis and HCC, respectively

analysis and consists of calculating a set of numerical parameters in order to measure different texture properties, e.g. coarseness, homogeneity, entropy, or local contrast. Such numerical descriptions of the texture are called *texture features*. The tissue characterization is firstly made separately for each of the three corresponding ROIs. Each of the ROIs is characterized with the same features. Then, features corresponding to the three ROIs (thus – to the three acquisition moments) are placed together in the one triphase complex vector, characterizing the triplet of ROIs. This vector contains the information about tissue properties that change over different contrast product concentrations in the vessels. The set of labeled feature vectors is called the training (or the learning) set. Often, not all features are equally useful for a tissue description. For this reason, the selection of the most suitable ones

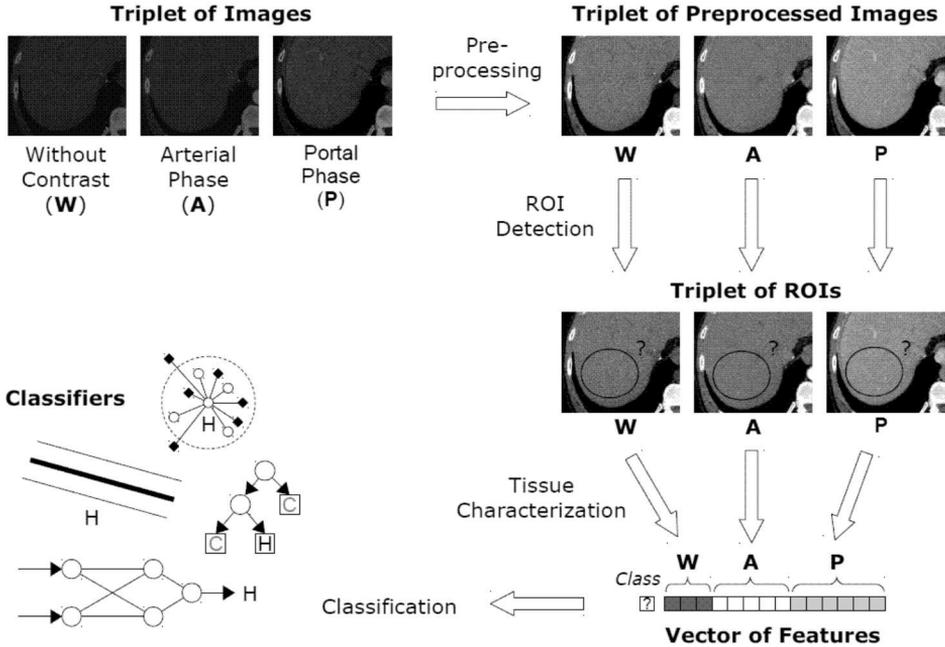


Figure 2. The system for texture-based classification of liver tissues.
Second stage of work: application of the classifiers to aid a diagnosis.
C and H stand for liver tissue classes, cirrhosis and HCC,
respectively

is made. The subsets of features selected for different acquisition moments could differ. Finally, one or several classifiers are constructed on the basis of the training set, composed of vectors of selected features.

The second stage of the system work, i.e., its application to aid a diagnosis, is presented in Figure 2. At this stage, a triplet of images, visualizing the same part of liver, but in the three typical phases of contrast propagation, is necessary. If the image preprocessing was performed at the first stage, now it is also performed, in order to obtain the same properties of the images, as they were in a previous stage. Then, a triplet ROIs is traced, one ROI on each image. Each ROI is characterized with the features that were selected in the previous stage. Features obtained for the three ROIs are placed in one triphase vector. Finally, the classifiers are used in order to propose the most probable tissue class. Such a class is one of the classes considered in the first stage of system work.

One of the major problems encountered during the construction of such systems is a choice of features for the proper tissue characterization. So far, a wide variety of methods for texture feature extraction have been proposed.

They allow the calculation of tens or even hundreds of features describing the various properties of visualized tissue. There is no such set of features that fits each problem, regardless of imaging techniques, acquisition conditions, or the organ that is visualized. However, applying all available features could be impossible for many reasons. Too many features, especially far in excess of the number of objects, could result in data overfitting. The presence of redundant or not reliable features could diminish the classification accuracy. Moreover, handling a relatively big set of observations, described by a large number of features, requires considerable memory resources and can be very time consuming. In the next part of the work, some important aspects for choosing the most relevant features will be discussed.

The ability to properly characterize tissue could be preliminarily assessed taking into account the instability (or the stability) of the feature. If a slight displacement of the ROI results in a significant change in the feature value, the feature can be considered unstable, thus not reliable in the tissue characterization process.

Another way to evaluate the usefulness of the feature could be assessment of its dependency on a ROI size. This is particularly important when the ROI size is different for different patients or for different images (which is true in the case of the database that we explore).

The instability of a feature and its dependency on the ROI size can be determined by a standard coefficient of variation (CV). The coefficient of variation is calculated from a set of feature values, obtained for ROIs located at very close positions, or having almost the same size. It expresses the importance of the variability of several feature values (given by the standard deviation) compared to the absolute value of their average. The more unstable the feature is, the greater its coefficient of variation and the lower its reliability in the process of tissue characterization. We consider that the CV of a feature should not exceed a certain threshold, in order to consider the feature as stable.

In our work, we use the same sets of square ROIs in order to assess both the stability of the feature and its dependency on the ROI size. In each case, the coefficient of variation is calculated on the basis of the same number of feature values obtained, respectively, from ROIs of almost the same locations (approach *Displace*) or from the ROIs of almost the same sizes (approach *Size Changing*). For the first approach, the initial ROI size is first slightly decreased, then the reduced ROI is successively displaced, in order to take all the possible positions within its initial boundaries. For the second approach, the successive ROI vertices are moved by one pixel towards the ROI center, in order to obtain smaller and smaller square ROIs.

The coefficient of variation, obtained with the approaches *Displace* and *Size Changing* is denoted, respectively, CV_D and CV_S . Only the features with relatively small values of both CV_D and CV_S , thus insensitive both to small ROI displacements and to small changes in ROI size, are considered in further analyses.

After identifying the features with a relatively low coefficient of variation, the next step is to assess which subset of them ensures the best possible classification accuracy. When a set of available texture features is large, an evaluation of all of its possible subsets is impossible; the number of subsets of an N -element set is 2^N . Also, there can exist many different feature combinations that ensure comparable classification results. Due to these facts, it is possible and sufficient to test only a small part of the whole subset space.

We perform the feature selection using two searching directions (*Forward*, and *Backward*) in order to test a certain number of feature subsets. Each time, the quality of candidate feature subset is assessed by the classification accuracy ensured with the proposed subset. When the classification accuracy does not grow significantly while adding (in the *Forward* method) or eliminating (*Backward*) the new features, the search stops.

In our work, we propose repeating the following procedure many times: At first, a certain part of observations is randomly chosen from the initial set of observations. Afterwards, a feature selection is performed only on this part of chosen observations. After a multiple repetition of this procedure, we count how often each feature occurs in the final subset of selected features. On the basis of this “feature incidence frequency” a ranking of features is made. Then, we examine how many first features from the ranking are sufficient to ensure the best possible classification accuracy, using the whole set of observations. We suspect that such an approach ensures a better generalization than one single execution of feature selection on the entire observation set. We conduct our experiments separately for *Forward* and *Backward* searching directions.

Experimental Setup

The images were gathered in two hospitals in Rennes (France): Eugene Marquis Anticancer Center (Department of Medical Imaging, Radio Diagnosis Unit) and the University Hospital Pontchaillou (Department of Radiology). The images from the first center were acquired on the *HiSpeed NX/i* device, produced by *GE Medical Systems*. In total, studies on 28 patients

from this center were available. 21 of them were characterized by a slice thickness of 7 mm, the remaining 7 – by a slice thickness of 5 mm. The images from the second center were acquired on the *LightSpeed16* device (*GE Medical Systems*). They corresponded to 48 patients, and were all characterized by a 5 mm slice thickness.

All the images from both hospitals (about 100 images for each patient examination) were acquired using a helical scanner. A standard amount of contrast agent (100 ml) was injected into the patient's arm vein, at a rate of 4 ml/s. The acquisition of images corresponding to the arterial phase of the contrast agent propagation started about 20 seconds after the beginning of the injection. The images corresponding to the portal phase were acquired from 30 to 40 seconds later.

The images were initially stored in DICOM format, of 4096 gray levels. The minimum gray level (zero) corresponded to -2048 Hounsfield Units (HU) and the maximum gray level (4095) corresponded to 2047 HU. All the images were of the size 512×512 pixels. Their reconstruction diameter was between 330 mm and 500 mm.

Four classes of liver tissue were distinguished: HCC (the most frequently observed primary liver tumor), cholangiocarcinoma, the cirrhotic liver, and the healthy liver.

ROIs were manually drawn, as large as possible, on the images, avoiding the big vessels. They entered into the two disjoint ROI sets. The first set (named *Square-ROI-Set*), composed of only square ROIs of the same size (60×60 pixels), was used for assessing the stability of the texture features. It included 62 ROIs for each of: 2 available slice thicknesses, 4 tissue classes, and 3 acquisition moments ($2 \cdot 4 \cdot 3 \cdot 62 = 1488$ ROIs in total). The second one (named *Circular-ROI-Set*) was composed of only circular ROIs with a diameter ranging from 20 to 130 pixels. This ROI set was created on the basis of 2961 images (987 images for each of the three acquisition moments). In total, it included 4533 ROIs (1511 triplets of ROIs), $3 \cdot 319$ ROIs for HCC, $3 \cdot 222$ ROIs for cholangiocarcinoma, $3 \cdot 433$ ROIs for the cirrhotic liver, and $3 \cdot 537$ ROIs for the normal liver.

All the considered images were pre-processed. Since it was found that the range of pixel values characterizing the pixels belonging to all considered ROIs was less than $256 = 2^8$ (more precisely, it was 248), the DICOM images were converted to a 8-bit BMP format. Such a conversion (*windowing*) was done taking the *window width* of 256 levels (HU), and the *window center* of 70 HU. Thus, all the differences in gray levels between all the pixel pairs belonging to considered ROIs were preserved in the resulting images.

On the basis of 8-bit BMP images, 61 texture features were calculated

separately for each of the three acquisition moments. For this purpose, the *MIP (Medical Image Processing) Application* was used. This application was created at the Faculty of Computer Science, Bialystok University of Technology (Poland) and at the Signal and Image Processing Laboratory, Rennes 1 University (France). One of its modules, *Texture Analyzer*, enabled us to extract the texture features, based on the following methods:

- First Order statistics, abbreviated FO,
- Gradient based, GB,
- Co-Occurrence Matrices, COM (Haralick et al., 1973),
- Run Length Matrices, RLM (Chu et al., 1990; Galloway, 1975),
- Gray Level Difference Matrices, GLDM (Weszka et al., 1976),
- Laws Texture Energy, LTE (Laws, 1980),
- Fractal based, FB (Chen et al., 1989),
- Texture Feature Coding Method, TFC (Horng et al., 1996),
- Autocorrelation, AC (Gonzalez et al., 2002).

Table 1 contains the names of features obtained from each method.

Table 1. Texture features chosen for evaluation

Method	Features
AC	<i>Autocorr</i>
COM	<i>AngSecMom, InvDiffMom, Entr, Corr, Contrast, DiffAvg, DiffEntr, DiffVar, SumAvg, SumEntr, SumVar</i>
FB	<i>FractalDim, FractalArea</i>
FO	<i>Avg, Var, Skew, Kurt</i>
GB	<i>GradAvg, GradVar, GradSkew, GradKurt</i>
GLDM	<i>DAvg, DEntr, DAngSecMom, DInvDiffMom, DContrast</i>
LTE	<i>E3L3, S3L3, S3E3, E3E3, S3S3, E5L5, S5L5, W5L5, R5L5, S5E5, W5E5, R5E5, W5S5, R5S5, R5W5, E5E5, S5S5, W5W5, R5R5</i>
RLM	<i>Fraction, RLNonUni, GLNonUni, LongEmp, ShortEmp, LowGLREmp, HighGLREmp, RLEnr</i>
TFC	<i>MeanConv, CodeEntr, Coarse, Hom, CodeVar, ResSim, CodeSim</i>

When applying the COM, GLDM, and RLM methods, the number of gray levels was reduced from 256, used initially, to 64. Other methods used the whole range of 256 levels.

The co-occurrence matrices and the gray level difference matrices were constructed separately for 4 standard directions (0° , 45° , 90° , 135°) and for 5 different distances between the pixel pairs, going from 1 to 5. From each

of 20 thus obtained matrices, the same features were calculated, 11 features by the COM method and 5 features by the GLDM method. Then, 20 values of the same feature, corresponding to different directions and pixel distances, were averaged in order to obtain only one value per feature. The run-length matrices were also constructed for 4 standard directions. Each of them served to calculate the 8 features. The 4 values of the same feature, obtained for different directions of pixel runs, were averaged.

The normalized autocorrelation coefficients (AC method) and the 2 among 7 TFC features (*CodeEntr*, *CodeSim*) were also calculated separately for 4 standard directions, and for 5 different pixel distances, going from 1 to 5. Then, an average of 20 feature values were used to characterize a liver tissue. The remaining 5 TFC features were obtained by averaging 4 values of the same feature, calculated separately for 4 standard directions.

The FB method was based on the fractional Brownian motion model (Chen et al., 1998) and considered 4 pixel distances (1, 2, 3, and 4).

The LTE method provided 19 features, obtained either by the application of 24 filtering masks of 5×5 size or the application of 8 masks of 3×3 size. In the first case, 4 symmetric masks and 10 pairs of asymmetric ones, each pair consisting of a mask and its transposition, were used, in the second – 2 symmetric masks and 3 pairs of asymmetric ones (a mask and its transposition). The sum of elements of each convolution matrix was equal to zero. For each pair of asymmetric masks, the resulting images were added. Images obtained with an application of symmetric masks were multiplied by two. Finally, the entropies of 14 or 5 thus obtained images (with, respectively, 5×5 masks, and 3×3 masks) served as texture features.

In order to assess the feature stability and its dependency on the ROI size, the *Square-ROI-Set* was used. The assessment of the coefficient of variation was performed separately for each of its 24 subsets. Two approaches were separately applied for calculating the coefficient of variation for each feature: *Displace* (giving the CV_D value) and *Size Changing* (for CV_S). The coefficient of variation was always calculated on the basis of 9 feature values. In the *Displace* approach, the ROI was reduced to a 58×58 square in order to take the 9 possible positions inside its initial boundaries. In the *Size Changing* approach, the 9 considered ROIs were of sizes going from 60×60 to 52×52 . For each of the 24 subsets, the average of 62 values of coefficient of variation, obtained separately for each of 62 ROIs, was used for further analyses.

Next, the experiment consisted of selecting a subset of features ensuring the best possible classification accuracy. In this step, only the triplets of

ROIs from the *Circular-ROI-Set* were used. Each triplet was described by 120 texture features, 40 features for each acquisition moment. Only the features insensitive both to small ROI displacements and to small changes in ROI size were considered. This experiment was performed with the *Weka* software (Hall et al., 2009). The following selection settings were applied: the wrapper method – as an evaluator of each tested subset of features (*WrapperSubserEval*), the *C4.5* tree (Quinlan, 1993) (called *J48* in *Weka*), as a classifier, and the *BestFirst* searching strategy with the two searching directions, *Forward* and *Backward* (tested separately). The selection was repeated 100 times. Each time $1007 = 1511 \cdot 2/3$ multiphase feature vectors were randomly chosen for the selection experiment. On the basis of the 100 obtained subsets of selected features, the feature incidence frequency ranking was made.

Finally, for the classification experiment, the entire *Circular-ROI-Set* was used. Each triplet of ROIs was described only by several or several dozen selected texture features, occupying the first positions of the feature incidence frequency ranking. Different numbers of first ranked features were used. The classification was performed with *Weka*, using an *Ensemble of Classifiers* with an adaptive boosting voting scheme (Freund et al., 1997), called *AdaBoostM1* in *Weka*, and a *C4.5* tree (*J48*) as the underlying algorithm. The number of iterations for the *AdaBoostM1* algorithm was 100. The classification accuracy was estimated by 10-fold cross-validation, repeated 10 times.

Results and Discussion

First, we assessed the feature stability and its dependency on the ROI size. Since the presentation of all the coefficients of variation obtained for each of the 61 features (24 values of CV_D , and 24 values of CV_S) would occupy too much space, we limit ourselves, for the moment, to the presentation of general conclusions, drawn from the whole set of results.

- Regardless of the approach (*Displace* or *Size Changing*) and, as far as the same slice thickness is considered, the three averaged coefficient of variation values (CV_D or CV_S), obtained for the three acquisition moments do not differ significantly. Nor do they differ between 4 tissue classes. However, for the *Size Changing* approach, the highest of 4 CV_S values, obtained for different classes, but corresponding to the same slice thicknesses and the same acquisition moment, is observed slightly more frequently for the tumor tissue.

Table 2. Maximum feature CV value (among the values corresponding to 2 approaches, 3 acquisition moments, and 4 classes) obtained for different slice thicknesses: 7 mm and 5 mm. The features are sorted by the increasing maximal CV values, corresponding to the slice thickness of 7 mm. Only the CV values not exceeding a threshold of 0.01 for both slice thicknesses are taken into account

Rank	Feature	7 mm	5 mm	Rank	Feature	7 mm	5 mm
1	<i>Autocorr</i>	0.0009	0.0010	21	<i>S3S3</i>	0.0034	0.0063
2	<i>FractalDim</i>	0.0013	0.0018	22	<i>DInvDiffMom</i>	0.0034	0.0049
3	<i>E3E3</i>	0.0014	0.0023	23	<i>Avg</i>	0.0036	0.0037
4	<i>CodeEntr</i>	0.0014	0.0040	24	<i>InvDiffMom</i>	0.0036	0.0047
5	<i>S5L5</i>	0.0016	0.0024	25	<i>W5S5</i>	0.0036	0.0079
6	<i>S3L3</i>	0.0017	0.0026	26	<i>MeanConv</i>	0.0037	0.0069
7	<i>E5E5</i>	0.0017	0.0025	27	<i>SumAvg</i>	0.0039	0.0039
8	<i>S5S5</i>	0.0017	0.0032	28	<i>CodeVar</i>	0.0044	0.0081
9	<i>ShortEmp</i>	0.0018	0.0017	29	<i>GradAvg</i>	0.0052	0.0060
10	<i>W5E5</i>	0.0019	0.0033	30	<i>DiffEntr</i>	0.0052	0.0034
11	<i>E3L3</i>	0.0021	0.0025	31	<i>DEntr</i>	0.0052	0.0033
12	<i>Fraction</i>	0.0022	0.0020	32	<i>R5R5</i>	0.0054	0.0083
13	<i>W5L5</i>	0.0022	0.0028	33	<i>DAngSecMom</i>	0.0054	0.0062
14	<i>R5S5</i>	0.0025	0.0050	34	<i>Entr</i>	0.0055	0.0042
15	<i>S3E3</i>	0.0026	0.0059	35	<i>RLEnr</i>	0.0058	0.0041
16	<i>R5E5</i>	0.0026	0.0050	36	<i>LongEmp</i>	0.0062	0.0053
17	<i>E5L5</i>	0.0027	0.0030	37	<i>R5W5</i>	0.0064	0.0089
18	<i>R5L5</i>	0.0027	0.0043	38	<i>DAvg</i>	0.0070	0.0072
19	<i>S5E5</i>	0.0030	0.0044	39	<i>SumEntr</i>	0.0072	0.0050
20	<i>W5W5</i>	0.0030	0.0054	40	<i>HighGLREmp</i>	0.0074	0.0076

- The CV_D values are almost always lower than the corresponding CV_S values. This may indicate that studied features are more influenced by the ROI size than by the ROI position. An exception to this rule is observed for the most unstable features (with the largest values of coefficient of variation), such as *Skew* (FO method) or *GradKurt* (GB). We can thus conclude that one should be particularly careful in the choice of features when analyzing the ROIs of different sizes.

As a measure of the feature stability, we finally took the maximum of the 24 CV values, obtained for two different approaches, three acquisition moments and four tissue classes. They are listed in Table 2. For each feature,

Table 3. Ranking of features according to their frequency rate [%], obtained with the *Forward* searching direction. Superscript indices indicate the corresponding acquisition moment. Only the first 33 features are considered

Rank	Feature	Frequency
1	$Avg^{(A)}$	3.94
2	$R5S5^{(P)}$	3.54
3	$SumAvg^{(W)}$	3.27
4	$LongEmp^{(A)}$	3.07
5	$HighGLREmp^{(A)}$	2.67
6	$SumAvg^{(A)}$	2.60
7	$HighGLREmp^{(W)}$	2.54
8	$Avg^{(W)}$	2.47
9	$R5S5^{(W)}$	1.94
10–11	$Avg^{(P)}, SumAvg^{(P)}$	1.87
12	$HighGLREmp^{(P)}$	1.67
13	$S3S3^{(A)}$	1.60
14	$R5E5^{(P)}$	1.40
15–17	$W5S5^{(W)}, ShortEmp^{(A)}, R5S5^{(A)}$	1.34
18–19	$R5E5^{(W)}, Fraction^{(A)}$	1.27
20–21	$R5E5^{(A)}, R5L5^{(P)}$	1.20
22–25	$W5W5^{(W)}, W5W5^{(A)}, RLEntr^{(P)}, S3E3^{(W)}$	1.13
26	$Entr^{(W)}$	1.00
27–33	$ShortEmp^{(W)}, S5S5^{(W)}, MeanConv^{(A)}, Fraction^{(P)}, R5R5^{(P)}, SumEntr^{(A)}, E3E3^{(P)}$	0.93

the maximum CV value was found separately for two different slice thicknesses: 7 mm and 5 mm. Only the most stable features (with the CV not greater than a fixed threshold, 0.01, for both slice thicknesses) are included in the table.

For most of the features, the greater stability (expressed in the lower maximal value of the coefficient of variation) can be observed for thicker slices – those with a thickness of 7 mm. An exception to this rule can be found for the majority of RLM features (*ShortEmp*, *Fraction*, *RLEntr*, *LongEmp*) and for the different measures of texture entropy: *Entr*, *DiffEntr*, *SumEntr* (COM method), *DEntr* (GLDM). When the slice thickness diminishes, smaller and smaller details in the vascular structure became

Table 4. Ranking of features according to their frequency rate [%], obtained with the *Backward* searching direction. Superscript indices indicate the corresponding acquisition moment. Only the first 32 features are considered

Rank	Feature	Frequency
1	<i>HighGLREmp</i> ^(P)	1.46
2	<i>R5R5</i> ^(P)	1.43
3	<i>HighGLREmp</i> ^(W)	1.33
4–6	<i>RLEntropy</i> ^(P) , <i>R5L5</i> ^(P) , <i>CodeEntr</i> ^(W)	1.30
7	<i>FractalDim</i> ^(A)	1.29
8	<i>SumAvg</i> ^(W)	1.25
9–10	<i>R5W5</i> ^(P) , <i>DInvDiffMom</i> ^(P)	1.19
11–12	<i>S5E5</i> ^(A) , <i>SumAvg</i> ^(A)	1.18
13	<i>S3S3</i> ^(P)	1.14
14–18	<i>HighGLREmp</i> ^(A) , <i>FractalDim</i> ^(P) , <i>SumAvg</i> ^(P) , <i>CodeEntr</i> ^(P) , <i>DEntr</i> ^(P)	1.13
19–20	<i>Autocorr</i> ^(A) , <i>Avg</i> ^(P)	1.11
21	<i>RLEnr</i> ^(A)	1.10
22–24	<i>S5E5</i> ^(W) , <i>Avg</i> ^(A) , <i>SumEntr</i> ^(P)	1.08
25–26	<i>LongEmp</i> ^(A) , <i>CodeEntr</i> ^(A)	1.07
27–30	<i>Avg</i> ^(W) , <i>FractalDim</i> ^(W) , <i>SumEntr</i> ^(A) , <i>S3S3</i> ^(A)	1.03
31–32	<i>W5S5</i> ^(A) , <i>R5S5</i> ^(P)	0.99

perceptible in the image. Its texture becomes more and more varied, and its “disorder” increases.

In the next experiment (selection of a subset of features ensuring the best possible classification accuracy), we used only the most stable 40 features (Table 2).

In the case of the *Forward* searching direction, the minimum number of selected features was 7, the maximum – 30, and the average number was about 15. The features corresponding to the arterial phase were slightly more frequently selected. On average, they represented 35.45% of the selected features. The features corresponding to the portal phase, with the frequency of 31.98%, followed.

An application of the *Backward* direction resulted in more numerous subsets of selected features. They included between 45 and 91 features. The average number of features was about 64. In this case, the percentages of features corresponding to each acquisition moment were almost similar.

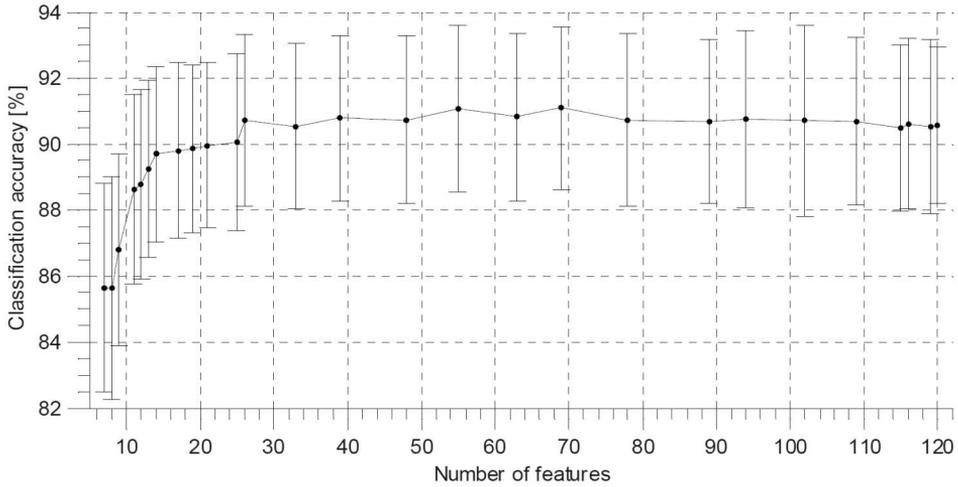


Figure 3. Classification accuracy (with standard deviation) obtained with different numbers of the most frequently selected features with *Forward* direction

The ranking of features, according to their incidence frequency in the entire selection experiment is presented in Tables 3 and 4, for the *Forward* and the *Backward* directions, respectively. For each feature name, superscript indices indicate the corresponding acquisition moment: without contrast (W), arterial phase (A), or portal phase (P). The rankings take into account only a certain number of the best features. In the next experiment, we will see that with such a number of features, it is possible to achieve the highest possible classification accuracy.

Figures 3 and 4 show the classification accuracy obtained for different numbers of first features taken from the rankings presented in Table 3 (concerning the *Forward* direction), and Table 4 (*Backward* direction).

From Figure 3, we can notice that the use of 26 features, most frequently selected with the *Forward* searching direction (Table 3), leads to the correct recognition of 90.71% of the observations. Even if the best possible result (91.10%) is obtained for a set of 69 features, using such a “better” and larger set of features might not be necessary. Taking into account the standard deviations (approximately 2.50%), we can conclude, that the improvement achieved in this case (0.39%) is not significant.

We can also observe that the 26 most selected features derive from only 4 extraction methods: LTE, RLM, COM, and FO (12, 7, 4, and 3 features, respectively). The most represented acquisition moment is, in this case, the arterial phase (10 of 26 features), then follows the no-contrast phase (9 features) and the arterial phase (7 features).

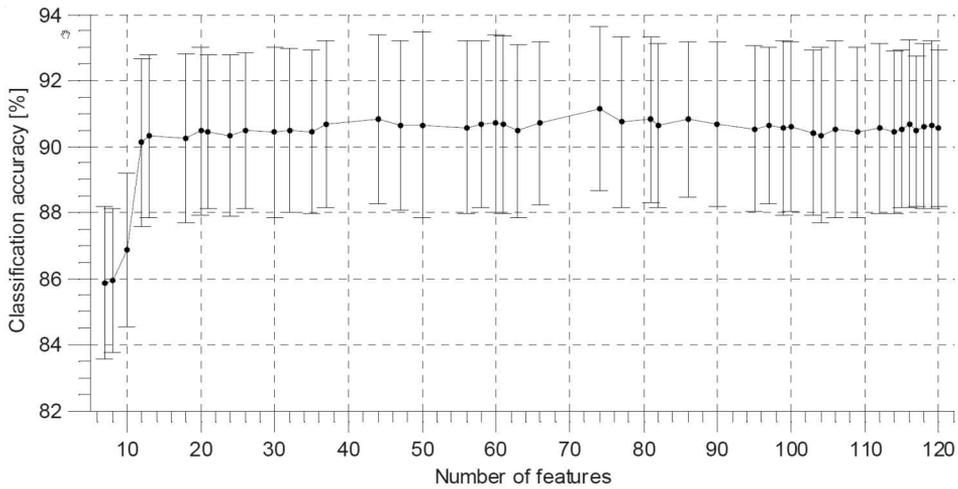


Figure 4. Classification accuracy (with standard deviation) obtained with different numbers of the most frequently selected features with *Backward* direction

In the case of the features most frequently selected with the *Backward* direction (Table 4, Figure 4), the classification accuracy exceeds 90% even with 12 features. The classification results do not change significantly even when more and more features are considered. The best result (91.15% with a standard deviation of 2.49%) is now obtained with the set of 74 features.

This time, the set of the most frequently selected features includes features obtained from the 6 extraction methods: LTE (4 features), RLM (3 features), COM (2 features), GLDM, FB, and TFC (one feature for each of the last three methods). The features corresponding to the portal phase acquisition are more numerous represented in this set. They constitute half of the set. Each of the remaining acquisitions (no-contrast, and the arterial phase one) is represented by 3 features.

For comparison, the classification accuracy obtained with the set of all the $3 \cdot 61 = 183$ features (including the unstable features, and dependent on ROI size) was 88.94%.

Conclusions and Future Work

In this paper, an image-based CAD system was presented. The advantage of the system is its ability to analyze a triplet of images simultaneously. Each image in a triplet visualizes the same liver slice, and corresponds to a different moment of contrast media propagation. The first one is taken

without contrast, the second – in the arterial phase, and the third – in the portal phase. At first, the liver texture in each acquisition moment is characterized. Afterwards – features corresponding to the three acquisition moments are united in one multiphase vector. Thus, the texture evolution over the contrast media propagation is characterized.

The study focused on a choice of the best texture features for description of triplets of textures. 61 features were tested for each of the three acquisition moments (183 features in total). At the first step, $3 \cdot 21$ features that were unstable or dependent on the ROI size were excluded. Then $3 \cdot 40 = 120$ remaining features, united in triphase vectors, were subjected to a selection. The classification experiments showed that a small set of the most frequently selected features (composed of 26 or even 12 features) is able to ensure the classification accuracy that is comparable to, or even better than, the accuracy achieved by using all of the 120 features. The features corresponding to each acquisition moment were different.

In the future, we plan to search for the features that are independent on the image resolution (or on the reconstruction diameter). It seems that it would also be interesting to apply a similar method of multiphase texture characterization for other classification tasks, wherever the organ is visualized repeatedly, each time under different acquisition conditions.

Acknowledgements

We thank Dr. Yann Roland, Dr. Andre Carsin, and Dr. Damien Olivie for their medical contribution to this study. This work was supported by the grant S/WI/2/2013 from Bialystok University of Technology.

R E F E R E N C E S

- Bilello, M., Gokturk, S. B., Desser, T., Napel, S., Jeffrey, R. B., & Beaulieu, C. F. (2004). Automatic detection and classification of hypodense hepatic lesions on contrast-enhanced venous-phase CT. *Med. Phys.*, 31(9), 2584–2593. DOI: 10.1118/1.1782674.
- Bruno, A., Collorec, R., Bezy-Wendling, J., Reuze, P., & Rolland, Y. (1997). Texture analysis in medical imaging. In C. Roux & J.-L. Coatrieux (Eds.), *Contemporary perspectives in three-dimensional biomedical imaging* (pp. 133–164). Amsterdam, Netherlands: IOS Press. DOI: 10.3233/978-1-60750-874-8-133.

- Chen, C.-C., DaPonte, J., & Fox, M. (1989). Fractal feature analysis and classification in medical imaging. *IEEE Trans. Med. Imag.*, 8(2), 133–142. DOI: 10.1109/42.24861.
- Chen, E.-L., choo Chung, P., Chen, C.-L., Tsai, H.-M., & Chang, C.-I. (1998). An automatic diagnostic system for CT liver image classification. *IEEE Trans. Biomed. Eng.*, 45(6), 783–794. DOI: 10.1109/10.678613.
- Chu, A., Sehgal, C., & Greenleaf, J. (1990). Use of gray value distribution of run lengths for texture analysis. *Pattern Recog. Lett.*, 11(6), 415–419. DOI: 10.1016/0167-8655(90)90112-F.
- Duda, D., Kretowski, M., & Bezy-Wendling, J. (2004). Texture-based classification of hepatic primary tumors in multiphase CT. In C. Barillot, D. Haynor, & P. Hellier (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI’2004* (Vol. 3217, pp. 1050–1051 Part II). Springer Berlin Heidelberg. DOI: 10.1007/978-3-540-30136-3_133.
- Duda, D., Kretowski, M., & Bezy-Wendling, J. (2006). Texture characterization for hepatic tumor recognition in multiphase CT. *Biocybern. Biomed. Eng.*, 26(4), 15–24. Retrieved July 31, 2013, from http://www.ibib.waw.pl/bbe/bbefulltext/BBE_26_4.015.FT.pdf.
- Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1), 119–139. DOI: 10.1006/jcss.1997.1504.
- Galloway, M. M. (1975). Texture analysis using gray level run lengths. *Comput. Graph. Image Process.*, 4(2), 172–179. DOI: 10.1016/S0146-664X(75)80008-6.
- Gletsos, M., Mougikakou, S., Matsopoulos, G., Nikita, K., Nikita, A., & Kelekis, D. (2003). A computer-aided diagnostic system to characterize CT focal liver lesions: design and optimization of a neural network classifier. *IEEE Trans. Inf. Technol. Biomed.*, 7(3), 153–162. DOI: 10.1109/TITB.2003.813793.
- Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing* (2nd ed.). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1), 10–18. DOI: 10.1145/1656274.1656278.
- Haralick, R., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Trans. Syst. Man Cybern.*, SMC-3(6), 610–621. DOI: 10.1109/TSMC.1973.4309314.
- Hornig, M.-H., Sun, Y.-N., & Lin, X.-Z. (1996). Texture feature coding method for classification of liver sonography. In B. Buxton & R. Cipolla (Eds.), *Computer Vision – ECCV’96* (Vol. 1064, pp. 209–218 Part I). Springer Berlin Heidelberg. DOI: 10.1007/BFb0015537.

- Husain, S., & Shigeru, E. (2000). Use of neural networks for feature based recognition of liver region on CT images. In *Neural Networks for Signal Processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop*, 11–13 December 2000. (Vol. 2, pp. 831–840). New York, USA: The IEEE, Inc. DOI: 10.1109/NNSP.2000.890163.
- Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., & Forman, D. (2011). Global cancer statistics. *CA: A Cancer J. Clin.*, 61(2), 69–90. DOI: 10.3322/caac.20107.
- Lambrou, T., Linney, A. D., & Todd-Pokropek, A. (2006). Wavelet transform analysis and classification of the liver from computed tomography datasets. In *Proceedings of the 6th International IEEE EMBS Special Topic Conference*. 26–28 October 2006. Retrieved July 31, 2013, from <http://medlab.cs.uoi.gr/itab2006/proceedings/medicalimaging/107.pdf>.
- Laws, K. I. (1980). *Textured image segmentation*. Unpublished doctoral dissertation, University of Southern California, Los Angeles, California, USA.
- Mala, K., Sadasivam, V., & Alagappan, S. (2006). Neural network based texture analysis of liver tumor from computed tomography images. *Int. J. Biol. Life Sci.*, 2(1), 33–40. Retrieved July 31, 2013, from <http://www.waset.org/journals/ijbels/v2/v2-1-5.pdf>.
- Mougiakakou, S., Valavanis, I., Mouravliansky, N., Nikita, K., & Nikita, K. (2009). Diagnosis: a telematics-enabled system for medical image archiving, management, and diagnosis assistance. *IEEE Trans. Instrum. Meas.*, 58(7), 2113–2120. DOI: 10.1109/TIM.2009.2015538.
- Quatrehomme, A., Millet, I., Hoa, D., Subsol, G., & Puech, W. (2013). Assessing the classification of liver focal lesions by using multi-phase computer tomography scans. In H. Greenspan, H. Muller, & T. Syeda-Mahmood (Eds.), *Medical Content-Based Retrieval for Clinical Decision Support – MCBR-CDS 2012* (Vol. 7723, pp. 80–91). Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-36678-9_8.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Smutek, D., Shimizu, A., Tesar, L., Kobatake, H., Nawano, S., & Svacina, S. (2006). Automatic internal medicine diagnostics using statistical imaging methods. In *19th IEEE International Symposium on Computer-Based Medical Systems (CBMS'2006)*, 22–23 June 2006 (pp. 405–412). Los Alamitos, California, USA: The IEEE Computer Society Press. DOI: 10.1109/CBMS.2006.56.
- Stoitsis, J., Valavanis, I., Mougiakakou, S. G., Golemati, S., Nikita, A., & Nikita, K. S. (2006). Computer aided diagnosis based on medical image processing and artificial intelligence methods. *Nucl. Instrum. Methods Phys. Res., Sect. A*, 569(2), 591–595. DOI: 10.1016/j.nima.2006.08.134.

- Wang, L., Zhang, Z., Liu, J., Jiang, B., Duan, X., Xie, Q., Hu, D., Li, Z. (2009). Classification of hepatic tissues from CT images based on texture features and multiclass Support Vector Machines. In W. Yu, H. He, & N. Zhang (Eds.), *Advances in Neural Networks ISNN 2009* (Vol. 5552, pp. 374–381 Part 2). Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-01510-6.43.
- Weszka, J. S., Dyer, C. R., & Rosenfeld, A. (1976). A comparative study of texture measures for terrain classification. *IEEE Trans. Syst., Man Cybern., SMC-6*(4), 269–285. DOI: 10.1109/TSMC.1976.5408777.
- Ye, J., Sun, Y., Wang, S., Gu, L., Qian, L., & Xu, J. (2009). Multiphase CT image based hepatic lesion diagnosis by SVM. In *2nd International Conference on Biomedical Engineering and Informatics (BMEI'2009)*, 17–19 October 2009 (pp. 1–5). New York, USA: The IEEE, Inc. DOI: 10.1109/BMEI.2009.5304774.