Proceedings of the 29th Annual International
Conference of the IEEE EMBS
Cité Internationale, Lyon, France
August 23-26, 2007.

Evaluation of Texture Features in Hepatic Tissue Characterization from Non-enhanced CT Images

Ioannis K. Valavanis, Stavroula G. Mougiakakou, Alexandra Nikita, Konstantina S. Nikita

Abstract—**Aim of this paper is to evaluate the diagnostic contribution of various types of texture features in discrimination of hepatic tissue in abdominal non-enhanced Computed Tomography (CT) images. Regions of Interest (ROIs) corresponding to the classes: normal liver, cyst, hemangioma, and hepatocellular carcinoma were drawn by an experienced radiologist. For each ROI, five distinct sets of texture features are extracted using First Order Statistics (FOS), Spatial Gray Level Dependence Matrix (SGLDM), Gray Level Difference Method (GLDM), Laws' Texture Energy Measures (TEM), and Fractal Dimension Measurements (FDM). In order to evaluate the ability of the texture features to discriminate the various types of hepatic tissue, each set of texture features, or its reduced version after genetic algorithm based feature selection, was fed to a feed-forward Neural Network (NN) classifier. For each NN, the area under Receiver Operating Characteristic (ROC) curves (A_z) was calculated for all one-vs-all discriminations of hepatic tissue. Additionally, the total A_z for the multi-class discrimination task was estimated. The results show that features derived from FOS perform better than other texture features (total A_z: 0.802±0.083) in the discrimination of hepatic tissue.**

I. Introduction

THE most common imaging techniques for the detection and diagnosis of hepatic lesions include Ultrasonography (US), Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and angiography. Although one of the most popular and highly accurate techniques is CT after administration of contrast agents, the use of iodinated contrast agents is related with renal toxicity and allergic reactions. Furthermore, in some cases the diagnosis has to be confirmed with invasive procedures (biopsies). Computer-Aided Diagnosis (CAD) systems based on the combined use of image processing and artificial intelligence techniques attract much attention, since they can provide diagnostic assistance to clinicians.

The use of image derived features has already been proposed for the discrimination of hepatic tissue in CT images. Mir *et al.* [1] proposed the use of Spatial Gray Level Dependence Matrix (SGLDM), Gray Level Run Length Method (GLRLM), and Gray Level Difference Method (GLDM) derived texture features in order to discriminate normal from malignant hepatic tissue. Chen et al. [2] used SGLDM texture features fed a Probabilistic Neural Network (P-NN) for the characterization of hepatic tissue (hepatoma and hemangioma), and Gletsos et al. [3] used SGLDM based features fed to a system of three sequentially placed Neural Networks (NNs) to classify hepatic tissue into four categories. In [4], Huang et al. proposed the use of auto-covariance texture features and a Support-Vector-Machine (SVM) for discriminating benign from malignant hepatic lesions. Seltzer et al. [5] used a wide set of image derived features (e.g. size, shape, diffuse of liver disease) to develop statistical prediction rules into a computer model to classify benign from malignant and hepatocyte-containing vs non–hepatocyte tissue. Although a lot of effort has been devoted to liver tissue characterization from CT images, the potential of different texture features has not been systematically assessed.

In this study, five distinct types of texture features are used and evaluated in the multi-class discrimination of hepatic tissue of Regions of Interest (ROIs) on abdominal non-enhanced CT images into four classes: normal liver (C1), cyst (C2), hemangioma (C3), and hepatocellular carcinoma (C4).

II. Methodology

In order to assess the potential of various texture features in the discrimination of hepatic lesions from CT images the following steps were preformed: i) five sets of texture features were estimated, ii) if the dimensionality of the features set was greater than a predefined threshold, a Genetic Algorithm (GA) based feature selection method was applied, iii) each of the initial and reduced feature sets was fed to a NN classifier, and iv) the area under Receiver Operating Characteristic (ROC) curves (A_z) was calculated.

A. Image Acquisition

Abdominal non-enhanced CT images with a spatial resolution of 512×512 pixels and 8-bit gray-level at the

Manuscript received April 16, 2007. This work was supported in part by the General Secretariat of Research and Technology / Ministry of Development (Greece).

I. K Valavanis is with Faculty of Electrical and Computer Engineering, National Technical University of Athens, 9 Heroon Polytechneiou Str. 15780 Zographou, Greece (e-mail: ivalavan@biosim.ntua.gr).

S. G. Mougiakakou is with Faculty of Electrical and Computer Engineering, National Technical University of Athens, 9 Heroon Polytechneiou Str. 15780 Zographou, Greece (corresponding author, phone: +30 210 772 2968; fax: +30 210 772 3557; e-mail: smougia@cc.ece.ntua.gr).

A. Nikita is with the Department of Radiology, Medical School, University of Athens, Athens, Greece (e-mail: anikita@cc.uoa.gr).

Prof. Konstantina S. Nikita is with Faculty of Electrical and Computer Engineering, National Technical University of Athens, 9 Heroon Polytechneiou Str. 15780 Zographou, Greece (e-mail: knikita@cc.ece.ntua.gr).

W150+60 window taken from both patients and healthy controls were acquired using a Philips CT LX Scanner. A total of 38 healthy controls were identified by the radiologist, while the diagnosed hepatic lesions from patients with C2 (15 patients), C3 (24 patients), and C4 (20 patients), were validated by needle biopsies, density measurements, and the typical pattern of enhancement after the intravenous injection of iodine contrast. The patient statistics are presented in Table I. The position and size of the lesions were defined in CT images by an experienced radiologist. It has to be noted that special care was taken in identifying ROIs in abnormal liver, so as to include the largest possible area of the lesion and avoid the margins of the lesion as well as hypodense areas, which could correspond to necrotic regions. For each patient a maximum of two free-hand ROIs were taken from different slices, resulting in a total of 147 ROIs, 76 of which corresponded to C1, 19 to C2, 28 to C3, and 24 to C4.

B. Feature Extraction

Five distinct sets of features using five texture analysis methods were calculated for each ROI:

First order Statistics: For each ROI, six features were obtained [6]: average gray level (*avg*), standard deviation (*sd*), entropy (*ent*), coefficient of variance (*cv*), skewness (*sk*), and kurtosis (*kur*).

Spatial Gray-Level Dependence Matrix: Eight texture characteristics were calculated from the SGLDM of each ROI [6]-[7]: angular second moment (*asm*), contrast (*con*), correlation (*corr*), sum of squares (*ss*), inverse difference moment (*idm*), entropy (*ent*), homogeneity (*hg*), cluster tendency (*clt)*. Six different values of intersample spacing d=1, 2, 4, 6, 8, 12 pixels were used for obtaining SGLDM, and the feature values were computed by averaging over four uniformly distributed angular directions, 0^o, 45^o, 90^o, and 135^o for each d. Thus, a total of 48 texture characteristics was obtained.

Gray-Level Difference Matrix: Application of the GLDM to each ROI resulted in five features: contrast (*con*), mean (*mn*), entropy (*ent*), inverse difference moment (*idm*), and angular second moment (*asm*) [8]. GLDM was calculated for d=1, 2, 3, 4 pixels and the feature values were calculated as the mean value of the feature estimations for the directions $(0, d)$, $(-d, d)$, $(d, 0)$, $(-d, -d)$ for each d. A total of 20 features was derived for each ROI.

TABLE I
PATIENT STATISTICS

	Age (years)			Sex (male/female)
	min	*max*	*mean*	
C1	35	70	55,5	20/18
C2	44	70	58,3	7/8
C3	45	63	52,7	11/13
C4	61	78	70,5	13/7

Laws' Texture Energy Measures: The TEM were derived from the vectors $L3$={1, 2, 1}, $E3$={-1, 0, 1}, and $S3$={-1, 2, -1} [9]. Convolving those vectors with themselves or with one another, the vectors $L5$={1, 4, 6, 4, 1}, $S5$={-1, 0, -2, 0, -1}, and $E5$={-1, -2, 0, 2, 1} are calculated. Multiplication of the column vectors of length 5 with the row vectors of the same length results in the 5x5 mask: $L5^TE5$, $L5^TS5$, $-E5^TS5$, and $R5^TR5$. After the convolution of each mask with the ROIs, texture statistics were applied, estimating absolute sum / # of pixels (*as*), sum of squares / # of pixels (*ss*), and entropy (*ent*). Thus, the resulting feature vector contained twelve texture features.

Fractal Dimension Texture Measurement: For each ROI, a 3-dimensional feature vector was estimated from the FDM. The components of the feature vector correspond to the parameters $H^{(1)}$, $H^{(2)}$, and $H^{(3)}$ of the multiresolution fractal feature vector [10].

C. Feature Selection

In order to reduce the dimensionality of feature sets with more than ten features (SGLDM, GLDM and TEM), feature selection based on GA was applied [3], [11]. According to [3], the use of GAs results in more robust feature vectors as compared to other deterministic feature selection techniques, in problems related to liver tissue classification from CT images.

D. Neural-Network Classifier

Each of the feature sets, full or reduced, was used as input to a NN based classifier in order to classify ROIs into one of the four classes. The used NNs consisted of an input layer with a number of input neurons equal to the number of the features of the used set, one hidden layer with variable number of neurons, and one output layer with two neurons. The output neurons encoded the four classes of hepatic tissue (00=C1, 01=C2, 10=C3, 11=C4). The back-propagation algorithm with adaptive learning rate and momentum [12] was used for NN training. In order to find the appropriate number of hidden neurons, and the values of learning rate and momentum for each NN, a trial-and-error process was applied.

In order to obtain reliable results on the potential of the feature sets to discriminate the ROIs, the development and evaluation of the NNs was based on the bootstrap method. To this end, a training set consisting of 147 ROIs was sampled with replacement from the available 147 ROIs. The ROIs not appearing in the training set were randomly allocated into two equally sized sets (validation and testing sets). The procedure was repeated for 50 times, resulting in 50 groups of training, validation and testing sets. For each of the 50 groups of sets, a NN using one of the texture feature sets was optimally constructed and trained using training and validation sets and finally tested in the testing set using ROC curves.

E. ROC Curves Generation

In the simplest case of a NN with one output neuron the

ROC curves are generated by thresholding the output neuron in the range [0, 1] and estimating the True Positive (*TP*) and False Positive (*FP*) rates.

In the case of NNs with two output neurons the thresholds for assigning 0 or 1 to the output neurons changed depending on the hepatic class tested again others: e.g. for C2 with encoding 01 both thresholds changed parallel from [1, 0] towards [0, 1] with a step equal to 0.02. Measurements for *TP* and *FP* rates were estimated, ROC curve was generated and area under ROC curve (A_z) was calculated. For each NN four ROC curves were generated for each binary classification task of type "one vs all" e.g. normal vs cyst, hemangioma, and hepatocellular carcinoma, and consequently four A_z $(A_{z,i}, i = 1,...,4)$ values were estimated. A total A_z value was calculated for the four class problem using the following equation [13]:

$$A_{z,total} = \sum_{i=1}^{4} A_{z,i} p_i \quad (1)$$

where p_i is the prevalence of the i^{th} class in testing set.

The whole process was repeated for all bootstrap testing sets.

III. Results and Discussion

ROC curves for the four binary classification tasks, the corresponding A_z measurements and total A_z measurements were calculated using as input each one of the total eight feature sets derived. These feature sets correspond to the FOS, FDM, TEM, GLDM and SGLDM features with full dimensionality and to the reduced TEM, GLDM and SGLDM sets obtained after feature selection. The average ROC curves obtained using the 50 bootstrap testing sets are depicted in Fig. 1, while Table II presents the A_z values (mean ± standard deviation) calculated for each feature set using the 50 bootstrap testing sets.

From both Fig. 1 and Table II it is shown that FOS feature set outperforms all other full or reduced feature sets in all binary decision cases and in the total A_z measurement. FOS feature set achieved the maximum mean total A_z measurement (0.802 ± 0.08). The second best performing feature set in terms of total A_z measurement is the TEM set in its full (0.754 ± 0.063 using 12 features) or reduced version (0.765 ± 0.066 using 8 features). Full or reduced TEM feature sets are the second best performing input sets in all binary decision cases, but in the case of "cyst vs all" where they are outperformed by GLDM feature sets. SGLDM and GLDM feature sets seem to follow in total performance with rather small differences in total A_z measurements (mean A_z within the range [0.67 0.72]), whereas FDM is the worst performing feature set (total A_z: 0.613±0.059 using only three features).

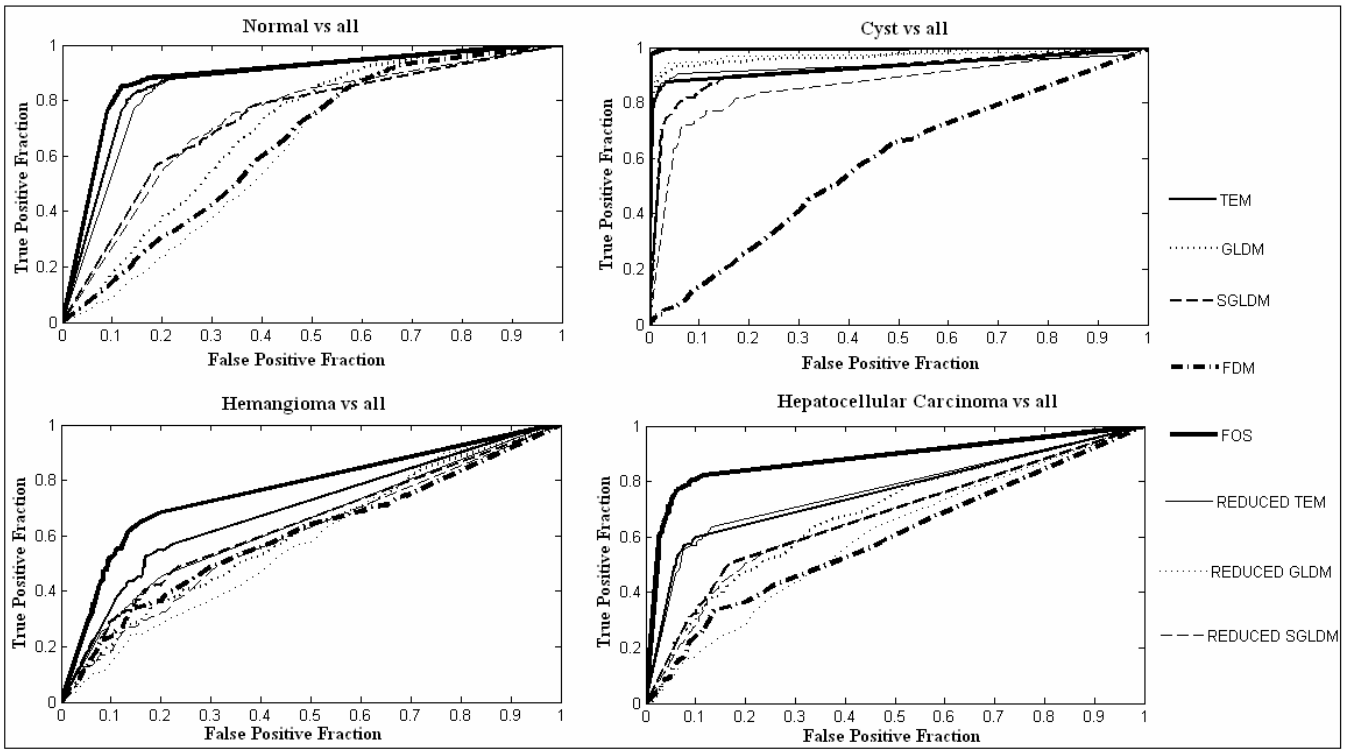

Fig. 1. Average ROC curves for all "one vs all" classifications using the obtained feature sets

TABLE II
A_Z MEASUREMENTS FOR ALL "ONE VS ALL" CLASSIFICATIONS AND TOTAL A_Z MEASUREMENTS (MEAN ± STANDARD DEVIATION)

Features	# of features	C1 vs all	C2 vs all	C3 vs all	C4 vs all	Total
TEM	12	0.868±0.069	0.929±0.095	0.696±0.144	0.759±0.095	0.754±0.063
GLDM	20	0.701±0.079	0.975±0.046	0.603±0.166	0.693±0.125	0.720±0.071
SGLDM	48	0.734±0.084	0.917±0.089	0.638±0.129	0.672±0.096	0.707±0.060
FOS	6	**0.887±0.057**	**0.996±0.016**	**0.765±0.131**	**0.882±0.101**	**0.802±0.084**
FDM	3	0.651±0.088	0.585±0.014	0.601±0.184	0.614±0.127	0.613±0.059
Reduced TEM	8	0.862±0.067	0.935±0.148	0.636±0.157	0.767±0.104	0.765±0.066
Reduced GLDM	5	0.636±0.085	0.980±0.059	0.580±0.144	0.601±0.111	0.666±0.075
Reduced SGLDM	8	0.741±0.091	0.864±0.111	0.595±0.150	0.670±0.140	0.703±0.083

It has to be noted that the application of feature selection keeps the high A_z measurements of TEM feature set, thus providing an equally performing set of lower dimensionality (eight out of 12 features were selected). Total A_z measurements imply the same for SGLDM set (eight out of 48 features were selected). A moderate reduce in total A_z measurement was observed for the reduced GLDM set (five out of 20 features were selected).

Finally, A_z values for the binary decisions show that C3 is the most difficult case to discriminate using the texture features used in this study. On the other hand, C2 is the easiest to discriminate with the maximum A_z value of all other liver tissue types (0.996 ± 0.016 obtained using FOS features). It has to be noted that high A_z values for the discrimination "C2 vs all" don't introduce bias in the total A_z values as the prevalence of C2 in bootstrap testing sets, which is used in (1), is rather small.

Future work will evaluate the use of the proposed texture features and classifiers in the discrimination of other types of liver lesions. Furthermore, a physiology based interpretation of the obtained results is in progress (e.g. why a texture feature set discriminates better liver tissue given the tissue properties it reflects).

IV. CONCLUSION

In the current study, the potential of texture feature sets was evaluated in characterizing liver ROIs from non-enhanced CT images. Average ROC and A_z were calculated using bootstrap derived testing sets. It has been shown that FOS based features provide superior results with a mean A_z equal to 0.802.

REFERENCES

[1] A. H. Mir, M. Hanmandlu, S. N. Tandon, "Texture analysis of CT images," *IEEE Eng. Med. Biol. Mag.*, vol. 14, no. 6, pp. 781-786, Nov/Dec 1995.

[2] E. L. Chen, P-C. Chung, C. L. Chen, H. M. Tsa, C. I. Chang, "An automatic diagnostic system for CT liver image classification," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 6, pp. 783-794, 1998.

[3] M. Gletsos, S. G. Mougiakakou, G. K. Matsopoulos, K. S. Nikita, A. Nikita, D. Kelekis, "A computer-aided diagnostic system to characterize CT focal liver lesions: Design and optimization of a neural network classifier," *IEEE Trans. Inform. Techn. Biomed.* vol. 7, pp. 153-162, 2003.

[4] Y. L. Huang, J. H. Chen, W .C. Shen, "Diagnosis of hepatic tumors with texture analysis in non-enhanced computed tomography images," *Academic Radiology*, vol. 13, pp. 713-720, 2006.

[5] S. E. Seltzer, "Multimodality diagnosis of liver tumors: Feature analysis with CT, Liver-specific and Contrast-enhanced MR, and a computer model," *Academic Radiology*, vol. 9, pp. 256-259, 2002.

[6] R. M. Haralick, L. G. Shaphiro, *Computer and Robot Vision*, vol. I. Addison-Wesley Publishing Company, 1992.

[7] R. M. Haralick, K. Shanmugan, I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, no. 6, pp. 610-622, 1973.

[8] J. S. Weszka, C. R. Dryer, A. Rosenfeld, "A comparative study of texture measures for terrain classification," *IEEE Transaction on System, Man, and Cybernetic*, vol. SMC-6, pp. 269-285, 1976.

[9] K. I. Laws, "Rapid Texture Identification," Proceedings of the *SPIE Conference for Missile Guidance*, vol. 238, pp. 376-380, 1980.

[10] C.-M. Wu, Y.-C. Chen, "Texture features for classification of ultrasonic liver images," *IEEE Trans. Med. Imag.*, vol. 11, no. 2, pp. 141-152, 1992.

[11] D. Goldberg, *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley, 1989.

[12] S. Haykin, *Neural networks: A comprehensive foundation*, Prentice-Hall, 1999.

[13] F. Provost, P. Domingos, 2001. "Well-trained PETs: Improving probability estimation trees," *CeDER Working Paper #IS-00-04,* Stern School of Business, New York University, NY 10012, 2001.