Computer Aided Diagnosis of CT Focal Liver Lesions based on Texture Features, Feature Selection and Ensembles of Classifiers

Stavroula G. Mougiakakou[1], Ioannis K. Valavanis[1], Alexandra Nikita[2], Konstantina S. Nikita[1]

[1] National Technical University of Athens, Faculty of Electrical and Computer Engineering, 9 Iroon Polytechniou Str., 15780 Zografou, Athens, Greece
{smougia, knikita}@cc.ece.ntua.gr, ivalavan@biosim.ntua.gr
[2] University of Athens, Medical School, Department of Radiology, 20 Papadiamantopoulou Str., 15228 Athens, Greece
anikita@cc.uoa.gr

Abstract A computer aided diagnosis system aiming to classify liver tissue from computed tomography images is presented. For each region of interest five distinct sets of texture features were extracted. Two different ensembles of classifiers were constructed and compared. The first one consists of five Neural Networks (NNs), each using as input either one of the computed texture feature sets or its reduced version after feature selection. The second ensemble of classifiers was generated by combining five different type of primary classifiers, two NNs, and three *k*-nearest neighbor classifiers. The primary classifiers of the second ensemble used identical input vectors, which resulted from the combination of the five texture feature sets, either directly or after proper feature selection. The decision of each ensemble of classifiers was extracted by applying voting schemes.

1 Introduction

Computer-Aided Diagnosis (CAD) systems, for the characterization of liver tissue, attract more and more attention, in order to assist clinicians in diagnosis, and reduce the number of required biopsies. Various approaches, most of them using ultrasound B-scan and Computed Tomography (CT) images, have been proposed based on different image characteristics, such as texture features, and fractal dimension estimators combined with various classifiers [1], [2], [3]. Texture analysis of liver

Please use the following format when citing this chapter:
Mougiakakou, Stavroula, Valavanis, Ioannis, Nikita, Alexandra, Nikita, Konstantina, 2006, in IFIP International Federation for Information Processing, Volume204, Artificial Intelligence Applications and Innovations, eds. Maglogiannis, I., Karpouzis, K., Bramer, M., (Boston: Springer), pp. 705–712

CT images based on Spatial Gray Level Dependence Matrix (SGLDM), Gray Level Run Length Method (GLRLM), and Gray Level Difference Method (GLDM) has been proposed in [4], in order to discriminate normal from malignant hepatic tissue. Texture features from SGLDM have been applied to a Probabilistic Neural Network (P-NN) in [5] for the characterization of hepatic tissue (hepatoma and hemangioma) from CT images. Additionally, SGLDM based texture features fed to a system of three sequentially placed Neural Networks (NNs) have been used in [6] for the classification of hepatic tissue into four categories.

The principal aim of the present paper is to assess the potential of ensembles of classifiers in the development of a CAD system able to discriminate four hepatic tissue types: normal liver (C1), hepatic cyst (C2), hemangioma (C3), and hepatocellular carcinoma (C4) from CT images.

2 Methodology

The generic design of a CAD system is presented in Fig. 1. Regions of Interest (ROIs) drawn by an experienced radiologist on CT images were driven to a feature extraction module, where five different texture feature sets were obtained. The full feature sets or their reduced versions obtained after proper feature selection in the feature selection module, were fed to two alternative ensembles of classifiers (EC). The primary classifiers of the first ensemble (EC1) were generated by applying a single learning algorithm to different data sets, while the classifiers of the second ensemble (EC2) were generated by using different learning algorithms on the same data set. The predictions of the primary classifiers of each ensemble were combined using appropriate voting schemes.

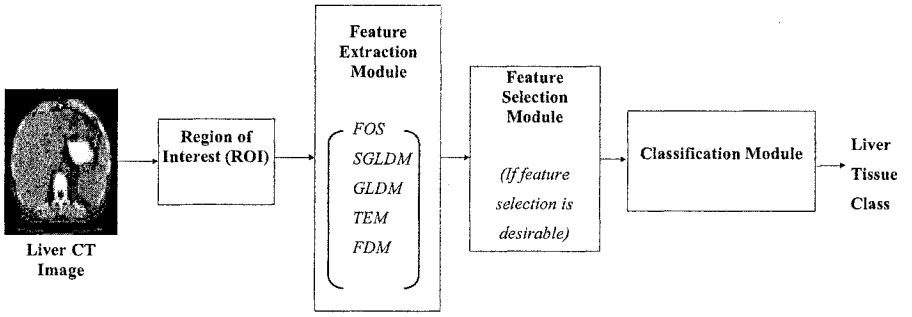

Fig. 1. Generic design of CAD1 ... CAD5

2.1 Image Acquisition

Abdominal non-enhanced CT images with a spatial resolution of 512×512 pixels and 8-bit gray-level at the W150+60 window taken from both patients and healthy controls were used. The diagnosed hepatic lesions from patients with C2, C3, and C4, were validated by needle biopsies, density measurements, and the typical pattern of enhancement after the intravenous injection of iodine contrast. The position, size

and extent of the lesions were defined in CT images by an experienced radiologist. A total of 147 free-hand ROIs were sampled and distributed into three disjoint data sets (training, validation, and testing), as presented in Table 1.

2.2 Feature Extraction

In the feature extraction module, five sets of features were calculated for each ROI.

Table 1. Distribution of the available samples in training, validation, and testing set

	Training Set	Validation Set	Testing Set
Normal (C1)	42	17	17
Cyst (C2)	11	4	4
Hemangioma (C3)	16	6	6
Hepatocellular Carcinoma (C4)	14	5	5

2.2.1 First Order Statistics

Features from FOS [7] are easily computed from the intensity function of the image. In our study, six features were calculated for each ROI: average gray level (avg_{FOS}), standard deviation (std_{FOS}), entropy (ent_{FOS}), coefficient of variation (cv_{FOS}), skewness (sk_{FOS}), kurtness (kur_{FOS}).

2.2.2 Spatial Gray-Level Dependence Matrices

Texture characteristics can be derived from SGLDM of the ROI [7], [8]. The features calculated in our experiments are: angular second moment (asm_{SGLDM}), contrast (con_{SGLDM}), correlation (cor_{SGLDM}), variance (var_{SGLDM}), inverse difference moment (idm_{SGLDM}), entropy (ent_{SGLDM}), homogeneity (hg_{SGLDM}), cluster tendency (clt_{SGLDM}). The features were calculated for intersample spacing of 1, 2, 4, 6, 8, and 12 pixels. For each value of intersample spacing, the feature values are computed by averaging over four uniformly distributed angular directions, 0°, 45°, 90°, and 135°. Thus, a total of 48 texture characteristics were obtained through SGLDM for each ROI.

2.2.3 Gray-Level Difference Matrix

Application of the GLDM to each ROI results in a 20-dimensional feature vector. Five texture features, were extracted based on the gray level difference density function [9]: contrast (con_{GLDM}), mean value (mn_{GLDM}), entropy (ent_{GLDM}), inverse difference moment (idm_{GLDM}), angular second moment (asm_{GLDM}). These features were calculated for distances of 1, 2, 3, and 4 pixels. The final feature value for each distance was computed by averaging over the feature values corresponding to the four angular directions.

2.2.4 Laws' Texture Energy Measures

Laws' TEM are derived from three simple vectors of length three [10]. In our study, the following four Laws' zero-sum masks were used: $L5E5 = L5^{T}E5$, $E5S5 = -E5^{T}S5$, $L5S5 = L5^{T}S5$, $R5R5 = R5^{T}S5$. After convolving each ROI image with each of the four masks, the following measures were calculated: Sum of absolute values/# of pixels (as_{TEM}), *sum of squares/# of pixels* (ss_{TEM}), entropy (ent_{TEM}). Thus, twelve

Laws' energy measures (4 masks × 3 statistics per mask) are calculated for each ROI.

2.2.5 Fractal Dimension Texture Measurements

For each ROI, a 3-dimensional feature vector is estimated from the FDM. The components of the feature vector correspond to the parameters $H1_{FDM}$, $H2_{FDM}$, $H3_{FDM}$, of the multiresolution fractal feature vector [3].

2.3 Feature Selection

For the purpose of feature selection, a GA based on [11], was used in the present paper. The algorithm makes use of a randomly created initial population of N chromosomes. Each chromosome is a binary mask, with 1 indicating that the feature is selected, and 0 that the corresponding feature is omitted. The chromosomes are mated, with possibility depending on their fitness, in order to propagate their genetic material to their offspring. $N/2$ pairs of chromosomes are selected using the elitist selection method. The selected chromosomes are mated using the two-point crossover (crossover probability, P_c). The mutation genetic operator (mutation probability, P_m) is applied in order to switch the value of chromosome bits. Thus, a new chromosome is produced replacing the old one. The fitness function is estimated for the new population, and the best results are stored. The procedure is repeated for a number of N_G generations. The maximum squared Mahalanobis distance was used as fitness function [6]. Since the number of selected features is not taken into account in computing the fitness function, a "penalty" function for feature sets exceeding a given dimensionality threshold was applied. Thus, the corresponding individuals were assigned a fitness value equal to 50% of the average population fitness. The GA was run for a dimensionality threshold equal to ten [6]. The GA parameters were: $N = 200$, $N_G = 250$, $P_c = 0.8$, and $P_m = 0.008$.

2.4 Classification

The estimated texture features sets were applied to either of two different ensembles (EC1 and EC2) of classifiers. EC1 was constructed by combining five Multilayer Perceptron NNs (MLP-NN), each trained with one out of the five distinct texture feature sets, while EC2 was constructed by combining one MLP-NN, one Probabilistic NN (P-NN), and three Nearest Neighbor (k-NN) classifiers, each trained with the combination of the five computed texture feature sets. For each ensemble of classifiers the final decision was generated by combining the outputs of the corresponding primary classifiers through appropriate voting schemes.

2.4.1 Multilayer Perceptron Neural Network

The MLP-NN classifier [12] used in this study is based on a feed-forward NN consisting of one input layer with a number of input neurons equal to the number of features fed into the NN, one hidden layer with variable number of neurons, and one

output layer consisting of two output neurons, encoding the different types of liver tissue (00=C1, 01=C2, 10=C3, and 11=C4). The MLP-NN was trained, using the training set, by the batched Back-Propagation (BP) algorithm with adaptive learning rate and momentum [12]. Moreover, the optimal number of hidden neurons, as well as the appropriate values of momentum and initial learning rate were estimated using a trial-and-error process, until no further improvement of classification accuracy in the validation set could be obtained.

2.4.2 Probabilistic Neural Network

The P-NN performs interpolation in multidimensional space [1]. The P-NN consists of one input layer, with number of neurons equal to the number of used features, a hidden layer, a summation unit layer, and an output layer. In order to classify a ROI the corresponding feature set is applied to the input layer and then into the hidden layer, followed by the summation layer. Finally, the neuron in the output layer classifies the ROI into the class with the highest probabilistic density function. The applied training procedure is the same as in the case of the MLP-NN classifier.

2.4.3 k-Nearest Neighbor Classifier

The k-NN classifier identifies the k nearest neighbors to the feature vector to be classified from the training set based on a distance measurement of the vectors. The feature vector is classified to the most frequent class occurring in the set of neighbors [13]. In this paper, a *1*-NN classifier along with two mk-NN (modified k-NN, $k>1$, classifiers), mk_1-NN ($2 \le k_1 \le 5$) and mk_2-NN ($6 \le k_2 \le 9$), have been developed [13]. The mk-NN classifiers differs from the k-NN, in that the classification result is based both on the frequencies of the classes occurring in the set of neighbors, and on the distances of the neighbors from the feature vector.

2.4.4 Voting Scheme

A plurality and a weighted voting scheme were used to combine the predictions of the primary classifiers of EC1 and EC2 [14]. According to the plurality voting scheme, each primary classifier gives a vote for its prediction. The prediction receiving the most votes is the final prediction. According to the weighted voting scheme, the primary classifier predictions are averaged by taking into consideration a set of weights, with which the classifiers participate in the final prediction.

2.5 CAD System Architectures

Five alternative architectures (CAD1, ..., CAD5) were developed based on the generic design of the CAD system presented in Fig. 1. CAD1 and CAD2 were constructed using EC1, while CAD3, CAD4, and CAD5 were based on EC2. In CAD1 (Fig. 3(a)), each of the full-dimensional FOS, SGLDM, GLDM, TEM, and FDM feature sets, estimated in the feature extraction module, is fed into one of the five primary classifiers of EC1. CAD2 (Fig. 3(b)) differs from CAD1 in that feature selection is applied to the feature vectors estimated from SGLDM, GLDM, and TEM, since they have high dimensionality. In CAD3 (Fig. 3(c)), each primary classifier uses as input the 89-dimensional feature set, which results from the

combination of the full-dimensional FOS, SGLDM, GLDM, TEM, and FDM features sets. CAD4 (Fig. 3(d)) differs from CAD3 in that feature selection is applied

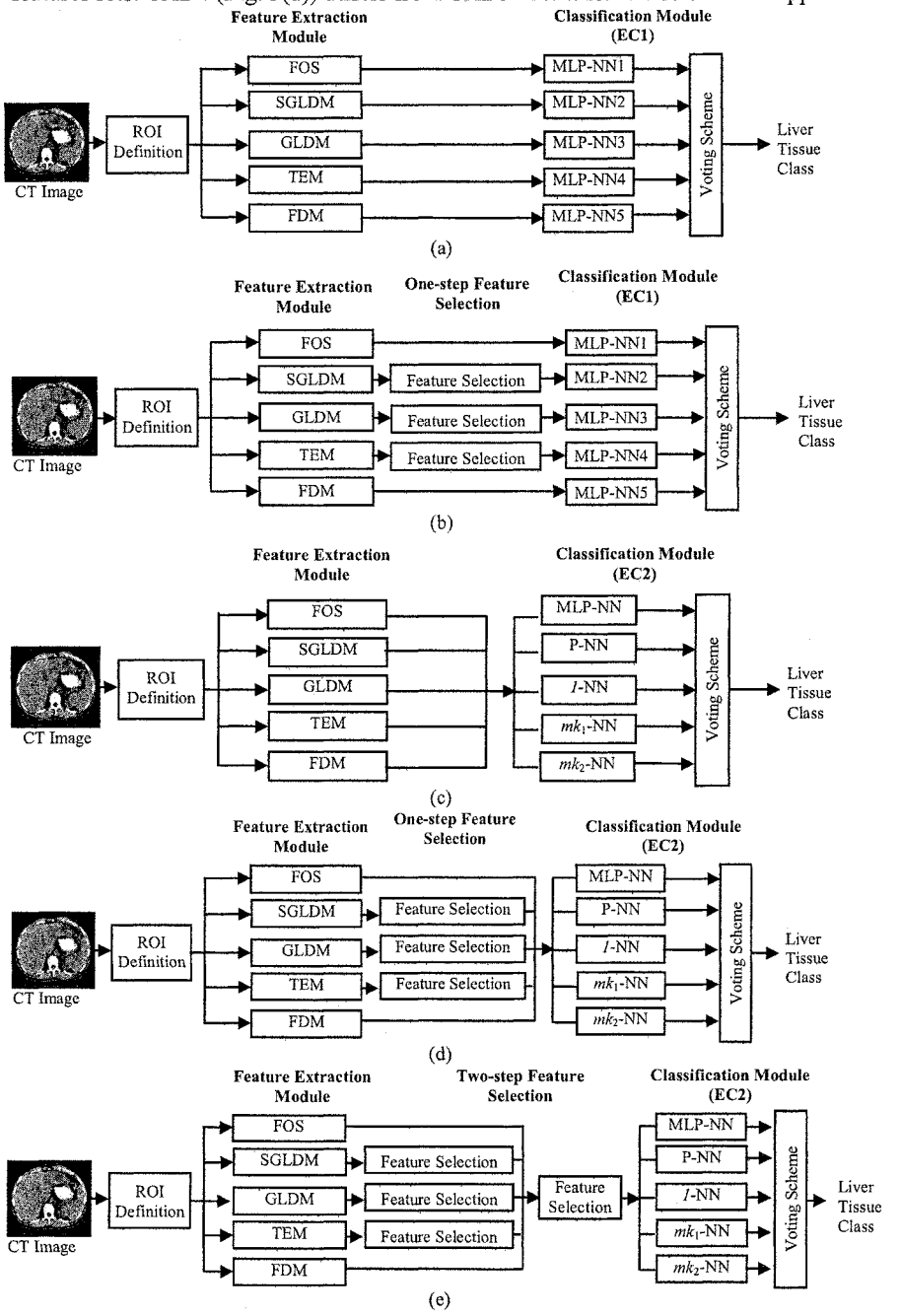

Fig. 2. Architecture of (a) CAD1, (b) CAD2, (c) CAD3, (d) CAD4, and (e) CAD5

to SGLDM, GLDM, and TEM feature sets prior to the combination with the full-dimensional FOS and FDM feature sets. The selected SGLDM, GLDM and TEM features are identical with the ones applied in CAD2. CAD5 (Fig. 3(e)) differs from CAD4 in that further feature selection is applied to the 30-dimensional feature set used by CAD4. The resulting 12-dimensional feature set provides input to each primary classifier of EC2 (CAD5). A plurality or weighted voting scheme extracts the final decision for each of EC1 and EC2.

3 Results and Discussion

In order to find the best performing CAD, the achieved classification rates of CAD1, ..., CAD5 in the testing set were comparatively assessed, along with their behavior to all the available datasets. All classification performances of the base classifiers and the ensembles of EC1 (CAD1 and CAD2) and EC2 (CAD3,..,CAD5) are presented in Tables 2 and 3, respectively. Based on Tables 2 and 3, it is observed that the primary classifiers of EC2 in CAD3, CAD4, and CAD5 classify better the liver regions, compared to the primary classifiers of EC1 in CAD1, and CAD2. CAD2, which uses EC1, and is the best of CAD1 and CAD2, achieved a classification performance in the testing set equal to 90.63%. This performance is slightly less than the best performance (93.75%) of CAD3, CAD4 and CAD5 which use EC2 and was achieved by CAD5. Furthermore, CAD2 with weighted voting scheme performs quite balanced in the datasets, while CAD5 had a quite unbalanced behavior. Thus the best architecture is CAD2 with weighted voting scheme.

Table 2. The individual and total classification performances of EC1 (CAD1, CAD2)

Classifier	Validation Set (%)		Testing Set (%)	
	CAD1	CAD2	CAD1	CAD2
MLP-NN1	90.63	(90.63)	87.50	(87.50)
MLP-NN2	65.63	71.88	62.50	56.25
MLP-NN3	65.63	65.63	53.13	43.75
MLP-NN4	87.50	84.38	81.25	90.63
MLP-NN5	65.63	(65.63)	59.38	(59.38)
Classifier Combination				
Plurality Voting Scheme	90.63	93.75	78.13	78.13
Weighted Voting Scheme	93.75	93.75	87.50	90.63

Table 3. The individual and total classification performances of EC2 (CAD3, CAD4, CAD5)

Classifier	Validation Set (%)			Testing Set (%)		
	CAD3	CAD4	CAD5	CAD3	CAD4	CAD5
MLP-NN	81.25	90.63	71.88	81.25	90.63	81.25
P-NN	93.75	96.88	90.63	81.25	81.25	90.63
1-NN	90.63	96.88	78.13	68.75	75.00	78.13
mk_1-NN	90.63	93.75	78.13	84.38	87.50	93.75
mk_2-NN	87.5	90.63	84.38	84.38	87.50	90.63
Classifier Combination						

Plurality Voting Scheme	90.63	96.88	81.25	84.38	84.38	93.75
Weighted Voting Scheme	90.63	96.88	84.38	84.38	84.38	93.75

In order to evaluate the classification ability of the proposed CAD architecture, to overcome problems encountered while assessing the various architectures, like the unusually better performances in the testing set compared to these on the validation set, and the fact that the ensembles did not outperformed all of the classifiers, it is under investigation the use of a larger image database, and more elaborate sampling schemes. Furthermore, the proposed system can be extended to other classes of liver lesions, and/or to liver images from other imaging devices.

4 Conclusion

The development of a CAD system aiming to discriminate four hepatic tissue types from non-enhanced CT images has been presented. An ensemble of classifiers has been constructed based on NN leading to a classification performance of 90.63%.

References

1. Kadah YM, Frag, AA, Zurada JM, Badawi AM, Youssef A-B M, Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images, *IEEE Trans. Med. Imag.*, vol. 15, no. 4, (1996) 466-478.
2. Sun YN, Horng MH, Lin XZ, Wang JY, Ultrasonic image analysis for liver diagnosis, *IEEE Engin. Med. Biol.*, vol. 11-12, (1996) 93-101.
3. Wu Ch-M, Chen Y-Ch, Sheng Hsieh K, Texture features for classification of ultrasonic liver images, *IEEE Trans. Med. Imaging*, vol. 11, no. 2, (1992) 141-151.
4. Mir AH, Hanmandlu M, Tandon SN, Texture analysis of CT images, *IEEE Eng. Med. Biol. Mag.*, vol. 14, no. 6, (1995) 781-786.
5. Chen EL, Chung P-C, Chen CL, Tsa HM, Chang CI, An automatic diagnostic system for CT liver image classification, *IEEE Trans. Biomed. Eng.*, vol. 45, no. 6, (1998) 783-794.
6. Gletsos M, Mougiakakou SG, Matsopoulos GK, Nikita KS, Nikita A, Kelekis D, 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, no. 3, (2003) 153-162.
7. Haralick RM, Shaphiro LG, *Computer and Robot Vision*, vol. I. Addison-Wesley 1992.
8. Haralick RM, Shanmugan K, Dinstein I, Textural features for image classification, *IEEE Trans. Systems, Man, and Cybernetics*, vol. 3, no. 6, (1973) 610-622.
9. Weszka JS, Dryer CR, Rosenfeld A, A comparative study of texture measures for terrain classification, *IEEE Trans. System, Man, and Cybernetic*, vol. SMC-6, (1976) 269-285.
10. Laws KI, Rapid texture identification, *Proc. of the SPIE Conference for Missile Guidance*, vol. 238, (1980) 376-380.
11. Goldberg D, *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley, (1989).
12. Haykin S, *Neural networks: A comprehensive foundation*, Prentice-Hall, (1999).
13. Wu Y, Ianakiev Kr, Govindaraju V, Improved k-nearest neighbor classification, *Pat. Recogn.*, vol. 35, (2002) 2311-2318.
14. Lam L, Suen Ch, Application of majority voting to pattern recognition: An analysis of its behavior and performance, *IEEE Trans. Systems, Man, and Cybernetics*, vol. 27, no. 5 (1997) 553-568.