The employment of textural and non textural image analysis algorithms in assessing the diffuse liver diseases

Sergiu Nedevschi, Cristian Vicas
Department of Computer-Science
Technical University of
Cluj-Napoca, Romania
sergiu.nedevschi@cs.utcluj.ro

Monica Lupsor, Radu Badea, Mircea Grigorescu
Medical Clinic no. 3
University of Medicine and Pharmacy
Cluj-Napoca, Romania

Abstract

This paper is focused on current progress of our research in improving diagnosis value of ultrasound imaging in the context of diffuse liver diseases. Image features are computed on ultrasound images and these features are used to train a classifier. The classifier is able to distinguish between various pathology grades. Present study shows that, based on ultrasound images, steatosis can be accurately graded and a qualitative assessment can be made in case of fibrosis. Further improvements can be made if we include more patients and consider non-imagistic features like clinical and biochemical analysis of the patient.

1. Introduction

There are two main diseases studied: nonalcoholic steatohepatitis and chronic hepatitis. Both diseases alter the liver by producing fatty infiltration (steatosis), fibrosis, inflammatory and necrosis processes. The way that these processes alter the liver varies from steatohepatitis to chronic hepatitis. For this reason, physicians use two scores in order to evaluate the liver biopsy [1]. First score is BRUNT and is used for steatohepatitis and the second is METAVIR and is used for chronic hepatitis. The aim of this research is to use noninvasive techniques (US imaging) to assess the BRUNT/METAVIR score or, at least, some components of these scores. Fibrosis is scored from 0 (healthy) to 4 (cirrhosis). Steatosis is graded from 0 (healthy) to 3 (>66% steatosis) Activity is graded also from 0 to 3.

We have four main patient groups. A clinically healthy group, steatohepatitis patients, chronic hepatitis patients (both groups have been diagnosed using liver biopsy) and a cirrhotic group which was only clinically diagnosed.

At liver tissue there are a number of pathological processes. Each of these processes could alter the aspect of the US images. As we shown in our previously work [2][3][4][5], steatosis has the biggest impact, followed by fibrosis and activity. Also we have shown that these processes can mask each other, i.e. strong fibrosis can mask a low grade steatosis and vice-versa [3][4]. The activity can be detected only if we consider two patient lots, having the same steatosis and fibrosis grade but different activity grade. If the steatosis grade or fibrosis stage varies, the activity cannot be detected.

Another interesting aspect discovered in our previous work [2][4] is that the fibrosis detection is better if we select only patients with low steatosis. The opposite is also true, detection of steatosis is better if the fibrosis stage is kept low.

The paper is structured as follows: Section 2 describes the methodology, section 3 focuses on steatohepatitis disease and section 4 focuses on fibrosis stage detection at chronic hepatitis. Section 5 concludes the paper.

2. Methodology

The methodology proposed in our research consists in five main steps: image acquisition, the establishment of region of interest, image attributes computation, feature selection and classification.

The algorithms used to generate image features are shown in Table 1, along with the number of features generated by each algorithm. Image acquisition, ROI establishment and the attribute computation algorithms are widely explained in [2][3][4].

Table 1. Algorithms used to generate image features and the number of features generated by each algorithm.

Algorithm	Number of features generated
Histogram statistics	6
Grey tone difference matrix	25
Grey level co-occurence matrix	200
Multifractal differential box counting	90
Morphological multifractal exponents	464
Multiresolution fractal dimension	3
Law's texture energy measures	56
Wavelet transform	255
Attenuation and backscattering coefficients	12

2.1. Feature selection algorithms

There are two algorithms used here, the correlation-based feature selection and chi square feature selection.

2.2. Correlation-based feature selection

In the center of this algorithm is a heuristic that evaluates a feature subset. It is considered the utility of an attribute to predict the class along with the correlation between attributes [6]. It is desired to have a set of features with high correlation with respect to class but with low inter-correlation.

The correlation between two attributes is measured using symmetric uncertainty (see equation 1 – 4)

$$U(A,B) = 2\frac{H(B) - H(B \mid A)}{H(A) + H(B)}$$

Equation 1. The evaluation of two attributes A şi B. H(A) is the entropy of A.

$$H(Y) = -\sum_{y \in Y} p(y) \log p(y)$$

Equation 2. Entropy formula for one attribute

$$H(Y \mid X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y \mid x) \log p(y \mid x)$$

Equation 3. Entropy formula for two dependent attributes.

$$Perf = \frac{\sum_j U(A_j, C)}{\sqrt{\sum_i \sum_j U(A_i, A_j)}}$$

Equation 4. Performance evaluation of an attribute subset with respect to a class C.

CFS technique allows us to evaluate a subset of features using Equation 4. One needs a method of searching through all possible combination of feature sets and retrieve the one having best performance.

We used the "best first" method proposed in [6]. The algorithm starts by building and evaluating all feature sets of length 1. The best feature set is selected. This subset is extended by adding one new attribute. The algorithm generates all possible feature sets and evaluates them. Again, best subset is selected, and the process continues until a new feature subset does not improve the results. A backtracking mechanism is used and the algorithm selects for extending the next best set found at previous step. In this manner, one can explore the entire subset space. However, the algorithm is stopped after k unsuccessful extensions. A value of 5 for k is used in our experiments.

One should note that the correlation based feature selection algorithm does not provide a ranking for attributes.

Using different datasets one can find that the selected attributes varies greatly. This happens when there are highly correlated features. The algorithm might choose one feature or another depending on the actual data.

2.3. Chi square feature selection

This method evaluates each attribute with respect to the class. The evaluation is performed using chi square statistical test (Equation 5).

$$\chi^2 = n \sum_{i,j} \frac{\left(N_{i,j} - N_i N_j\right)^2}{N_i N_j}$$

Equation 5. χ2 test. n is the number of instances, Ni is the observed number of events i, and Nij denotes the number of times that events i and j co-occur.

The attributes are sorted according to the calculated correlation grade and selected based on a threshold [6].

One should note that this algorithm provides us with a ranking of the attributes regardless of their correlation with each other.

2.4. Classification algorithm

Support Vector Machines (SVM) or kernel machines are a family of learning methods that can represent complex, nonlinear functions [6].

The basic idea for this algorithm is to map a lower dimension feature space where the classes are not linearly separable into a higher dimensional feature space where these classes are linearly separable.

2.5. Performance evaluation

The SVM classifier is evaluated using Cohen's Kappa statistic [7]. This allows us to globally assess the classification performance regardless of class numbers or class volumes.

Another performance indicator is the Area under Curve because it is widely used in medical research area.

Cross validation (10 fold) is used to assess the real performances of the trained classifiers. The feature selection algorithms (when they were used) are applied into the cross validation loop. The cross validation is run 5 times and the results are averaged.

Also, a parameter search for SVM was performed using another cross-validation loop.

3. Steatosis detection at patients with steatohepatitis

For steatosis detection we used healthy patients and affected (steatohepatitis) patients having fatty infiltration ≥ 1. First group has 22 patients and the second has 91 patients. As image features we used attenuation and backscattering coefficients.

60 parameter sets for SVM algorithm were tested using cross validation. All 12 computed features are used. The results are shown in Table 2.

Table 2. Fatty infiltration detection. Kappa and AUROC values obtained when comparing healthy and ill patients.

Computed statistic	Value
Kappa value	0.71
AUROC for healthy group	0.93
AUROC for ill patients	0.93

We investigated the discrimination rates between steatosis grades. The patients were divided in groups having steatosis grade 0 (healthy patients) 1,2 and grade 3. Kappa and AUC values were computed as described above. The results are presented in Table 3.

Table 3. Fatty infiltration grade detection. The patients were divided in three classes: healthy, fatty grade 1,2 and 3.

Computed statistic	Value
Kappa statistic	0.47
AUC for healthy lot	0.934
AUC for steatosis grade 1	0.709
AUC for steatosis grade 2	0.712
AUC for steatosis grade 3	0.85

From the previous patient lot we've selected only the patients having low fibrosis stage. The results are presented in Table 4.

It is noted a marginally improvement in detection when we select only patients having low fibrosis stage.

The most relevant features were the slope, offset and R^2.

The ROI establishment process, performed manually by the physician is operator dependent and very time consuming. However, an automatic ROI generation might not be able to capture the relevant aspects of steatosis. We studied the variability of the image features on the entire lot, on a class and on a ultrasound image.

Table 4. Fatty infiltration grade detection. The patients were divided in three classes, healthy, fatty grade 1,2 and 3 but selecting only low (0 or 1) fibrosis stage.

Computed statistic	Value
Kappa statistic	0.51
AUC for healthy lot	0.921
AUC for steatosis grade 1	0.793
AUC for steatosis grade 2	0.732
AUC for steatosis grade 3	0.931

We propose the following methodology: A simple algorithm proposes up to 5 ROI per image. A physician selects only the proper ROI's [2][4][8]. The image features are computed for each ROI. Next we compute the standard deviation (SD) of each feature with respect to each class (Table 5). The computed standard deviation for all the patients is larger than the SD for each class. This is an expected result because using these features we've obtained excellent steatosis detection. In Table 6 is computed the SD for each

ultrasound image. Then, the mean of these SD is computed according to patient's pathology. The obtained values are relatively close. These results show a low variability of the studied features when considering various regions of the same image and a high variability when considering different pathology classes.

Table 5. The SD for the relevant features computed on each studied patient lot.

Feature	All patients	Healthy	S1 lot	S2 lot	S3 lot
Slope	0.106	0.080	0.079	0.072	0.058
Offset	29.01	17.60	22.306	23.180	21.29
R^2	0.327	0.189	0.303	0.279	0.181

Table 6. The mean of the SD for the relevant features.

Feature	All patients	Healthy	S1 lot	S2 lot	S3 lot
Slope	0.050	0.060	0.052	0.048	0.044
Offset	15.25	13.79	15.589	15.66	15.86
R^2	0.173	0.166	0.185	0.173	0.147

For each ultrasound image was computed the mean of each feature. The resulting values were used to train and evaluate a classifier. In Table 7 are the obtained results. When comparing these figures with those in Table 3 one can see that they are almost identical.

Table 7. Fatty infiltration grade detection. The feature vector was built by calculating the average of each individual feature over the ROI's image.

Computed statistic	Value
Kappa statistic	0.48
AUC for healthy lot	0.91
AUC for steatosis grade 1	0.75
AUC for steatosis grade 2	0.77
AUC for steatosis grade 3	0.9

This indicates that an automatic ROI selection algorithm could replace the manual procedure, reducing operator dependent factors, reducing the workload and keeping the discrimination performance to the same levels.

4. Fibrosis stage detection

In order to determine the fibrosis stage, and to keep a relatively high patient count for each group we divided the patients in three classes, N, F23, F4. (Table 8) Preliminary results sown that steatosis can mask an existing fibrosis pathology [3][9]. As a result we produced a new selection holding only the patients with steatosis grade 0 or 1. (Table 8)

Table 8. The division of the patient lot in classes according to fibrosis stage.

Class	Inclusion Criteria	Number of Patients		Number of US Images from right lobe at 16 cm	
		All fatty grades	Fatty grade 0 or 1	All fatty grades	Fatty grade 0 or 1
N	Control patients along with fibrosis stage 1	211	168	2730	2228
F23	Fibrosis stage 2 and 3 patients	258	196	3225	2461
F4	Cirrhotic patients	81	81	1013	1013

One can notice that the volume of the data is reduced when selecting only patients with low steatosis grade by almost 25%.

1100 feature were computed using textural algorithms shown in Table 1. A parameter search was performed for each classifier, and the feature selection algorithm was applied into the cross-validation loop.

The results for the lot where all the patients are selected are presented in table.

Table 9 Discrimination rates for fibrosis stage when considering three class divisions and including all the patients, regardless of steatosis grade.

Computed statistic	All features	Chi Square	CFS
Kappa statistic	0.161	0.42	0.257
AUC for N class	0.61	0.77	0.68
AUC for F23 class	0.62	0.78	0.67
AUC for F4 class	0.63	0.82	0.728

The results in [3][4] indicated that a high fatty infiltration can lower the detection of steatosis. Using the same grouping as in table 9 we selected only the patients having low steatosis grade. The results are presented in Table 10.

From Table 9 one can see that using Chi square feature selection algorithm the kappa value is 0.42. From Table 10, the best kappa is 0.31 and is obtained

using CFS feature selection algorithm. This behavior is somehow opposed to our previous findings.

Even high fatty infiltration could mask the fibrosis, the increased number of samples in first data set (Table 9) gave us better results than those when we considered fewer patients and low fatty (Table 10).

The results shown in both tables are better than those obtained in our previous experiments. The main reason is the increased data volume.

We can conclude that it is required to include more patients into the study of fibrosis and chronic hepatitis.

Table 10. Discrimination rates for fibrosis stage when considering three class division and including patients with low steatosis grade.

Computed statistic	All features	Chi Square	CFS
Kappa statistic	0.172	0.23	0.31
AUC for N class	0.69	0.691	0.73
AUC for F23 class	0.66	0.693	0.72
AUC for F4 class	0.68	0.73	0.75

Table 11. Discrimination rates for normal/ill comparison when considering all the patients.

Computed statistic	All features	Chi Square	CFS
Kappa statistic	0.15	0.31	0.25
AUC for N class	0.6	0.71	0.67
AUC for F234 class	0.6	0.71	0.67

Table 12. Discrimination rates for normal/ill comparison when considering patients with low steatosis grade.

Computed statistic	All features	Chi Square	CFS
Kappa statistic	0.26	0.35	0.38
AUC for N class	0.68	0.77	0.79
AUC for F234 class	0.68	0.77	0.79

In order to differentiate healthy from ill (chronic hepatitis) patients we constructed two classes:

- N class, having control patients and fibrosis 1 patients

- F234 class having the rest of the chronic hepatitis and cirrhotic patients.

We also constructed a group where only low steatosis grade patients were selected. The results are shown in Table 11 and 12.

This time the expected behavior was noted. Selecting only low steatosis grade patients resulted into an increased detection rates. Grouping all the patients into only two classes had increased the data volume.

4.1. Activity grade detection

The initial results show that activity could be detected if we eliminate the effect of fibrosis and steatosis [10]. However, after we trained a classifier with various activity grades (form normal to A3) the results were disappointing. In every training case, all the samples were classified into one class.

4.2. Feature selection

From the tables 9 to 12 one can see that the performance of the classifier is improved when is used a feature selection schema.

CFS and Chi Square feature selection schemas were applied on the entire data set in order to investigate the relevant features. The features selected by the CFS algorithm were provided by Grey level co-occurrence matrix, grey tone difference matrix, histogram statistic, wavelet and multirezolution fractal dimension. The Chi Square algorithm evaluates each feature and can provide a ranking. The best features were multirezolution fractal dimension, followed by GLCM-correlation at several directions, wavelet, standard deviation computed on histogram and contrast computed from grey tone difference matrix algorithm.

5. Conclusions

Attenuation, backscattering and correlation coeficients can be successfuly employed to distinguish between normal liver and moderate/severe steatohepatitis. They allow us to discriminate between various steatosis grades even when there are overlapping pathologies.

Textural coefficients are able to capture the alterations produced by fibrosis. Using a subset of these coefficients it is possible to train a classifier that will be able to distinguish relatively healthy patients from moderate and severe affected patients.

As the study progressed it was clear that increased number of patients improve the results. Although it

was shown that severe steatosis can sometimes mask a low stage fibrosis, in present paper we've obtained better results when we considered all the patients, including those with high fatty infiltration. By including all the patients the volume of data is increased by almost 25% which leads to better trained classifiers.

A feature selection algorithm is very useful. Both studied algorithms improved the results. No clear distinction could be made in terms of improved classification rates between CFS and Chi Square algorithms. The best algorithm depends on the studied problem. One can note that CFS feature selection algorithm performed better when dealing with datasets where only patients having steatosis grade 0 and 1 were selected. Chi square feature selection algorithm performed better when dealing with datasets that were constructed regardless of the patient steatosis grade.

The methodology presented here might be positively biased by the fact that we used several ROI extracted from each patient. It is possible that the classification algorithm learns the specifics of a patient and not the characteristics of the disease.

The results obtained by averaging the attribute values from each ROI indicates that is possible to compute one feature vector for a patient, feature vector that is computed based on all the available ultrasound images.

These results were tested only on attenuation coefficients and not on textural attributes. It might be possible that the average should not be the best approach when dealing with textures. Some texture processing algorithms can be easily altered in such a way that multiple ROI's can be considered. For example, it's easy to build a co-occurrence matrix using pixels from two or more ROI. Unfortunately this approach cannot be easily applied to other algorithms (i.e. fractal dimension or wavelet)

The inclusion of other parameters (like clinical or biochemical analysis) as features might increase the detection performance obtained using only image processing features. These parameters are obtained during routine patient investigation by the physician.

In case of steatohepatitis disease, the fatty infiltration can be accurately graded, but, in case of chronic hepatitis only a qualitative assessment of the fibrosis can be made.

More patients have to be investigated in order to improve detection rates of the fibrosis.

6. Acknowledgement

This work was sponsorised by the Romanian National Authority for Scientific Research as part of the CEEX program FINALISM 94/31.07.2006 and SIDEF 71/17.07.2006.

7. References

[1] T. Poynard, V. Ratziu, P. Bedossa. "Appropriateness of liver biopsy." Can J Gastroenterol, vol. 14, nr. 6, pp. 543-548, 2000

[2] S. Nedevschi, C. Vicas, D. Mitrea M. Lupsor, M. Grigorescu, R. Badea, "Usefulness Of Attenuation And Backscattering Coefficients In Investigating Complex Nonalcoholic Steatohepatitis From Ultrasound Images. Preliminary Results", Proceedings of 19'th Biosignal Conference, 29-30 jun. 2008, Brno , paper ID 126

[3] C. Vicas, S. Nedevschi, M. Lupsor, R. Badea, H. Stefanescu "Fibrosis detection from ultrasound imaging. The influence of necro-inflammatory activity and steatosis over the detection rates." *Journal of Automation, Computers, Applied Mathematics* vol. 16, no. 3, pp 27-33, 2007

[4] C. Vicas, S. Nedevschi, M. Lupsor, R. Badea, M. Grigorescu "Steatohepatitis Detection from Ultrasound Images Using Attenuation and Backscattering Coefficients" *Journal of Automation, Computers, Applied Mathematics* vol. 16, no. 3, pp 19-25, 2007

[5] M. Lupşor, R. Badea, S. Nedevschi, C. Vicaş, S. Tripon, H., C., M. Grigorescu. "The assessment of liver fibrosis using the computerized analysis of ultrasonographic images. Is the virtual biopsy appearing as an option?" Acta Electrotehnica, *Proceedings of the 1st Internationa Conference on Advancements of Medicine and Health Care Through Technology* Meditech 2007, Volume 48, Number 4, pp245-251, 2007

[6] H. Ian., W. Frank "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

[7] J. Cohen, "A coefficient of agreement for nominal scales." *Educational and Psychological Measurement,* vol. 20, pp 37–46.

[8] C. Gui-Tao, S. Peng-fei, H. Bing, "Liver fibrosis identification based on ultrasound images captured under varied imaging protocols" *Journal of Zhejiang University SCIENCE* vol. 6B(11), pp. 1107-1114, 2005

[9] J. Kenneth, W. Taylor, A.C. Riely, L. Hammers "Quantitative US Attenuation in Normal Liver and in Patients with Diffuse Liver Disease: Importance of Fat" *Radiology*, vol. 160, pp. 65-71, 2006.

[10] Lupşor M, Vicaş C, Nedevschi S, Badea R, Stefănescu H, Grigorescu M. "Imbunătătirea valorii diagnostice a examinării ultrasonografice în bolile hepatice difuze prin aplicarea unor metode de procesare a imaginilor." *Simpozionul Național de Cercetare Stiințifică Medicală de Excelență*, Sibiu, Romania, 25-26 octombrie 2007