

Steatohepatitis Detection from Ultrasound Images Using Attenuation and Backscattering Coefficients

Cristian Vicas, Sergiu Nedevschi,
Technical University of Cluj-Napoca
{cristian.vicas, sergiu.nedevschi}@cs.utcluj.ro

Monica Lupsor, Radu Badea
3rd Medical Clinic Cluj Napoca
{mmlupsor, rbadea2003}@yahoo.com

Abstract

Diagnosing liver nonalcoholic steatohepatitis (NASH) using non invasive procedures is challenging because the visual aspects in US imaging between healthy and affected liver are very much alike. Attenuation and backscattering coefficients are employed to determine the behavior of the US beams into the liver tissue. The values of these coefficients generate features for a classifier. We try to determine if is possible to discriminate between healthy and steatohepatitis patients and if is possible to assess the steatohepatitis grade.

Support vector machine is used as a classifier and the Cohen's kappa statistic evaluates the performance of the classifier. Experiments performed on US images acquired from a lot of 51 biopsied patients shown that these coefficients can successfully be used to distinguish between healthy and moderate/severe steatohepatitis. The role of fibrosis stage and fatty is also investigated.

Brunt score is used to assess the steatohepatitis grades. Fibrosis stage and steatosis grade are also determined.

1. Introduction

Nonalcoholic Steatohepatitis (NASH) is an important disease because of its prevalence in general population (over 20%) and because it could have a progressive course leading to cirrhosis and hepatic failure [1] [2]. Simplicity and noninvasiveness make ultrasonography an excellent diagnostic tool, albeit one that depends more on the ability and experience of the examiner and is affected by more unstable factors than any other imaging method. Several studies have discussed tissue characterization, which involves measuring the physical parameters of ultrasonograms to obtain quantitatively objective ultrasound information on which a diagnosis can be based [3].

There are two main directions one based on texture attributes of the generated US image and the other one based on the attenuation of the US beam into the liver tissue. This article focuses on the last approach. The behavior of the ultrasound beam into the liver can be assessed using two coefficients: backscattering and attenuation. Backscattering coefficient measures the quantity of energy that is returned by the tissue back to the receiver. Attenuation measures the ratio of absorption with respect to the depth.

Some of the methods employed to compute these coefficients are based on measuring the speckle distribution [4] directly from the RF signal [5] [6] by spectral shift central frequency method [7] or using a high frequency transducer [8].

B-mode ultrasound imaging implies transforming the amplitude of the received signal into intensity gray levels. Based on this property the behavior of the ultrasound into the liver can be assessed by examining the ultrasound image [1] [3] [9]. These methods are employed in this paper to differentiate between normal liver and steatohepatitis and to distinguish between various steatohepatitis grades. The role of fibrosis stage and fatty grade is investigated. It is shown that these pathologies affect the US propagation and therefore the discrimination between steatohepatitis grades.

In chapter 2 are presented the attenuation and backscattering coefficients, in chapter 3 is a short presentation of the classification algorithm. Chapter 4 presents the experimental results on a lot of 51 patients and in chapter 5 some conclusions are drawn based on the results.

2. Attenuation and backscattering coefficients

A straight line is fitted along the ultrasound image. This line represents the ultrasound beam path into the

liver tissue. (Figure 1) The average gray level values for each point along this line are calculated by averaging a certain number of pixels from left and right of this line (including the pixel beneath the line)[1].

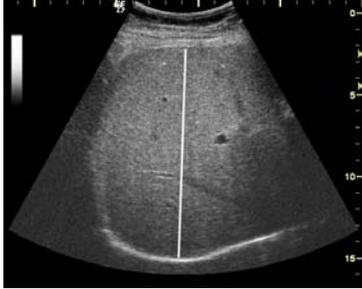


Figure 1 Ultrasound image from right lobe at 16 cm. The mean gray levels are computed along the white line.

For each point of the line there were stored two values: the average gray level computed as above and the depth. Linear regression by least-squares approximation was applied to this dataset.

The slope represents the attenuation coefficient and the offset (intercept) represents the backscattering coefficient.

Two additional features were calculated, the correlation coefficient r^2 (Equation 1) and the fitting error (Equation 2)

$$r^2 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x}) \sum_i (y_i - \bar{y})}$$

Equation 1. Correlation coefficient

$$s^2 = \sum \frac{(y_i - (a + b \cdot x_i))^2}{n - 2}$$

Equation 2. Fitting error: a is the offset and b is the slope of the linear regression. n represents the number of points

The numbers of pixels that are averaged to the left and right of the line is a parameter to the algorithm.

The slope, offset, r^2 and s^2 represent the output of the algorithm

3. Classification Algorithm

Support Vector Machines (SVM) or kernel machines are a family of learning methods that use an

efficient training algorithm and can represent complex, nonlinear functions [10] [11].

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.

Kernel machines usually find the optimal linear separator, the one that has the largest margin between it and the positive examples on one side and the negative examples on the other side.

3.1 Assessing the classification performance

One of the common methods of assessing the classification performance is the cross-validation. The basic idea is to estimate how well each hypothesis will predict unseen data. This is done by setting aside some fraction of the known data and using it to test the prediction performance of a hypothesis induced from the remaining data. K-fold cross-validation means that one run k experiments, each time setting aside a different $1/k$ of the data to test on, and average the results. Popular values of k are 5 and 10. The extreme is $k=n$, also known as leave-one-out cross-validation.

Cross validation can also be used when we want to determine the best parameters for a classification algorithm.

In order to compare two classification algorithms, or the performance among some datasets the kappa statistic is used [12]. This coefficient measures the agreement between predicted and observed categorizations of a dataset, while correcting for agreement that occurs by chance (Figure 2). The maximum value for kappa is 1 corresponding to a perfect classifier. The value 0 corresponds to a random classifier.

$$\kappa = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)},$$

Figure 2 Computing the kappa statistic

where $\text{Pr}(a)$ is the relative observed agreement among raters, and $\text{Pr}(e)$ is the probability that agreement is due to chance.

In order to compare two kappa values the Student's t-test is used [11] with significance level set at 0.05

4. Experimental Results

4.1 Steatohepatitis evaluation

Brunt score is used throughout this paper to assess the steatohepatitis grade.

Brunt gradation (or score) consists in 3 grades plus the healthy “0” grade: grade 1 - mild ; grade 2 – moderate and grade 3 - severe.

When a physician evaluates the Brunt score he examines the following components: steatosis, hepatocyte ballooning, lobular inflammation and portal inflammation.

The fibrosis grade is also evaluated using 4 grades.

In this study for each patient the following information were determined by liver biopsy: brunt score, fatty (steatosis) grade and fibrosis stage.

4.2. Patients and ultrasound images

Over 500 patients were investigated using various techniques including US examination and liver biopsy. Only 54 patients have NASH pathology and are suited for our study. From each patient a number of US images were acquired from right lobe. The US machine parameters were set in such a way that the quantity of the information received from the tissue was maximized. The post processing settings were set to off. TGC curve set to neutral and *Depth* set to 16 cm. All US images were acquired using exactly the same machine settings.

According to Brunt score, fatty and fibrosis stage the patients are divided in several categories. On each US image (where it was possible) a straight line was fitted in such a way that it will avoid artifacts. The line has to be as parallel as possible with the US rays, preferably vertical. The fitted line is the region of interest. Some of patient groups have less than 30 regions of interest and they were eliminated from the study.

The control or healthy group consists of patients who do not have any liver illness. This group wasn't biopsied so Brunt 0 grade for these patients cannot be firmly presumed. In Table 1 is shown the patient lots taken into consideration.

Table 1 Patient and ultrasound image distribution

Brunt	Fatty	Fibrosis	Patients	US images
Healthy	N/A	N/A	18	137
1	1	1	17	121
1	2	2	3	36
2	2	1	4	38
2	2	2	5	48
3	3	1	4	41

One can notice that only several combinations of scores are present. We avoided comparing two groups

that have only one (or two) patient each because the results will yield no statistical significance.

Datasets were created by including presented US images along with special created datasets where one or two parameters were ignored (i.e. a dataset consisting in Brunt grade 1 patients regardless of the fatty or fibrosis stage). One dataset consists in healthy patients only. 31 such datasets were generated.

4.3. Features

The four presented features were implemented. The gray level for each depth level was computed by averaging 7, 11 and 15 pixels on a horizontal line centred on the established line.

These settings give us 12 features for each ultrasound image.

Image features were computed using a framework developed by the authors.

4.4. Classification algorithm

A SVM implementation [13] is used. Weka [11] data mining system provides the SMO algorithm that is used to train a support vector machine. The most important SMO parameters are as following: RBF kernel, gamma set to 1, complexity parameter set to 4 and the features are normalized. Cross validation with 10 folds is used to establish the true positive rates for each class. (the ratio between correctly classified instances and the total number of instances in that class). 5 runs of cross validation with different splits are performed.

In order to get a better view of the results the classifier has to discriminate only two classes at a time.

As presented previously, 31 datasets were generated. Each dataset was compared with every other dataset using the SMO algorithm. The mean kappa statistic over the 10 cross-validations and 5 runs evaluates the performance of the algorithm.

4.5. Results

4.5.1 Detecting steatohepatitis. First problem is discriminating between control group (healthy patients) and steatosis groups. In Table 2 is shown the kappa statistic when comparing the three steatosis grades. When building steatosis group we considered only the Brunt score.

Table 2 Comparison between normal lot and Brunt grades

Comparison case	Kappa
Normal – Brunt 1	0.7
Normal – Brunt 2	0.85
Normal – Brunt 3	0.99

One can notice a very good discrimination ratio between normal group and severe steatosis group.

Lower kappa value is obtained when comparing normal lot versus mild steatosis (Brunt score 1)

This lot is further divided according to fibrosis level in order to study the impact of the fibrosis grade on the detection rates.

In Brunt 1 lot we have two fibrosis levels. (F1 and F2) We will try to discriminate between normal group and those two groups. The results are shown in Table 3

Table 3 Normal – Brunt 1 comparison. The effect of fibrosis stage

Comparison case	Kappa
Normal – Brunt 1 Fibrosis 1	0.73
Normal – Brunt 1 Fibrosis 2	0.76

One can notice a better results when we separate the lot according to fibrosis grade. This difference however have no statistical significance ($p < 0.05$) when is compared with the performance of Normal – Brunt 1 case.

We divide the Brunt 1 group according to steatosis level. We find that there are two levels, fatty 1 and fatty 2. (Table 4) Again, we note the kappa statistic when differentiating between normal lot and these two lots.

Table 4 Normal – Brunt 1 comparison. The effect of steatosis grade

Comparison case	Kappa
Normal – Brunt 1 Fatty 1	0.66
Normal – Brunt 1 Fatty 2	0.95

In fatty grade 1 lot the kappa statistic is not significantly lower than those obtained when we considered the fibrosis grade.

In fatty grade 2 lot the kappa statistic is significantly greater than those when considering fibrosis grade, or when comparing the whole Brunt 1 lot.

Unfortunately we cannot divide the lot furthermore according to each fibrosis grade because the volume of data decreases considerably. Only two comparisons were available (Table 5):

Table 5 Mixed effect of fatty and fibrosis on Normal – Brunt 1 discrimination

Comparison case	Kappa
Normal – Brunt 1 Fatty 1 Fibrosis 1	0.7
Normal – Brunt 1 Fatty 2 Fibrosis 2	1

One can notice the importance of both fibrosis grade and fatty grade in discriminating the normal patients from Brunt grade 1 patients.

Studying the importance of fatty and fibrosis grades over the discrimination between normal and Brunt grade 2 we find the results in Table 6

Table 6 Normal – Brunt 2 discrimination. Importance of fatty and fibrosis

Comparison case	Kappa
Normal – Brunt 2 Fibrosis 0	0.84
Normal – Brunt 2 Fibrosis 1	0.96
Normal – Brunt 2 Fibrosis 2	0.83
Normal – Brunt 2 Fatty 2	0.87
Normal – Brunt 2 Fatty 2 Fibrosis 1	0.96
Normal – Brunt 2 Fatty 2 Fibrosis 2	0.76

These values are compared with the Normal – Brunt 2 comparison case.

Only when we consider the Fibrosis 1 component the detection rates increases significantly ($p < 0.05$).

When we consider the fatty, only grade 2 has a sufficient data volume for comparison. The detection rate remains unchanged. Again, when considering fibrosis 1 grade one can notice a significant increase in detection rates. It is not clear why when considering the fibrosis grade 2 we don't obtain better results.

4.5.2 Detecting steatohepatitis grade. Another area of interest is differentiating between steatohepatitis grades. In Table 7 are shown the results when comparing Brunt 1, Brunt 2 and Brunt 3 steatohepatitis grades.

Table 7 Discriminating between Brunt grades

Comparison case	Kappa
Brunt 1 – Brunt 2	-0.01
Brunt 1 – Brunt 3	0.2
Brunt 2 – Brunt 3	0.16

The results are significantly lower than those obtained when comparing with normal lot. The Brunt grade 1 is not discriminated at all from Brunt grade 2. In the following (Table 8) we will focus on the impact of fibrosis and fatty grades on comparison Brunt 1 and Brunt 2 lots.

Table 8 The effect of fibrosis stage on Brunt 1 - Brunt 2 detection

Comparison case	Kappa
Brunt 1 Fibrosis 1 – Brunt 2 Fibrosis 1	0.18
Brunt 1 Fibrosis 2 – Brunt 2 Fibrosis 2	0.5
Brunt 1 Fibrosis 1 – Brunt 3 Fibrosis 1	0.4

The results are significantly greater than those when we do not consider the fibrosis. Unfortunately the Brunt 1 Fibrosis 0 did not have enough samples so it was discarded. In Table 9 are shown all the comparison cases involving Brunt 1, 2 and 3 grades and fibrosis grades.

Table 9 The effect of various fibrosis stages over Brunt 1 – Brunt 2/3 discrimination

Comparison case	Kappa
Brunt 1 Fibrosis 1 - Brunt 2 Fibrosis 0	0
Brunt 1 Fibrosis 1 - Brunt 2 Fibrosis 1	0.18
Brunt 1 Fibrosis 1 - Brunt 2 Fibrosis 2	0.03
Brunt 1 Fibrosis 2 - Brunt 2 Fibrosis 0	0.41
Brunt 1 Fibrosis 2 - Brunt 2 Fibrosis 1	0.49
Brunt 1 Fibrosis 2 - Brunt 2 Fibrosis 2	0.5
Brunt 1 Fibrosis 1 - Brunt 3 Fibrosis 1	0.4
Brunt 1 Fibrosis 2 - Brunt 3 Fibrosis 1	0.12
Brunt 2 Fibrosis 0 - Brunt 3 Fibrosis 1	0.41
Brunt 2 Fibrosis 1 - Brunt 3 Fibrosis 1	0.75
Brunt 2 Fibrosis 2 - Brunt 3 Fibrosis 1	0.55

Fibrosis has a major impact in Brunt score discrimination especially when discriminating between lower/higher brunt grade combined with higher/lower fibrosis grade: Brunt 1 Fibrosis 2 – Brunt 3 Fibrosis 1 comparison yields a kappa=0.12, but when the first group has also fibrosis 1, kappa is significantly greater (0.4) The same results can be found in Brunt 1 Fibrosis 1 - Brunt 2 Fibrosis 0. Unfortunately the same results are not found when comparing Brunt 2 Fibrosis 2 versus Brunt 3 Fibrosis 1 where kappa is larger than expected.

Discriminating Brunt 1 Fibrosis 1 versus Brunt 2 grades is difficult regardless of Fibrosis grade in Brunt 2 lot. On the other hand, the Brunt 1 Fibrosis 2 group is better discriminated. The significance of this result is unclear, but this phenomenon is present when comparing Brunt 1 Fibrosis 1 with Brunt grade 3 group. Because no clear rule on how the fibrosis alters the detection performance can be assessed, one might conclude that other pathologies like hepatocyte ballooning, inflammation and necroses have their effects on the US image and alters the discrimination process.

These results show that fibrosis has a major role in steatosis grade discrimination. A more comprehensive picture could be drawn if we were able to compare combinations like Brunt 1 Fibrosis 3, Brunt 3 Fibrosis 0, etc.

In Table 10 is shown the impact of fatty over the Brunt 1 – Brunt 2 discrimination. Again, the volumes for some comparison cases are lower than the selected threshold.

Table 10 The effect of the fatty over Brunt 1 – Brunt 2 comparison

Comparison case	Kappa
Brunt 1 Fatty 1 – Brunt 2 Fatty 2	0.12
Brunt 1 Fatty 2 – Brunt 2 Fatty 2	0.2

The results are significantly lower than those in fibrosis. One might conclude that fatty has a lower impact on discrimination accuracy than fibrosis. However the reduced dataset volumes and the lack of some interesting fibrosis/fatty combinations diminish the viability of this conclusion. Another interesting observation is that in Brunt 1 – Brunt 2 the results were significantly lower than those when we consider only the fatty grade 2 patients.

4.5.3 Detecting fibrosis grade. We try to distinguish between normal lot and fibrosis grade. In Table 11 are shown the results.

Table 11 Detection of fibrosis stage

Comparison case	Kappa
Normal – Fibrosis 0	0.9
Normal – Fibrosis 1	0.78
Normal – Fibrosis 2	0.78

One might conclude that using attenuation/backscattering coefficients the fibrosis grade can be accurately assessed. Unfortunately at a closer inspection one notices that Fibrosis 0 is mostly associated with Brunt 2 group, and the rest of fibrosis grades have equal distributions among Brunt 1 and 3 grades.

4.5.4 Detecting fatty (steatosis) and fatty grade. We compare normal patients with various grades of fatty. In Table 12 is shown the results

Table 12 Detection of steatosis grade

Comparison case	Kappa
Normal – Fatty 1	0.65
Normal – Fatty 2	0.9
Normal – Fatty 3	1

In discriminating between healthy patients fatty grade 2 is better detected than brunt grade 2. This result is statistical significant. In case of discriminating between normal – fatty 1 the lower kappa value is not statistical significant.

These findings suggest that other phenomena like fibrosis or necrosis might interfere and alter the attenuation/backscattering coefficients. These interferences are best viewed at mild fatty grades. When the steatosis is greater (fatty grade 2 and 3) these interferences have little impact on studied features.

5. Discussions and conclusions

Attenuation and backscattering coefficients can be successfully employed to distinguish between normal liver and moderate/severe steatohepatitis. The detection rates are almost 100% in detecting severe steatohepatitis. Mild steatohepatitis have good discrimination rates (kappa = 0.7) but there is room for improvements. A tuned classifier or association with texture features might give us improved results.

The fibrosis grade doesn't affect much the steatohepatitis detection. The results were improved when considering the fibrosis stage but the difference wasn't statistical significant.

Steatosis (fatty) have an interesting impact over the steatohepatitis discrimination. Grade 1 steatosis does not affect the discrimination performance between normal and Brunt 1 patients.

Grade 2 steatosis on the other hand yields a kappa=0.95. As a result, patients with brunt grade 1 associated with fatty grade 2 can be successfully distinguished from healthy patients.

Another important issue is separating the patients according to their steatohepatitis grades. The brunt 1 grade patients were not distinguished from brunt 2 patients. (kappa is almost 0, which means that the classifier randomly distributed the patients in the two groups) The brunt 1 – brunt 3 and brunt 2 – brunt 3 comparison cases yields better results, but kappa does not exceed 0.2. We suspected that fibrosis and fatty has a major impact on US image characteristics but more patients have to be examined in order to populate the excluded patient groups. In some combinations of brunt/fatty/fibrosis the discrimination was higher than in the rest of the cases. This suggest that a classifier could be trained to assess the brunt steatohepatitis grade at least when specific fibrosis/fatty grades are present.

The hepatocyte ballooning, inflammation and necrosis have their effect on the US image. The way that these phenomena alter the image remains unclear

but their effects affect the discrimination between steatohepatitis grades and between healthy and mild steatohepatitis grade.

More patients have to be investigated in order to clearly assess the impact of fibrosis, inflammation and fatty on the ultrasound image. Understanding these effects will allow us to make important steps throughout steatohepatitis grade detection based on ultrasound images.

6. Acknowledgements

This work was sponsored 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 D. Gaitini, Y. Baruch, E. Ghersin, E. Veitsman "Feasibility Study Of Ultrasonic Fatty Liver Biopsy:Texture Vs. Attenuation And Backscatter" *Ultrasound In Med. & Biol.*, Vol. 30, No. 10, Pp. 1321-1327, 2004.
- 2 T. Poynard, V. Ratziu, P. Bedossa. "Appropriateness of liver biopsy." *Can J Gastroenterol*, vol. 14, no. 6, pp. 543-548, 2000.
- 3 A. Szebenia, G. Tolvajb, A. Zalatnaic "Correlation of ultrasound attenuation and histopathological parameters of the liver in chronic diffuse liver diseases" *European Journal of Gastroenterology & Hepatology*, vol. 18, pp. 37-42, 2006.
- 4 Y. Fujii, N. Taniguchi, Y. Wang, "Clinical Application Of A New Method That Segments The Region Of Interest Into Multiple Layers For Rf Amplitude Histogram Analysis In The Cirrhotic Liver" *J Med Ultrasonics*, vol. 3, pp. 91-98, 2004.
- 5 Z.F. Lu, J.A. Zagzebski, F.T. Lee "Ultrasound Backscatter And Attenuation In Human Liver With Diffuse Disease" *Ultrasound In Med. & Biol.*, Vol. 25, No. 7, Pp. 1047-1054, 1999.
- 6 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.
- 7 Y. Fujii, N. Taniguchi, K. Itoh, K. Shigeta, "A New Method for Attenuation Coefficient Measurement in the Liver" *J Ultrasound Med*, vol. 21, pp. 783-788, 2002.
- 8 W.C. Yeh, Y.M. Jeng, C.H. Li, P.H. Lee, And P.C. Li "Liver Steatosis Classification Using High-Frequency Ultrasound" *Ultrasound In Med. & Biol.*, vol. 31, no. 5, pp. 599-605, 2005.

9 M. Yasser, A. Kadah, A. Farag, M. Jacek, "Classification Algorithms for Quantitative Tissue Characterization of Diffuse Liver Disease from Ultrasound Images" *Ieee Transactions On Medical Imaging*, vol. 15, no. 4, august 1996.

10 S. Russell, P. Norvig "Artificial Intelligence: A Modern Approach," *Prentice Hall Series in Artificial Intelligence*, Englewood Cliffs, New Jersey, 1995.

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

12 J. Cohen, "A coefficient of agreement for nominal scales." *Educational and Psychological Measurement*, vol. 20, pp 37-46.

13 J. Platt, "Fast training of support vector machines using sequential minimal optimization." *Advances in kernel methods: Support vector learning*. Cambridge, MA: MIT Press. 1998.