

DETECTION AND STAGING OF LIVER FIBROSIS USING ADDITIVE LOGISTIC MODELS

Cristian Vicas, Sergiu Nedevschi
Technical University Cluj Napoca,
28, Gh. Baritiu St. 400027, Cluj Napoca, Romania
{cristian.vicas, sergiu.nedevschi}@cs.utcluj.ro

Monica Lupsor, Radu Badea
3rd Medical Clinic Cluj Napoca,
19-23, Croitorilor St. 400162, Cluj Napoca, Romania
{monica.lupsor, rbadea}@umfcluj.ro

ABSTRACT

Fibrosis and cirrhosis are the main complications of chronic liver diseases. At present, liver biopsy is the golden standard for evaluating liver fibrosis. However, this is an invasive procedure, hence the interest in developing non-invasive approaches. The present study identifies novel possibilities for non-invasive fibrosis evaluation. We included 591 hepatitis C patients. Fibrosis was assessed using the Metavir score. A number of 93 features were obtained from each patient using B-mode ultrasound, Doppler ultrasound, transient elastography and common biochemical and cytological measurements. The patients were grouped according to fibrosis stages and additive logistic regression models were built. Cross-validation along with Area Under Curve (AUROC) was used to measure the classification performance. The AUROC of 0.90 was recorded when discriminating between fibrosis stage ≤ 3 and fibrosis stage 4.

KEY WORDS

Liver fibrosis, non-invasive diagnosis, medical diagnostics, additive logistic regression models

1. Introduction

The accurate diagnosis and staging of hepatic fibrosis is crucial for the prognosis and treatment of liver diseases. For the moment, liver biopsy remains the gold standard for fibrosis assessment in hepatitis C. Liver biopsy presents several important drawbacks. In [1] authors note morbidity, observer variability and sampling variation.

There are several non-invasive alternative methods to investigate the liver: ultrasound investigation including simple B-mode and Doppler mode, usual cytological and biochemical measurements used to investigate the liver function, special serum markers that are involved in fibrosis generation, transient elastography etc.

In [2], using Doppler ultrasonography, authors record the maximum velocity of blood at the portal vein and hepatic artery in 19 patients with cirrhosis, 61 with chronic hepatitis and 20 healthy patients. For each patient the arterio-portal (A/P) ratio was computed. The authors noted that the levels of A/P ratio were significantly higher in patients with liver cirrhosis.

Another approach presented in [3] investigated the serum levels of hyaluronic acid, type III procollagen, N-terminal procollagen etc. in patients with hepatic fibrosis. The authors included 114 serum samples from biopsied patients divided in two groups according to their fibrosis stage ($S \geq 1$). They computed an AUROC of 0.8 for procollagen III peptide marker.

The aspartate aminotransferase to platelet ratio index was proposed in [4] as a non-invasive index to predict significant fibrosis and cirrhosis. Using one training set including 192 patients and a validation set consisting of 78 patients the authors used this index to predict significant fibrosis (Ishak score ≥ 3) and cirrhosis. The AUROC for significant fibrosis in validation set was 0.87 and 0.93 for cirrhosis.

Ultrasound based transient elastography is a non-invasive procedure that determines the stiffness of the liver tissue. Liver tissue stiffness is associated with the degree of fibrosis. In [5] the authors performed a review on the performance of transient elastography in detecting liver fibrosis. Their conclusion was that transient elastography is clinically useful in assessing the presence or the absence of cirrhosis. Nevertheless, discriminating between other fibrosis stages is still a challenging task.

In another paper [6] the authors reviewed a series of non-invasive tests including various commercially available serum kits, transient elastography and blood flow investigation. Combining biomarkers that measure different fibrosis components may advance diagnosis accuracy. The authors also note that the imperfect golden standard might impose some limitations in developing a non-invasive test. This could be compensated by developing longitudinal tests.

In [7] the authors reviewed 153 papers that approach non-invasive fibrosis detection. They included all papers referring to imaging modalities and use liver biopsy as reference. They concluded that the diagnostic performances of all described non-invasive radiologic modalities were better in distinguishing between patients with cirrhosis and patient affected by lesser degrees of fibrosis. However, fibrosis was rarely reliably staged. Again, transient elastography appears to be the most promising technique.

In the present paper we included most of the non-invasive features analyzed in [2, 4-5, 8]. A number of 93 features were acquired from the 591 patients.

These features included transient elastography evaluation, several common serum markers, B-mode ultrasound to measure various liver/spleen dimensions and Doppler measurements to evaluate the blood flow. The Metavir score was used to assess the fibrosis stage.

The problems related to fibrosis detection problem were examined from various points of view by grouping the patients in six 2-class problems and two multiclass problems. We used additive logistic models as a classifier. The performance of the classification schema was assessed using cross validation.

The additive logistic models employed here are able to perform an automatic attribute selection. We identified the relevant features using a ranking schema.

The present paper brings several contributions to the non-invasive diagnosis field. We investigated the usefulness of a large number of features including transient elastography. In addition, we investigated the discrimination power of a set of features that can be acquired in any medical facility without the need of special equipment or access to a laboratory.

This paper included a large volume of patients, one of the largest met in fibrosis detection literature.

We also identified new research directions involving non-invasive fibrosis detection. Several visual scores, like capsule regularity were found to be good predictors for fibrosis staging.

The paper is structured as follows: Chapter 2 briefly describes the classification algorithm and the performance evaluation methodology. Chapter 3 introduces the features recorded for each patient and the feature ranking algorithm. Chapter 4 presents the experimental results while Chapter 5 the discussions on these results.

2. Learning Schema and Performance Evaluation

In this section we will present the additive logistic models employed, as well as the performance evaluation methodology. Patients are divided in two or multi-class problems. The division of patients is performed based on their fibrosis stage. The biological and ultrasound data recorded for each patient are used as features for the classifier. The classifier is trained to predict the class for each patient. Performance is evaluated using the Cross Validation method.

2.1 Additive Logistic Models

The classifier employed here is based on additive logistic models [9].

Let (x,y) be an instance, $x=(x_1,x_2, \dots x_n)$ represents a feature vector and y represents the class of the instance.

The logistic regression model is shown in Equation 1.

$$p(x) = P(y = 1|x) = \frac{e^{f(x)}}{1 + e^{f(x)}} \quad (1)$$

where

$$f(x) = \beta_0 + \sum_{i=1}^n \beta_i x_i \quad (2)$$

Fitting a logistic regression model means finding the optimum parameter vector $\beta=(\beta_0,\beta_1,\dots,\beta_n)$. The standard procedure is to search for the maximum likelihood. There are iterative algorithms that asymptotically approach the maximum likelihood solution.

An additive logistic model has the form presented in Equation 3.

$$F_M(x) = \sum_{m=1}^M f_m(x) \quad (3)$$

where $f_m(x)$ represents a logistic regression model and M represents the number of logistic models that build up the additive model at step M .

Friedman et al. presented in [10] the *LogiBoost* algorithm designed to build additive logistic models.

This algorithm performs iterative forward stage wise fitting. In every iteration, it computes the response variables that encode the error of the currently fit model on the training examples. Then it tries to improve the model by adding a function $f_m(x)$ to the $F(x)$.

In the special case where the functions $f_m(x)$ are linear functions of the input variables, the additive logistic regression model is equivalent to a linear logistic model.

Based on *LogiBoost*, Landwehr et al. proposed in [9] the *SimpleLogistic* algorithm. For $f_m(x)$ the authors used a simple regression function that performs a regression on only one attribute. At each step of the algorithm, only one attribute is added to the final model, namely the most significant one.

The algorithm is stopped when adding new attributes does not improve the prediction accuracy on unseen instances. The accuracy for each model is assessed using a 5 fold cross-validation loop.

After determining the best number of iterations the algorithm is run on the entire training set using the already determined number of iterations.

In the present paper, the *SimpleLogistic* algorithm was used. This algorithm has the advantage of producing a model that can be observed and examined. Another advantage is that *SimpleLogistic* algorithm performs an intrinsic attribute selection.

2.2 Performance evaluation

In order to evaluate the classification performance as close as possible to real-life situations, the following procedure is used: The available dataset is split using 10 fold stratified cross validation [11]. At each step one fold is kept for testing and the rest of 9 folds are used to build the training set. For each training set missing values are replaced using mean imputation. Normalization is then performed and the resulting dataset is used as training set for the classification schema. The mean values and normalization parameters inferred on the training set are applied to the corresponding test set.

At each of the 10 steps the predictions on the unseen test sets are collected. After the cross-validation schema is ended, the Area Under Curve (AUROC) and the variation σ^2 of AUROC is computed on the collected prediction set using Mann-Whitney-Wilcoxon U statistic [12].

Confidence intervals for the mean AUROC (95%) are computed using Equation 4:

$$CI = 1.96 \frac{\sigma}{\sqrt{n}} \quad (4)$$

For binary problems we reported only one AUROC value. For multiclass problems we reported the individual AUROC value for each class.

3. Features and Feature Ranking

3.1 Biological and Ultrasound Features

A number of 93 features were recorded for each patient.

The features acquired from each patient are grouped into five categories depending on the acquisition process. Table 1 presents the main categories along with several examples of features.

Another classification of the features could be achieved by means of their operator-dependent variability. Some examinations are objective measurements such as biochemical and serum markers, others are measurements performed by a human expert using an ultrasound scanner (liver dimensions, blood flow measurements, etc) and others are purely subjective scores established by the same human expert (liver capsule regularity, liver homogeneity, etc). The measurement unit of each feature was kept constant throughout the experiments. However, the exact measurement unit is not important because each feature underwent a normalization process.

Biological parameters are acquired by physicians during routine examinations. Depending on the clinical condition of the patient, the physician may choose to exclude certain investigations. As a result, it is very likely that not all attributes should be available for all patients.

One needs to infer the values for these missing features. In the present paper we chose mean imputation to infer values for missing attributes. For each attribute the mean is computed and the missing attributes are replaced with the computed mean.

We also investigated the power of the features that can be measured using only B-mode ultrasound and Doppler. We chose only objective features that can be acquired in any medical facility without the need of special equipment. We excluded biochemical measurements and transient elastography. We also excluded subjective visual scores such as liver homogeneity, capsule regularity and hepatic angles. Out of the patient characteristics we only included age and sex. The remaining dataset consisted of 42 features.

3.2 Feature Ranking Schema

The additive logistic model employed in the present paper has the ability to perform an intrinsic feature selection.

The usefulness of a feature is given by the $|\beta_i|$. The greater $|\beta_i|$ for a feature, the greater its importance in the model. In order to evaluate the performance of each feature independently from the dataset variability the following experiment was performed for each binary problem: 66% of the available patients were randomly selected. On this fold, an additive logistic regression model was fit, and the relevant features were recorded.

After 10 runs, we ranked the attributes based on the number of models in which the feature was selected. A ranking of ten for a feature means that the feature was selected in all 10 models. For each feature, we computed the mean and standard deviation for the β_i coefficient.

The ranking schema used here allowed us to identify the attributes that are useful in fibrosis prediction. We did not examine directly the particular logistic model trained on a full dataset because this particular model overfits the particular dataset used for training.

Table 1. Examples of features used in this paper. Only several features are listed from each category.

Category	Example of features	Number of features
Patient characteristics	Age, sex, height, weight, body mass index	7
Common serological markers	aspartate aminotransferase, urea, γ -glutamyl-transpeptidase, triglycerides, cholesterol, HDL cholesterol, platelets, sideremia	25
B-mode ultrasound evaluation	Liver dimensions, Spleen dimensions, liver capsule regularity, liver homogeneity, lymph node in the hepatic hilum, visceral fat thickness, thickness of the abdominal aortic wall	28
Doppler measurements	Velocity in portal vein, Time averaged maximum velocity in hepatic artery, Time averaged mean velocity in hepatic artery, hepatic artery peak systolic velocity, flow acceleration in hepatic arter	30
Transient elastography	Liver stiffness	3

4. Experimental Results

A number of 591 patients were included in the present study. The patients provided written informed consent before the beginning of the study, in accordance to the principles of the Declaration of Helsinki (revision of Edinburgh, 2000).

The present study was approved by the local Ethical Committee of the University of Medicine and Pharmacy of Cluj-Napoca. For 535 patients liver fibrosis was assessed using the Metavir score after liver biopsy. In addition to biopsied patients a number of 24 clinically healthy patients and 32 known cirrhotic patients were also included in the

study. For ethical considerations these patients were not biopsied. Table 2 shows patient distribution according to the fibrosis stage. Each patient provided one set of features. This set, along with the corresponding fibrosis stage was considered an instance in our datasets.

On average, each feature was missing in 109.7 instances (standard deviation 119.3) and in each instance there were on average 17.2 features missing (standard deviation of 12.6)

The patients were grouped into 6 synthetic binary problems and in two realistic, multiclass problems. In addition to this, there were two sets of features used for each problem. As a result, the additive logistic model was evaluated 16 times. The relevant features were determined for each 6 binary problems using both feature sets. Only features with a minimum ranking of 8 were shown.

Table 2. Patients grouped according to Metavir findings.

Patient condition	Abbreviation	Volume
Fibrosis stage 0 or healthy patients	F0	62
Fibrosis stage 1	F1	187
Fibrosis stage 2	F2	183
Fibrosis stage 3	F3	84
Fibrosis stage 4 or cirrhotic patients.	F4	75

4.1 Experimental results on binary problems

Experiments presented in this section can give an important insight into what the relevant features for detecting a certain fibrosis stage are. Table 3 shows the six experiments and the corresponding performance evaluation. For each binary problem we used the full and the reduced feature set.

The relevant features for each combination of binary problem and feature set are shown in Table 4 through Table 15.

Table 3. AUROC results on synthetic datasets for initial and for reduced set of features.

Dataset	Initial feature set		Reduced feature set	
	AUROC	Confidence intervals	AUROC	Confidence intervals
F0/F4	0.97	±0.03	0.90	±0.06
F01/F34	0.90	±0.04	0.84	±0.04
F0/F1234	0.81	±0.04	0.74	±0.06
F01/F234	0.77	±0.04	0.70	±0.05
F012/F34	0.86	±0.04	0.84	±0.04
F0123/F4	0.90	±0.05	0.85	±0.05

Table 4. Features for F0/F4 dataset sorted according to their ranking. The β_i coefficients are for model that predicts class F4.

Ranking	Feature	Mean β_i	Std Dev β_i
10	Liver homogeneity	11.64	5.85
9	Liver capsule regularity	1.69	0.63
8	Age	2.56	1.92

Table 5. Features for F01/F34 dataset sorted according to their ranking. The β_i coefficients are for model that predicts class F34.

Ranking	Feature	Mean β_i	Std Dev β_i
10	Liver stiffness	10.60	5.4
10	Aspartate aminotransferase	3.61	1.09
10	Age	1.99	0.59
10	Liver capsule regularity	1.71	0.20
9	Cholesterol	-1.88	1.15
8	Left lobe diameter	1.40	0.79

Table 6. Features for F0/F1234 dataset sorted according to their ranking. The β_i coefficients are for model that predicts class F1234.

Ranking	Feature	Mean β_i	Std Dev β_i
10	Liver stiffness	8.40	1.8
10	Age	1.29	0.23
10	Left lobe diameter	1.92	0.15
8	Cholesterol	-1.91	1.32

Table 7. Features for F01/F234 dataset sorted according to their ranking. The β_i coefficients are for model that predicts class F234.

Ranking	Feature	Mean β_i	Std Dev β_i
10	Liver stiffness	5.67	2.26
10	Aspartate aminotransferase	2.70	0.53
9	Cholesterol	-1.13	0.51
9	Age	1.03	0.46
9	Liver capsule regularity	0.75	0.44

Table 8. Features for F012/F34 dataset sorted according to their ranking. The β_i coefficients are for model that predicts class F34.

Ranking	Feature	Mean β_i	Std Dev β_i
10	Liver stiffness	5.24	1.94
10	Aspartate aminotransferase	2.27	0.40
10	Age	1.27	0.13
8	Left lobe diameter	1.01	0.63
8	Liver capsule regularity	0.95	0.57

Table 9. Features for F0123/F4 dataset sorted according to their ranking. The β_i coefficients are for model that predicts class F4.

Ranking	Feature	Mean β_i	Std Dev β_i
10	Liver stiffness	3.92	0.63
10	Liver capsule regularity	1.28	0.13

Table 10. Features for F0/F4 dataset sorted according to their ranking when using the reduced set of features. The β_i coefficients are for model that predicts class F4.

Ranking	Feature	Mean β_i	Std Dev β_i
10	Age	2.91	0.65
4	Spleen perimeter	1.27	1.65
3	Spleen area	3.05	4.93

Table 11. Features for F01/F34 dataset sorted according to their ranking when using the reduced set of features. The β_i coefficients are for model that predicts class F34.

Ranking	Feature	Mean β_i	Std Dev β_i
10	Left lobe diameter	2.33	0.33
10	Age	2.09	0.12
9	Right lobe diameter	-1.05	0.44
8	Caudate lobe diameter	1.42	0.85
8	Gallbladder wall thickening	1.18	0.64

The best discriminated case was that of the F0 vs F4 problem. In this dataset were included only healthy patients or having fibrosis stage 0 along with cirrhosis patients.

Taking the model trained for F0/F4 problem we tried to predict the other patients, having fibrosis stage 1, 2 and 3. Table 16 shows the confusion matrix.

Table 12. Features for F0/F1234 dataset sorted according to their ranking when using the reduced set of features.

The β_i coefficients are for model that predicts class F1234.

Ranking	Feature	Mean β_i	Std Dev β_i
10	Spleen perimeter	1.52	0.22
9	Age	1.32	0.65
8	Left lobe diameter	1.12	0.74
8	Splenic vein diameter	1.90	1.11

Table 13. Features for F01/F234 dataset sorted according to their ranking when using the reduced set of features.

The β_i coefficients are for model that predicts class F234.

Ranking	Feature	Mean β_i	Std Dev β_i
10	Age	1.57	0.26
9	Spleen perimeter	1.46	0.61
7	Left lobe diameter	1.01	0.76

Table 14. Features for F012/F34 dataset sorted according to their ranking when using the reduced set of features.

The β_i coefficients are for model that predicts class F34.

Ranking	Feature	Mean β_i	Std Dev β_i
10	Left lobe diameter	1.94	0.21
10	Age	1.81	0.30
10	Gallbladder wall thickening	1.64	0.28
10	Right lobe diameter	-1.25	0.17
9	Caudate lobe diameter	1.63	0.89
8	Splenic vein diameter	2.06	1.20

Table 15. Features for F0123/F4 dataset sorted according to their ranking when using the reduced set of features.

The β_i coefficients are for model that predicts class F4.

Ranking	Feature	Mean β_i	Std Dev β_i
10	Gallbladder wall thickening	1.93	0.47
10	Age	1.69	0.52
8	Splenic vein diameter	1.86	1.09
8	Spleen congestion	0.71	0.43

Table 16. Predictions for the patients having fibrosis stage 1,2 and 3. Each row represents a fibrosis stage. On F0 column is shown the number of patients predicted as being fibrosis stage 0. In F4 column is shown the number of patients predicted as being cirrhosis.

	F0	F4	Total
F1	123 (66%)	64 (34%)	187
F2	86 (47%)	97 (53%)	183
F3	19 (23%)	65 (77%)	84

4.2 Experiments on realistic datasets

The experiments presented in this section mimic the situation where a classifier is used in clinical practice to predict the fibrosis stage. We designed two multiclass scenarios. In the first scenario the data was partitioned in five classes, each fibrosis stage with its own class. The results are shown in Table 17. In the second scenario the data was partitioned in 3 classes: no or mild fibrosis, medium fibrosis and severe fibrosis. The results are shown in Table 18. For each problem we used the full and the reduced feature set.

Table 17. AUROC on a dataset containing one class for each Metavir score.

Class	Initial feature set		Reduced feature set	
	AUROC	Confidence intervals	AUROC	Confidence intervals
F0	0.85	± 0.04	0.71	± 0.03
F1	0.70	± 0.03	0.58	± 0.03
F2	0.58	± 0.02	0.56	± 0.04
F3	0.65	± 0.05	0.61	± 0.02
F4	0.88	± 0.03	0.83	± 0.04

Table 18. AUROC on a dataset containing the patients grouped in three classes. First class contained mild fibrosis (F0 or F1), the second class included patients with fibrosis stage 2 and the third class included severe fibrosis stages 3 and 4.

Class	Initial feature set		Reduced feature set	
	AUROC	Confidence intervals	AUROC	Confidence intervals
F01	0.77	± 0.04	0.67	± 0.05
F2	0.55	± 0.03	0.56	± 0.03
F34	0.84	± 0.04	0.78	± 0.04

5. Discussions

The AUROC's values enumerated in Table 3, Table 17 and 18 represent the estimated performance of the classification schema employed here. The real performance of a schema trained on these datasets might be different. The reader should note that the confidence intervals for AUROC curves could be biased. They are computed assuming a normal distribution of the performance evaluations, assumption which is not always true [13]. Moreover, in [14] the authors noted that cross validation produces biased variance estimation. We employed this assumption here because we did not

perform any comparison or model selection based on the performance estimations.

The missing attributes were not uniformly distributed especially with respect to fibrosis stage. Some measurements were excluded based on clinical decisions regarding the patient's health. For example, most of the serological measurements were missing from the healthy patients. As a result, the logical choice to replace missing values was by mean imputation.

Other schemas that infer the missing attribute value based on the existing attributes at other instances could introduce a severe bias because of the inhomogeneous fibrosis stage representation in the dataset.

The first two classification problems, namely F0/F4 and F01/F34 have a theoretical significance, because the real patient distributions are continuous. However, they show that there are detectable changes between healthy and cirrhotic patients in our datasets. The rest of the classification problems include all available patients. The difference between them is the splitting point. These models could be applied in the clinical practice. They distinguish between cirrhotic/non cirrhotic patients (F0123/F4) with relatively high performance. The discrimination performance decreases as the split point is between lower fibrosis stages. These findings are similar to those found in literature [7].

The results presented in Table 16 indicated that the performance of a non-invasive fibrosis staging cannot be determined only by the discrimination performance of adjacent fibrosis stages. Fibrosis 1 patients were classified mostly as belonging to fibrosis 0. Fibrosis 3 patients were classified mostly as fibrosis 4. However, there are a significant number of fibrosis stage 1 patients (34% of the entire F1 lot) that were unexpectedly classified as fibrosis stage 4. The same fact can be observed in fibrosis stage 3 lot.

Liver stiffness measured using transient elastography is one of the most significant feature. It has the highest rank (see Tables 4-9) confirming the findings in literature.

In this paper were used two set of features. The second set was built by keeping attributes that can be evaluated without the need of specialized equipment or with access to a clinical laboratory. The fibrosis detection and staging performance was lower when using only these features. However, the greatest advantage is that these features could be acquired by any clinician in one session. They are cheap to acquire because portable Doppler capable ultrasound machines are relatively common. Future experiments in this direction could identify a simple and reliable method to screen the patients and to monitor them along the evolution of the disease. In [6] authors suggested using the non-invasive tests in order to measure the changes within individual patients because it is possible that the limited performance in predicting lower fibrosis stages are inherited from the biopsy interpretation errors. A longitudinal study, in which patients are investigated during the evolution of the disease could offer

a better validation method of the non-invasive fibrosis detection tests.

Some of the features, such as liver capsule regularity prove to be very powerful. In Table 4 through Table 9 one can see that the liver homogeneity and liver capsule scores have a high relevance. In the present paper these scores were assessed by a human expert. Work still has to be done in order to establish a procedure that can give objective measures to these scores. In the case of other diffuse liver disease, liver steatosis, there was significant progress in quantifying such visual scores [15].

6. Conclusion

The present paper showed the fact that is possible to build a classification schema that can be used in the current clinical practice in order to predict severe fibrosis.

New research directions were identified, automatic quantification of relevant visual scores.

Acknowledgement

Part of this work was funded by the National Council for Scientific Research in Higher Education, Grant nr. 41-071/2007: SONOFIBROCAST.

References

- [1] P. Bedossa, D. Dargere, and V. Paradis, Sampling variability of liver fibrosis in chronic hepatitis C, *Hepatology*, 38(6), 2003, 1449-1457.
- [2] M. Hirata, S.M.F. Akbar, N. Horiike, and M. Onji, Noninvasive diagnosis of the degree of hepatic fibrosis using ultrasonography in patients with chronic liver disease due to hepatitis C virus, *European Journal of Clinical Investigation*, 31(6), 2001, 528-535.
- [3] M. Zheng, W.M. Cai, H.L. Weng, and R.H. Liu, ROC curves in evaluation of serum fibrosis indices for hepatic fibrosis, *World J Gastroenterol*, 8(6), 2002, 1073-1076.
- [4] C.T. Wai, J.K. Greenon, R.J. Fontana, J.D. Kalbfleisch, J.A. Marrero, H.S. Conjeevaram, and A.S.F. Lok, A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C, *Hepatology*, 38(2), 2003, 518-526.
- [5] J.A. Talwalkar, D.M. Kurtz, S.J. Schoenleber, C.P. West, and V.M. Montori, Ultrasound-based transient elastography for the detection of hepatic fibrosis: systematic review and meta-analysis, *Clin Gastroenterol Hepatol*, 5(10), 2007, 1214-1220.
- [6] I. Guha and W.M. Rosenberg, Noninvasive Assessment of Liver Fibrosis: Serum Markers, Imaging, and Other Modalities, *Clin Liver Dis*, 12(4), 2008, 883-900.
- [7] S. Bonekamp, I. Kamel, S. Solga, and J. Clark, Can imaging modalities diagnose and stage hepatic fibrosis and cirrhosis accurately?, *J Hepatol*, 50(1), 2009, 17-35.
- [8] Z. Jiang, K. Yamauchi, K. Yoshioka, K. Aoki, S. Kuroyanagi, A. Iwata, J. Yang, and K. Wang, Support vector

C. Vicas, S. Nedevschi, M. Lupsor, and R. Badea, "Detection and Staging of Liver Fibrosis Using Additive Logistic Models," in Eleventh IASTED Int. Conf. Artificial Intelligence and Applications, Innsbruck, Austria, 2011, pp. 95-100.

DRAFT

machine-based feature selection for classification of liver fibrosis grade in chronic hepatitis C, *J Med Syst*, 30(5), 2006, 389-94.

[9] N. Landwehr, M. Hall, and E. Frank, Logistic model trees, *Mach Learn*, 59(1-2), 2005, 161-205.

[10] J. Friedman, T. Hastie, and R. Tibshirani, Additive logistic regression: A statistical view of boosting, *Ann Stat*, 28(2), 2000, 337-374.

[11] I. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques* (Morgan Kaufmann Pub, 2005).

[12] M. Pepe, *The statistical evaluation of medical tests for classification and prediction* (Oxford: Oxford University Press, 2003).

[13] S. Macskassy, F. Provost, and S. Rosset, ROC confidence bands: an empirical evaluation, *Proceedings of the 22nd international conference on Machine learning*, Bonn, Germany, 2005, 537-544

[14] Y. Bengio and Y. Grandvalet, No unbiased estimator of the variance of k-fold cross-validation, *The Journal of Machine Learning Research*, 5(2004, 1089-1105.

[15] C. Vicas, S. Nedevschi, M. Lupsor, and R. Badea, Automatic detection of liver capsule using Gabor filters. Applications in steatosis quantification, *IEEE Proc.Intelligent Computer Comm. Process.*, 2009, 133-140.