Fibrosis detection using ultrasound and serological markers as features for additive logistic regression

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

Monica Lupsor, Radu Badea
3rd Medical Clinic 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 the fibrosis. Present study tries to identify novel possibilities for noninvasive fibrosis evaluation. We included 591 patients. From each patient a number of 93 features were obtained using B-mode ultrasound, Doppler ultrasound, transient elastography and serum markers. The patients were grouped in five two-class problems that were presented to two classification schemas, one based on Support Vector Machines and the other based on logistic regression. Cross-validation along with area under curve (AUROC) is used to measure the classification performance. The AUROC of 0.90 was recorded when discriminating between fibrosis stage ≤3 and fibrosis stage 4.

1. Introduction

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

There are several noninvasive alternative methods to investigate the liver: ultrasound investigation including simple B-mode and Doppler mode, current serum markers used to investigate 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 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] investigates the the serum levels of hyaluronic acid, type III procollagen, N-terminal procollagen etc. in patients with hepatic fibrosis. The authors used 114 serum samples from biopsied patients divided in two groups accroding to their fibrosis stage (S≥1). They computed an AUROC of 0.8 for procollagen III peptide marker.

The aspartate aminotransferase to platet ratio index is proposed in [4] as an nonivasive index to predict significant fibrosis and cirrhosis. Using one training set containing 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.

Support Vector Machines along with the sequential forward floating selection scheme was used in [5] to discriminate between patients with no or mild fibrosis (F≤1) and patients with significant fibrosis (F≥2). 204 patients were investigated using liver biopsy (METAVIR score) and 34 serum markers. The mild fibrosis lot has 86 patients and the severe fibrosis lot 118 patients. The authors used leave one out cross validation and four selected serum markers (total bilirubin, platelet count, prothrombin time, hyaluronic acid). They achieved an accuracy of 96% in predicting severe fibrosis and 79% in predicting mild fibrosis. The SVM parameters were fixed during the experiments. No special measures were taken to compensate for unbalanced datasets.

Ultrasound based transient elastography is a noninvasive procedure that determines the stiffness of the liver tissue. The liver tissue stiffness is associated with the degree of fibrosis. In [6] the authors perform a review on the performance of transient elastography in detecting liver fibrosis. Their conclusion is that transient elastography is clinically useful in assessing

the presence or the absence of cirrhosis. Discriminating between other fibrosis stages is still a challenging task.

In another paper [7] authors review a series of noninvasive tests including various commercially available serum kits, transient elastography and blood flow investigation. Combining biomarkers that measure different components of the fibrosis may advance the diagnostic accuracy. The authors also note that imperfect golden standard might impose some limitations in developing a noninvasive test. This could be compensated by developing longitudinal tests.

In [8] the authors review 153 papers that approach the non invasive fibrosis detection. They included all papers referring to imaging modalities and use liver biopsy as reference. The authors investigate B-mode ultrasonography, Doppler ultrasonography, contrast-enhanced ultrasonography, sonographic elasiticity, transient elastography, tissue strain imaging, supersonic shear imaging, magnetic resonance imaging, contrast-enhanced MRI, diffusion-weighted MRI, magnetic resonance elastography, magnetic resonance spectroscopy, computed tomography, positron emission tomography, and single photon emission computed tomography. They conclude that the diagnostic performances of all described noninvasive radiologic modalities were better in distinguishing patients with cirrhosis from lesser degrees of fibrosis. However, staging of fibrosis was rarely achieved reliably. Transient elastography appears to be the most promising technique.

In our previous work [9] we tried to determine a method of detecting fibrosis using image processing algorithms applied on B-mode ultrasound images. In [9] the basic idea was to quantify the liver homogeneity by the means of texture descriptors. The results indicated the possibility of detecting fibrosis by means of non invasive investigations.

In this paper we tried to include most of the noninvasive features analyzed in [1]-[7]. A number of 93 features are acquired from 591 patients. These features include transient elastography evaluation, several common markers, B-mode ultrasound to measure various liver/spleen dimensions and Doppler measurements to evaluate the blood flow. METAVIR score is used to assess the fibrosis stage.

Fibrosis detection problem was examined from various points of view by grouping the patients in five different 2-class problems. Using logistic regression we determined the most relevant features in each case.

The additive logistic models employed here are able to perform an automatic attribute selection. We studied the impact of different patient distributions over the selected features.

From the original dataset, 42 features were used to build a reduced feature set with the goal to identify the power of the features that can be measured using only B-mode ultrasound and Doppler. Again we investigated the power of selected features with respect to changes in the dataset.

Original contributions of this paper are that a large variety of noninvasive features are included simultaneously in the study. Classification algorithms were used to predict the fibrosis stage. Models for fibrosis detection are built. These models are studied in order to identify the relevant features. We also investigated the power of relevant features regarding changes in the dataset.

The paper is structured as following: Chapter 2 briefly describes the two classification algorithms and the evaluation methodology. Chapter 3 introduces the features recorded for each patient. Chapter 4 presents the experimental results and in Chapter 5 conclusions to this paper are drawn.

2. Classification algorithms

There are two classification schemas used. Support Vector Machines and a schema based on logistic regression.

Mean imputation is used to infer values for missing attributes. For each attribute the mean is computed using existing values in the training set.

The missing values are replaced with the computed mean. If there are missing values in the test set, these values are replaced with previously computed mean. Normalization is performed on each training set. The normalization parameters computed on train set are used to normalize the test set.

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. For each training fold missing values are replaced using mean imputation. Normalization is then performed and the resulting dataset is used as training set for each classification schema.

The predictions on the unseen test set are collected. After the cross-validation schema is ended, on the collected predictions is computed the Area Under Curve (AUROC).

The results from each fold are used to compute the standard deviation. These values are averaged over 10 runs using different random cross validation splits.

Both classification schemas receive the same pair of random training and testing set generated by the cross-validation fold.

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

$$ci = 1.96\frac{\sigma}{\sqrt{n}}, \quad (1)$$

where σ is the standard deviation and n is the total number of samples. In our case, n is 10. We assume that the results follow a normal distribution.

Additive logistic regression performs an implicit attribute selection.

Using the entire dataset, such a classifier is trained and the selected attributes are observed.

2.1. Support Vector Machines

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

In the dataset is possible that the classes are not equally distributed. In order to compensate for the unbalanced dataset we alter the cost for each class.

Suppose that we have a dataset with two classes, each class has the volume n_1 and n_2. The cost weight for each class is c_1 and c_2. According to [10] the costs should be in the relation:

$$\frac{c_1}{c_2} = \frac{n_2}{n_1}. \quad (2)$$

Let $n_2 > n_1$ and $c_2 = 1$. Then, $c_1 = \frac{n_2}{n_1}$.

A radial basis function kernel is used. The cost coefficient and the gamma parameter for the kernel are determined using a grid search strategy [11].

2.2. Additive Logistic Model

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

The logistic regression model is shown in Equation 3.

$$p(x) = P(y = 1 \mid x) = \frac{e^{f(x)}}{1 + e^{f(x)}}, \quad (3)$$

where

$$f(x) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i. \quad (4)$$

Fitting a logistic regression model means estimating the parameter vector $\beta=(\beta_0, \beta_1, \ldots, \beta_n)$. The standard procedure in statistics is to search the maximum likelihood [12].

There are iterative algorithms that asymptotically approach the maximum likelihood solution.

An additive logistic model has the form presented in Equation 5

$$F_M(x) = \sum_{m=1}^{M} f_m(x), \quad (5)$$

where $f_m(x)$ represents a logistic regression model.

Friedman et al presents in [13] 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 z_k 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 $F(x)$.

The *LogiBoost* algorithm:

1) Start with weights $w_k=1/K$, $F(x)=0$ and probability estimates $p(x_k)=0.5$

2) Repeat for $m=1,2, \ldots, M$:

a) Compute the working response and weights

$$z_k = \frac{y_k - p(x_k)}{p(x_k)(1 - p(x_k))}, \quad (6)$$

$$w_k = p(x_k)(1 - p(x_k)). \quad (7)$$

b) Fit the function $f_m(x)$ by a weighted least square regression of z_k to x_k, using weights w_k:

$$\beta_m = \arg\min_{\beta} \sum_{k=1}^{K} (f_m(w_k x_k) - z_k)^2. \quad (8)$$

c) Update $F(x) \leftarrow F(x) + 0.5 f_m(x)$ and $p(x)$ using

$$p(x) = \frac{e^{F(x)}}{e^{-F(x)} + e^{F(x)}}. \quad (9)$$

3) Output the classifier

$$sign[F(x)] = sign[\sum_{m=1}^{M} f_m(x)]. \quad (10)$$

where K is the number of instances in the training set, x_k represents an instance, y_k its class and M are the number of iterations.

In special case that $f_m(x)$ 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 [13] 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, 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 evaluated using a 5 fold cross-validation loop [13].

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

In present paper, the *SimpleLogistic* algorithm is used. One of the greatest advantages of this algorithm is that it performs an intrinsic attribute selection.

3. Biological and ultrasound features

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

Table 1. Examples of features used in this paper. From each category only several features are listed.

Category	Example of features	Total number of features
Patients 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 avgeraged mean velocity in hepatic artery, hepatic artery peak systolic velocity, flow acceleration in hepatic arter	30
Transient elastography	liver stiffness	3

Another classification of the features could be achieved by the means of their operator dependent variability. Some examinations are objective measurements like serum markers, other are measurements performed by a human expert using an ultrasound scanner (liver dimensions, blood flow measurements, etc) and other are subjective scores established by the same human expert (liver capsule regularity, liver homogeneity, etc). In present paper all features were treated equally. The measurement unit of each feature was kept constant throughout the experiments. However, the exact measurement unit is not important because each feature undergoes a normalization process. Another important characteristic of the acquired data is that there are missing values.

In average, each feature is missing in 109.7 instances (standard deviation 119.3) and in each instance there are in average 17.2 features missing (standard deviation of 12.6)

The missing attributes are 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 are missing from the healthy patients. As a result, the logical choice to replace missing values is 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.

4. Experimental results

A number of 591 patients were included in present study. 62 patients are healthy patients or have fibrosis stage 0 at liver biopsy. 187 were diagnosed with fibrosis stage 1, 183 with fibrosis stage 2, 84 with fibrosis stage 3 and 75 with cirrhosis. These patients were divided in five 2-class problems according to Table 2. Each of these five 2-class problems was presented to the two classification schemas. There were two main data sets, first containing the full spectra of 93 features and another dataset containing only 42 features obtained by selecting only certain features. The selection criteria will be specified later in this chapter.

4.1. Results for the full dataset

The recorded mean AUROC and confidence intervals are summarized in Table 3. In these experiments the full dataset was used. The additive logistic regression model allows for an automated feature selection. Running this algorithm for each 2-class problem the logistic model for the second class was noted. In Table 4-8 are enumerated the features and the β_i parameter.

In the cases where the generated model has many features, only first most relevant 10 values are shown. The relevance of a feature is given by $|\beta_i|$. The greatest $|\beta_i|$ for a feature, the greatest is its importance in the model. A β_i coefficient with a positive value for a feature means that the increased value for that feature predicts the class for the model. On the other hand, a negative coefficient means that a lower feature value will predict the class.

Table 2. The patients are grouped into five two-class problems. In the last column are shown the class volume and the weights for SVM cost parameter.

Description	Abbreviation	Volume/SVM weights
Healthy vs cirrhotic patients	F0/F4	62;75 1.2;1
Mild fibrosis (F<2) vs severe fibrosis (F>2)	F01/F34	249;159 1;1.56
Non cirrhotic (F<4) vs cirrhotic (F=4)	F0123/F4	516;75 1;6.88
Medium fibrosis (F<3) vs severe fibrosis (F≥3)	F012/F34	432;159 1;2.71
Mild fibrosis (F<2) vs the rest of the dataset (F≥2)	F01/F234	249;342 1.37;1

Table 3. Mean AUROC results and the confidence interval. The confidence intervals are computed using Equation 1.

Dataset abbreviation	Support Vector Machines		Additive logistic model	
	Mean AUROC	Confidence intervals	Mean AUROC	Confidence intervals
F0/F4	0.90	±0.05	0.97	±0.03
F01/F34	0.84	±0.04	0.90	±0.04
F0123/F4	0.83	±0.05	0.90	±0.05
F012/F34	0.78	±0.04	0.86	±0.04
F01/F234	0.71	±0.04	0.77	±0.04

Table 4. Additive logistic model for F0/F4 dataset. The model predicts class F4.

Rank	Feature	β_i
1	Pulsatility index in hepatic artery	-28.89
2	Liver homogeneity	14.93
3	Aspartate aminotransferase	12.13
4	Cholesterol	-5.18
5	Age	3.95
6	Protrombine index	-2.92
7	Splenic vein diameter	2.78
8	Caudate lobe diameter	2.42
9	Alanine aminotransferase	2

One should note that in tow class problems, the model for the first class is equal with the negated model for the second class. In the F0/F4 logistic model Pulsatility index in hepatic artery has a negative coefficient in the model. This means that an increased value for pulsatility index will predict the first class (F0). A decreased value for the index will predict the second class (F4). The aspartate aminotransferase feature has a positive coefficient in the same model. An increased aspartate aminotransferase value will predict the second class (F4) and a lower value will predict the first class (F0). The same reasoning can be applied to all the features included in presented models.

We must stress that the models presented in tables 4-8 were built on the entire dataset available and the exact actual performance of each feature might vary for another set of patients.

Table 5. Additive logistic model for F01/F34 dataset. The model predicts class F34.

Rank	Feature	β_i
1	Liver stiffness	8.29
2	Spleen area	4.32
3	Quik time	3.34
4	Aspartate aminotransferas	2.94
5	Systolic acceleration time in hepatic artery	-2.85
6	Activated partial thromboplastin time	-2.3
7	Proteins	2.1
8	Cholesterol	-1.8
9	Left lobe diameter	1.77
10	Hepatic artery peak systolic velocity	1.67

Table 6. Additive logistic model for F0123/F4 dataset. The model predicts the class F4

Rank	Feature	β_i
1	Liver stiffness	3.73
2	Splenic longitudinal diameter	2.66
3	Proteins	2.46
4	Splenic vein diameter	2.25
5	International normalized ratio	1.81
6	Gallbladder wall thickening	1.62
7	Liver capsule regularity	1.3

Table 7. Additive logistic model for F012/F34 dataset. The model predicts the patients having fibrosis stage ≥3.

Rank	Feature	β_i
1	Creatinine	-68.78
2	Liver stiffness	5.83
3	splenic vein diameter	2.41
4	Aspartate aminotransferase	2.17
5	Proteins	2.14
6	Mean corpuscular volume	1.98
7	Gallbladder wall thickening	1.89
8	Collateral circulation	-1.82
9	Gallbladder cholesterol polyps	1.8
10	Age	1.62

In order to evaluate the performance of each feature with respect to the dataset variability the following experiment was performed for each five problems: 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 rank the attributes based on the number models in which the feature is selected. A ranking of ten for a feature means that the feature was selected in all of the 10 models. For each feature, we computed the mean and standard deviation for β_i

coefficient. In Table 9-13 are shown the results. Only features with ranking at least 8 were shown.

Table 8. Additive logistic model for F01/F234 dataset. The model predicts the class containing the patients having fibrosis stage ≥ 2

Rank	Feature	β_i
1	Liver stiffness	3.83
2	Hepatic artery blood flow	2.51
3	Mean corpuscular hemoglobin concentration	-2.02
4	Mean corpuscular hemoglobin	-1.97
5	Bilirubine	1.68
6	Hepatic artery blood flow	1.42
7	Cholesterol	-1.27
8	Age	1.21
9	Time averaged maximum velocity in hepatic artery	1.11
10	White cells	-1.01

Table 9. 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 10. 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 11. 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

The best discriminated case was 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 in Table 4 we try to predict the other patients, having fibrosis stage 1, 2 and 3. In Table 14 are shown the results. 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 1 patients (34% of the entire F1 lot) that were unexpectedly classified as fibrosis stage 4. The same fact can be observed in F3 lot. These results indicate that the performance of a noninvasive fibrosis staging cannot be determined only by the discrimination performance of adjacent fibrosis stages.

Table 12. 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 13. 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 14. 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. Results for reduced dataset

We try to investigate the power of the features that can be measured using B-mode ultrasound and Doppler. We excluded biochemical measurements and transient elastography. From patient characteristics we included only age and sex.

We also excluded visual scores like liver homogeneity, capsule regularity and hepatic angles. The remaining dataset consists of 42 features.

The patients were divided in the same five 2-class problems. Only additive logistic model was evaluated. The results are shown in Table 15. As in previous experiment we tried to identify the ranking of the features by performing 10 random splits of the dataset

and recording the selected features in each model. The results are enumerated in tables 16-20.

Table 15. Mean AUROC results and the confidence interval when using the reduced set of features. Only the results from additive logistic regression are shown. The confidence intervals are computed using Equation 1.

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

Table 16. 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 17. 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

Table 18. 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

5. Discussions and conclusions

The mean AUROC's values enumerated in Table 3 represent the estimated performance of the classification schemas employed here. We must stress that the models presented in tables 4-8 were built on the entire dataset available and the exact actual performance of these models is unknown. The actual performance of the models listed in tables 4 through 8 can be assessed only by examining new patients. The performance of each individual feature is influenced by random coincidences in patient distribution, by the missing attributes and other particularities. For example, in Table 7 the creatinine biochemical feature has a $|\beta_i|$ coefficient of 68.7 far greater than any other feature in the model. However, when studying the ranking of features with respect to modifications in patient distributions (Table 12) one can note that the creatinine no longer appears as a significant feature. When looking at the ranking results we found that it was selected as a feature in only one model.

Table 19. 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 20. 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

Each of the 93 features in the database should be scrutinized in order to remove outliers, to remove as much as possible the missing values (even with the risk of removing patients from the study) and to identify and correct eventual errors. When taking a closer look to creatinine value one could note that there are several outliers. These values affect the normalization process. There are other features which exhibit this behavior. It was a medical decision not to remove those patients because the attribute values were plausible. Other alternatives to handle these features are standardization, where the feature values are scaled in such a way that they follow a normal distribution of 0 mean and standard deviation equal to 1. Another approach is to logarithmically scale the values.

Liver stiffness measured using transient elastography is one of the most significant feature. It has the highest rank (see tables 10-13). In [6],[7],[8] the authors note that transient elastography could be the modality with greatest chance to be used as a

noninvasive fibrosis assessment method. In Table 15 the results are significantly lower when using reduced feature set (the one that doesn't contain the transient elastography measurements).

The first two classification problems, namely F0/F4 and F01/F34 have a theoretical significance, because the real patient spectra are continuous. However, they show that there are detectable changes between healthy and cirrhotic patients in our datasets.

Three out of five classification problems include all available patients. The difference between them is the splitting point.

These models could be applied in clinical practice. They distinguish between cirrhotic/non cirrhotic patients (F0123/F4) with relatively high performance.

One should keep in mind that this performance is an estimated one. More patients have to be investigated in order to confirm this performance and to validate these results.

The discrimination performance decreases as the split point is between lower fibrosis stages. These findings are similar to those found in literature [8].

No distinction was made between objective features (like measurements) and qualitative features like visual scores. Each of the relevant features must undergo a scrutiny to determine the link to the fibrosis mechanism and to determine their operator dependent variability.

Some of these features, like liver capsule regularity prove to be very powerful. In Tables 9-13 one can see that the liver homogeneity and liver capsule scores are relatively high as relevance. In present paper this homogeneity was assessed using a visual score established by a human expert. Work has to be done in order to establish a procedure that can give objective measures for these scores.

Present paper show the fact that is possible to build classification schema that can be used in current clinical practice in order to predict severe fibrosis.

There are few directions that can improve the detection rates. First, more patients have to be examined. The most relevant features identified here should be recorded for each patient. Reducing the missing data in the future datasets will increase the accuracy of the performance evaluation.

In [7] authors suggest using the noninvasive 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, where patients are investigated during the evolution of the disease could offer a better validation method of the noninvasive fibrosis detection tests.

In [9] textural descriptors were used to quantify the liver homogeneity with relative success in predicting fibrosis. Image processing techniques could be further employed to eliminate the human expert in establishing the visual scores in case of liver capsule regularity.

6. References

[1]P. Bedossa, D. Dargere, V. Paradis, "Sampling Variability of Liver Fibrosis in Chronic Hepatitis C," *Hepatology*, Vol. 38, No. 6, pp: 1449-1457, 2003

[2]M. Hirata, S.M. Akbar, N. Horiike, M. Onji, "Noninvasive diagnosis of the degree of hepatic fibrosis using ultrasonography in patients with chronic liver disease due to hepatitis C virus," *Eur J Clin Invest.*, Vol. 31, No. 6, pp:528-35, 2001

[3]M. Zheng, W.-M. Cai, H.-L. Weng, R.-H. Liu, "ROC curves in evaluation of serum fibrosis indices for hepatic fibrosis," *World J. Gastroenterol*, Vol. 8, No. 6, pp:1073-1076, 2002

[4]C.-T. Wai, J.K. Greenson, R.J. Fontana, J.D. Kalbfleisch, J.A. Marrero, H.S. Conjeevaram, A. S.-F. Lok, "A Simple Noninvasive Index Can Predict Both Significant Fibrosis and Cirrhosis in Patients With Chronic Hepatitis C," *Hepatology*, Vol. 38, No. 2, pp: 518-526, 2003

[5]Z. Jiang, K. Yamauchi, K. Yoshioka, K. Aoki, S. Kuroyanagi, A. Iwata, J. Yang, K. Wang, "Support Vector Machine-Based Feature Selection for Classification of Liver Fibrosis Grade in Chronic Hepatitis C," *J. Med. Syst.*, vol. 30, pp. 389-394, 2006.

[6]J. A. Talwalkar, D.M. Kurtz, J.S. Schoenleber, P.C. West, V.M. Montori, "Ultrasound-Based Transient Elastography for the Detection of Hepatic Fibrosis: Systematic Review and Meta-analysis," *Clin. Gastroenerol. Hepatol.*, vol. 5, pp: 1214-1220, 2007

[7]I.N. Guha, W.M. Rosenberg, "Noninvasive Assessment of Liver Fibrosis: Serum Markers, Imaging, and Other Modalities," *Clin. Liver Dis.* vol. 12, pp: 883-900, 2008

[8]S. Bonekamp, I. Kamel, S. Solga, J. Clark, "Can imaging modalities diagnose and stage hepatic fibrosis and cirrhosis accurately," *J. Hepatology* vol. 50, pp: 17-35, 2009

[9]S. Nedevschi, C. Vicas, M. Lupsor, R. Badea, M. Grigorescu, "Usefulness of image processing techniques in assessing the diffuse liver diseases," in *Proc. ICCP 2008, 2nd Workshop Computers in Medical Diagnosis*, Cluj-Napoca, pp.10-15, Aug., 2008.

[10]C. W. Hsu, C. C. Chang, C. J. Lin. (2008, Oct. 2). "A practical guide to support vector classification," [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin

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

[12]N. Landwehr, M. Hall, E. Frank, "Logistic Model Trees," *Mach. Learn.*, vol. 59, pp. 161-205, May 2005.

[13]J. Friedman, T. Hastie , R. Tibshirani, "Additive Logistic Regression: a Statistical View of Boosting," *Ann. Statist.*, vol. 28, no. 2, pp.337-407, 1998