Online Submissions: wjg.wjgnet.com
wjg@wjgnet.com
doi:10.3748/wjg.15.2617

World J Gastroenterol 2009 June 7; 15(21): 2617-2622
World Journal of Gastroenterology ISSN 1007-9327

Effective use of FibroTest to generate decision trees in hepatitis C

Dana Lau-Corona, Luís Alberto Pineda, Héctor Hugo Avilés, Gabriela Gutiérrez-Reyes, Blanca Eugenia Farfan-Labonne, Rafael Núñez-Nateras, Alan Bonder, Rosalinda Martínez-García, Clara Corona-Lau, Marco Antonio Olivera-Martínez, Maria Concepción Gutiérrez-Ruiz, Guillermo Robles-Díaz, David Kershenobich

Dana Lau-Corona, Gabriela Gutiérrez-Reyes, Blanca Eugenia Farfan-Labonne, Rafael Núñez-Nateras, Alan Bonder, Rosalinda Martínez-García, Maria Concepción Gutiérrez-Ruiz, Guillermo Robles-Díaz, David Kershenobich, Department of Experimental Medicine, School of Medicine, Universidad Nacional Autonoma de Mexico, General Hospital of Mexico, Mexico City 06726, Mexico
Luís Alberto Pineda, Héctor Hugo Avilés, Department of Computer Sciences, Institute for Applied Mathematics and Systems, Universidad Nacional Autonoma de Mexico, Mexico City 01000, Mexico
Clara Corona-Lau, Marco Antonio Olivera-Martínez, Department of Gastroenterology, Lomas Altas Clinic, Mexico City 11950, Mexico
Author contributions: Lau-Corona D, Pineda LA, Gutiérrez-Reyes G, Farfan-Labonne BE, Gutiérrez-Ruiz MC, Robles-Díaz G and Kershenobich D designed the research; Pineda LA and Avilés HH performed and analyzed the decision trees; Olivera-Martínez MA and Kershenobich D were responsible for patient enrollment; Lau-Corona D, Núñez-Nateras R, Bonder A and Martínez-García R created and analyzed the patients database; Corona-Lau C performed the FibroTests; Lau-Corona D, Pineda LA and Kershenobich D wrote the manuscript.
Supported by A grant of the Universidad Nacional Autonoma de Mexico SDI.PTID.05.6
Correspondence to: David Kershenobich, MD, PhD, Professor of Medicine, Chief, Department of Experimental Medicine, School of Medicine, UNAM, General Hospital of Mexico, Mexico City 06726, Mexico. kesdhipa@yahoo.com
Telephone: +52-55-56232673 Fax: +52-55-57617651
Received: March 17, 2009 Revised: May 6, 2009
Accepted: May 13, 2009
Published online: June 7, 2009

Abstract

AIM: To assess the usefulness of FibroTest to forecast scores by constructing decision trees in patients with chronic hepatitis C.

METHODS: We used the C4.5 classification algorithm to construct decision trees with data from 261 patients with chronic hepatitis C without a liver biopsy. The FibroTest attributes of age, gender, bilirubin, apolipoprotein, haptoglobin, α2 macroglobulin, and γ-glutamyl transpeptidase were used as predictors, and the FibroTest score as the target. For testing, a 10-fold cross validation was used.

RESULTS: The overall classification error was 14.9% (accuracy 85.1%). FibroTest's cases with true scores of F0 and F4 were classified with very high accuracy (18/20 for F0, 9/9 for F0-1 and 92/96 for F4) and the largest confusion centered on F3. The algorithm produced a set of compound rules out of the ten classification trees and was used to classify the 261 patients. The rules for the classification of patients in F0 and F4 were effective in more than 75% of the cases in which they were tested.

CONCLUSION: The recognition of clinical subgroups should help to enhance our ability to assess differences in fibrosis scores in clinical studies and improve our understanding of fibrosis progression.

Key words: Hepatitis C; FibroTest; Decision trees; C4.5 algorithm; Non-invasive biomarkers

Peer reviewer: Wasim Jafri, Professor, The Aga Khan University Hospital, Stadium Road, Karachi 74800, Pakistan

Lau-Corona D, Pineda LA, Avilés HH, Gutiérrez-Reyes G, Farfan-Labonne BE, Núñez-Nateras R, Bonder A, Martínez-García R, Corona-Lau C, Olivera-Martínez MA, Gutiérrez-Ruiz MC, Robles-Díaz G, Kershenobich D. Effective use of FibroTest to generate decision trees in hepatitis C. *World J Gastroenterol* 2009; 15(21): 2617-2622 Available from: URL: http://www.wjgnet.com/1007-9327/15/2617.asp DOI: http://dx.doi.org/10.3748/wjg.15.2617

INTRODUCTION

One of the most widely used non-invasive markers to stage liver fibrosis is the FibroTest (FT; BioPredictive, Paris, France), which involves the measurement of surrogate markers, α2 macroglobulin (A2M), haptoglobin, γ-glutamyl transpeptidase (γGT), total bilirubin, and apolipoprotein A1 (APO-A1), which,

in combination, have a high predictive value for the diagnosis of significant fibrosis[1-5]. The correlation of these markers with liver fibrosis involves a formula that was derived through logistic regression using data from 339 patients. The data were also used for the construction of a classifier using neural networks, but logistic regression was favored[1]. Its efficacy has been further validated by comparing the predictions made by the formula to those obtained with histological scoring of liver biopsies[6-10].

FT is employed to evaluate a complex situation under conditions of uncertainty, e.g. evaluating the degree of fibrosis in a patient with hepatitis C. A recent review of FT performance in 6549 patients and 925 controls supported the recommendation in clinical practice of FT as an alternative to liver biopsy for the first-line assessment of liver injury in patients with chronic hepatitis C. This review concludes that neither biomarkers nor biopsies are sufficient alone to provide the information necessary to make definitive decisions in a given patient, but rather, all the clinical and biological data must be taken into account[11].

Based on experience obtained, acquiring new information about the behavior of FT in clinical practice will be useful to assess changes to the evaluation of patients with liver fibrosis. In addition, there is a rich set of automatic classification techniques developed within the context of machine learning using Artificial Intelligence, which can be used to simplify the classification process and to provide additional information to support the classification rational. One such technique is the automatic generation of decision trees. Decision trees provide explicit rules to relate the range of values of the biomarkers with fibrosis scores, and they might help to gain a better grasp of the importance and significance of the test.

MATERIALS AND METHODS

Patients

A total of 261 patients with chronic hepatitis C, who were HCV RNA+, not receiving any antiviral or antifibrotic treatment, in whom a liver biopsy could not be obtained and who had been submitted to the FibroTest as part of their first evaluation profile, were included in the study. The patients were recruited from the liver unit of Clinica Lomas Altas, Mexico City from January 2003 to December 2007.

For each patient, we retrospectively gathered data on age, gender, γGT, ALT, AST, total bilirubin, hemoglobin, white cell counts and platelets. All the analytical studies were performed independently of the present study and their results had been reported previously. The interval between routine blood test and FT was less than 5 d.

FibroTest

The FibroTests were performed according to published recommendations. This method provides a quantitative estimate of liver fibrosis ranging from 0.00 to 1.00. The FT cutoffs for presumed fibrosis stages were 0.00-0.21 (F0), 0.22-0.27 (F0-F1), 0.28-0.31 (F1), 0.32-0.48 (F1-F2), 0.49-0.58 (F2), 0.59-0.72 (F3), 0.73-0.74 (F3-F4) and > 0.75 (F4)[12]. Each attribute (component) of the FibroTest was considered and included in the construction of decision trees.

Decision trees

Decision trees are a diagrammatic representation of a decision process, where nodes represent questions about attribute values or ranges of values, and edges represent the possible answers that link question nodes with other nodes down the tree, which represent further questions. Nodes at the bottom of the tree represent classes: the class of an object satisfying all the questions associated to the nodes in the path from the top question node to the bottom class node. In the case of FibroTest, each question node in the tree represents an FT biomarker (e.g. bilirubin and apolipoprotein A1) value or value interval, and the bottom nodes represent the FT scores (i.e. F0, F1, F1-F2, *etc*). In the present study, decision trees were constructed using the C4.5 classification algorithm[13,14]. The C4.5 Algorithm, often referred to as statistical classifier, is based on the concepts of information entropy and information gain. Intuitively, information entropy is the number of bits required to code an event (i.e. a random variable), where the higher the probability of the event the lower the number of bits required to code it. Information gain, in turn, is the reduction of entropy when additional information is available. C4.5 uses the fact that each attribute of data can be used to make a decision that splits the data into smaller subsets, which have reduced information entropy. The C4.5 algorithm is freely available, and for the purpose of this study we used the code supplied directly by Quinlan at http://www.rulequest.com/Personal/.

The simplified pseudo-code for the algorithm is as follows: (1) Find the most informative attribute (i.e. the one with the lowest entropy or the largest information gain) in relation to the set of samples provided. (2) Create a decision node that splits on the selected attribute; this node will have a decision question on an attribute's value or value interval, and will partition the samples in relation to such a value (yes/no) or value interval (i.e. the ones with a lower value, the ones in range, and the ones with a higher value.) If all the samples belong to the same partition, the corresponding node becomes a class node. (3) Create a daughter node for each remaining case, and apply the procedure from (1) for all samples that remain in the corresponding partition.

For the experiments, the following predictive attributes were used: age, gender, bilirubin, Apo A1, A2M, GGT, and haptoglobin. The target was the FibroTest score. The classifier was built with data from the 261 patients. The algorithm also selects the best rules (i.e. paths) from the trees, and makes a set of compound rules, which are tested against all samples in the test-data, and provide a confidence factor. In the present study we tested the classification performance with these compound rules and computed the corresponding confusion matrix.

In order to enhance the confidence of the classifier, we used all the data for training and also for testing

Table 1 Clinical and laboratory data of the study population

Patients with chronic hepatitis C (n = 261)	Mean (range/SD)
Age (yr)	52 (20-78)
Age at infection (yr)	26.74 (birth-69)
Time of progression (yr)	26.41 (2-69)
Hemoglobin (g/dL)	14.83 ± 2.54
Platelet count (10^3/mm^3)	203 ± 82
White cell count (10^3/mm^3)	5.39 ± 1.82
γGT (IU/L)	77 ± 81
ALT (IU/L)	96.4 ± 108.1
Bilirubin (mg/dL)	0.9 ± 0.6

Table 2 Fibrosis stage (FibroTest) distribution according to gender in the study population n (%)

FibroTest score	Female	Male	Total
F0	16	4	20 (8)
F0-F1	5	4	9 (3)
F1	5	4	9 (3)
F1-F2	29	18	47 (18)
F2	21	13	34 (13)
F3	26	15	41 (16)
F3-F4	3	2	5 (2)
F4	44	52	96 (37)
Total	149 (57)	112 (43)	261 (100)

through 10-fold cross-validation as follows: Partition the whole set of available empirical data in 10 randomly generated designated equal subsets; use 90% as train-data and the remaining designated 10% as test-data and compute the classifier's performance. The procedure is repeated using the remaining designated 10% partitions as test-data. The performance of the classifier was presented as the average of the 10 tests.

RESULTS

Demographic data

Of the 261 patients, there were 149 (57%) female and 112 (43%) male. Their mean age was 52 years (range 20-78 years). The mean age at infection was 26 years (range birth to 69 years). 75.1% were genotype 1. The average time from exposure to risk factor to their first FibroTest was 26.4 years. The mean values for the following parameters was as follows: Hb 14.8 ± 2.5 g/dL, platelets 203 000 ± 82 000 10^3/mm^3, leukocytes 5399 ± 1821 10^3/mm^3, serum bilirubin 0.9 ± 0.67 mg/dL, ALT 96.4 ± 108 IU/L, and GGT 77.5 ± 81.0 IU/L (Table 1).

FibroTest

The reported FT scores indicate that 45% of the patients (n = 117) had either F4 (37%) or F0 (8%). The remaining 55% (n = 144) had intermediate stages of fibrosis (Table 2).

Decision trees

The C4.5 algorithm was used to construct ten decision trees. The algorithm selected a number of rules relating attribute values with the fibrosis score and the percentage of times that each rule was successfully applied for each tree. In addition, the algorithm produced a set of 26 compound rules out of the ten classification trees and these rules were used to classify the 261 patients. The plausible rules for the classification of patients in F0 and F4 are shown in Table 3.

Table 3 Compound rules generated for classes F0 and F4 and the percentage of times that each rule was successfully applied

F0	F4
GGT ⩽ 108 IU/L	Bilirubin > 1.2 mg/dL
A2M ⩽ 280 g/L	GGT > 26 IU/L
Apo A1 > 144 g/L	A2M > 216 g/L
Age > 36 yr	Class F4 (95.6%)
Age ⩽ 53 yr	
Class F0 (79.4%)	
	A2M > 335 g/L
	Haptoglobin ⩽ 54.6 g/L
	Age > 53 yr
	Class F4 (95.5%)
Bilirubin ⩽ 1.1 mg/dL	
A2M ⩽ 243 g/L	
Apo A1 > 126 g/L	
Haptoglobin > 73.5 g/L	Gender = M
Age ⩽ 50 yr	Bilirubin > 0.5 mg/dL
Class F0 (77.7%)	A2M > 396 g/L
	Class F4 (91.7%)
	Bilirubin > 0.6 mg/dL
	Haptoglobin ⩽ 16.2 g/L
	Class F4 (90.9%)
	GGT > 40 IU/L
	A2M > 372 g/L
	Age > 59 yr
	Class F4 (90.2%)

Table 4 Confusion matrix relating real FT scores (rows) to predicted FT scores (columns) in 261 patients with chronic hepatitis C

	F0	F0-F1	F1	F1-F2	F2	F3	F3-F4	F4
F0	18					2		
F0-F1		9						
F1			7			2		
F1-F2	2			37	1	7		
F2				3	18	13		
F3						38		3
F3-F4						1	3	1
F4			1			3		92

FT cases with true scores of F0 and F4 were classified with very high accuracy (18/20 for F0 and 92/96 for F4), which indicated that in the extreme stages of fibrosis, decision trees produced the correct classification in approximately 93% of the cases, as shown in the confusion matrix in Table 4.

The overall classification error was 14.9% (accuracy 85.1%). We observed that the largest confusion relates to false positives in F3 (43.7%). However, the chances that the predicted value is right for all FT scores, except F3, are quite high.

DISCUSSION

The FT scores used to evaluate the degree of liver fibrosis were generated using a formula obtained

through logical regression that can be thought of as a classification device or classifier. From this perspective the FT formula is a black box that has the FT attributes as inputs and produces the corresponding FT score as its associated output. However, it would be convenient to be able to look into the internal classification process and have access to the classification rational. In addition, it is also important to reinforce the reliability of the FibroTest and ensure that the classification results are independent of contingent features of the classification technique. Indeed, in the original formulation of the FibroTest, the use of neural networks, another black box classification technique, was also explored but logistic regression was preferred for clarity[1].

As with all machine learning techniques, decision trees such as those employed in the present work, are deduced from empirical data and the success of a given application depends on the quantity and quality of these data. In this respect, these algorithms are analogous to statistical regression techniques, such as logical regression, but rely on classification heuristics. They have proved to behave well not only in linear problems, but also in non-linear or unstable domains. For this, classifiers need to be built with a portion of the data, which is usually called the "train-data", and tested with a different portion of the data, which is usually called the "test-data", and it is essential that these two sets are distinct. For the induction process proper, the values of the attributes of each sample, "the predictors", are associated with its corresponding class, "the target", and the process is repeated iteratively for all the samples in the train-data. At the end of this process, each class is associated with a combination of values or value intervals of the attributes. The classifier can then be used to predict the class of a sample not used in the training process. In particular the performance of the algorithm can be assessed by comparing the known class of each sample in the test-data with the class predicted by the decision tree for such sample. The specifics of these procedures, with the heuristics employed, give rise to a large variety of classification techniques, one of which is decision trees. Decision trees can be created through a diversity of algorithms and the field as a whole is quite mature and has been applied to a large diversity of application domains with very positive results. In particular, in the clinical setting, decision trees have been applied, for instance, to proteomic data analysis in pancreatic cancer[15], to the prediction of interferon treatment effects based on microarray gene expression profiles[16], and to the prediction of diagnosis and outcome of dengue fever based on clinical, hematological and virological data[17].

Ranking the seven attributes used by the FT, we generated a learning set that allowed for the determination of a second classifier using an ensemble of decision trees. To identify the decision algorithms, we used the C4.5 decision tree classifier, which has several advantages over other statistical tools. Indeed, decision algorithms so generated are simple to understand, and they are able to handle missing values. In contrast, logistic regression and discriminant analyses require much more data preparation and more extensive handling of missing values for reliable calculations[18]. Decision algorithms are also easy to interpret and validate using common statistical techniques, which facilitates their use to predict the diagnosis and prognosis in different clinical settings.

The decision tree analysis of our data produced a set of seven plausible rules that correctly predicted the fibrosis score in more than 75% of the cases for F0 and F4. Interestingly, the rules for predicting F4 were precise in 90% of the cases. Therefore, the rules generated herein can be considered as having great accuracy. These findings add support to the impression that fibrosis at the extremes of the disease is more predictable, a notion that applies both to non-invasive markers and liver biopsy[4,19].

Of the markers employed in the decision trees and the derived rules, the most relevant were age and α-2 macroglobulin as independent predictors. If a patient has an FT score of F0 with a normal A2M and is below 53 years of age, our results suggest the presence of mild disease, adding to the clinical decision of not performing a liver biopsy and holding back the timing and selection of antiviral treatment. On the other hand, if a patient with hepatitis C and FT score of F4 is above 53 years and has an increased A2M, a diagnosis of significant fibrosis is presumed.

Fibrosis progression tends to vary between patients and even in the same person, for reasons that are not yet understood. However, age has consistently been reported as an important risk factor for fibrosis progression in chronic viral hepatitis, either at the onset of the disease or during its evolution. The changes with age tend mostly to be subtle and are consistent with a disease of long duration associated with the progression of normal aging[20-23]. How and why variants arise probably relates to changes in extracellular matrix[24,25], liver regeneration[26] and repair mechanisms[27]. In this regard, and based on our results, A2M is an extremely useful attribute. It is a protease inhibitor and a major carrier of cytokines synthesized by hepatic stellate cells and hepatocytes. Furthermore, its expression might inhibit matrix remodeling during fibrosis[28,29].

Neither age nor A2M alone has been proven to be an adequate marker of fibrosis, which makes it important to apply a more comprehensive approach in the use of non-invasive markers for liver fibrosis. Undoubtedly, not knowing which markers are the most predictive has been one of the main obstacles impeding their integration into clinical practice and patients' management. Our study indicates that the combination of the markers used in the FT is reliable and performs well, independently of, age or gender. Although ethnicity was not an inclusion criteria for our study, all the patients were Hispanics with an age range of 20-78 years.

There are pitfalls and caveats to FT use, and the decision tree analysis was not able to generate accurate rules to predict intermediate FT scores (particularly F3), suggesting that the FT attributes for these particular

stages exhibit considerable noise or are inconclusive. This was reflected in the analysis of the confusion matrix in which F3 was the most ambiguous; nonetheless, our study showed that restricting the biopsies to the patients with intermediate scores (F1-F3) could have prevented liver biopsies in 42% of the patients while maintaining an accuracy level above 75%. This is a strong argument to include the use of a non-invasive marker of fibrosis, such as the FT, in the profile evaluation of a patient with chronic hepatitis C, if for any reason; a liver biopsy cannot be performed.

Outside of clinical trials, with the advancement of new laboratory techniques, such as PCR and non-invasive biomarkers, more patients are treated for chronic fibrotic liver diseases without a liver biopsy[30], which remains the best predictor but is not necessarily the gold standard.

Analysis, such as the one performed in the present work, could help to further classify preclinical subgroups and identify subclasses of rapid or slower fibrosers. This classification should enhance our ability to assess differences in fibrosis scores in clinical studies and improve our understanding of fibrosis progression.

This work was presented as a poster at the AASLD Liver Meeting in November 2008 in San Francisco, California.

COMMENTS

Background

One of the most widely used non-invasive markers to stage liver fibrosis is the FibroTest which involves the measurement of a set of surrogate markers that, in combination, have a high predictive value for the diagnosis of significant fibrosis.

Research frontiers

The FT score is obtained with a formula generated through logical regression that can be thought of as a classification device or classifier. It would be convenient to be able to look into the internal classification process and have access to the classification rational. In addition, it is also important to reinforce the reliability of the FibroTest and ensure that the classification results are independent of contingent features of the classification technique.

Innovations and breakthroughs

The authors used the automatic generation of decision trees to simplify the classification process and to provide additional information to support the classification rational of the FibroTest as a non-invasive marker of liver fibrosis. Decision trees provide explicit rules to relate the range of values of the biomarkers with fibrosis scores, and they might help in gaining a better grasp of the importance and significance of the test.

Applications

Analysis, such as the one performed in the present work, could help to further classify preclinical subgroups and identify subclasses of rapid or slower fibrosers. This classification should enhance our ability to assess differences in fibrosis scores in clinical studies and improve our understanding of fibrosis progression.

Terminology

Decision trees are diagrammatical representations of decision processes, where nodes represent questions about attribute values or range of values, and edges represent the possible answers that link question nodes with other nodes down the tree, which represent further questions. Nodes at the bottom of the tree represent classes: the class of an object satisfying all the questions associated to the nodes in the path from the top question node to the bottom class node.

Peer review

This is a timely, well written article on FibroTest.

REFERENCES

1 **Imbert-Bismut F**, Ratziu V, Pieroni L, Charlotte F, Benhamou Y, Poynard T. Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study. *Lancet* 2001; **357**: 1069-1075

2 **Poynard T**, Imbert-Bismut F, Munteanu M, Messous D, Myers RP, Thabut D, Ratziu V, Mercadier A, Benhamou Y, Hainque B. Overview of the diagnostic value of biochemical markers of liver fibrosis (FibroTest, HCV FibroSure) and necrosis (ActiTest) in patients with chronic hepatitis C. *Comp Hepatol* 2004; **3**: 8

3 **Boursier J**, Bacq Y, Halfon P, Leroy V, de Ledinghen V, de Muret A, Bourlière M, Sturm N, Foucher J, Oberti F, Rousselet MC, Calès P. Improved diagnostic accuracy of blood tests for severe fibrosis and cirrhosis in chronic hepatitis C. *Eur J Gastroenterol Hepatol* 2009; **21**: 28-38

4 **Shaheen AA**, Wan AF, Myers RP. FibroTest and FibroScan for the prediction of hepatitis C-related fibrosis: a systematic review of diagnostic test accuracy. *Am J Gastroenterol* 2007; **102**: 2589-2600

5 **Myers RP**, Benhamou Y, Imbert-Bismut F, Thibault V, Bochet M, Charlotte F, Ratziu V, Bricaire F, Katlama C, Poynard T. Serum biochemical markers accurately predict liver fibrosis in HIV and hepatitis C virus co-infected patients. *AIDS* 2003; **17**: 721-725

6 **Halfon P**, Bourliere M, Deydier R, Botta-Fridlund D, Renou C, Tran A, Portal I, Allemand I, Bertrand JJ, Rosenthal-Allieri A, Rotily M, Sattonet C, Benderitter T, Saint Paul MC, Bonnot HP, Penaranda G, Degott C, Masseyeff MF, Ouzan D. Independent prospective multicenter validation of biochemical markers (fibrotest-actitest) for the prediction of liver fibrosis and activity in patients with chronic hepatitis C: the fibropaca study. *Am J Gastroenterol* 2006; **101**: 547-555

7 **Rossi E**, Adams L, Prins A, Bulsara M, de Boer B, Garas G, MacQuillan G, Speers D, Jeffrey G. Validation of the FibroTest biochemical markers score in assessing liver fibrosis in hepatitis C patients. *Clin Chem* 2003; **49**: 450-454

8 **Sebastiani G**, Vario A, Guido M, Noventa F, Plebani M, Pistis R, Ferrari A, Alberti A. Stepwise combination algorithms of non-invasive markers to diagnose significant fibrosis in chronic hepatitis C. *J Hepatol* 2006; **44**: 686-693

9 **Wilson LE**, Torbenson M, Astemborski J, Faruki H, Spoler C, Rai R, Mehta S, Kirk GD, Nelson K, Afdhal N, Thomas DL. Progression of liver fibrosis among injection drug users with chronic hepatitis C. *Hepatology* 2006; **43**: 788-795

10 **Sène D**, Limal N, Messous D, Ghillani-Dalbin P, Charlotte F, Thiollière JM, Piette JC, Imbert-Bismut F, Halfon P, Poynard T, Cacoub P. Biological markers of liver fibrosis and activity as non-invasive alternatives to liver biopsy in patients with chronic hepatitis C and associated mixed cryoglobulinemia vasculitis. *Clin Biochem* 2006; **39**: 715-721

11 **Poynard T**, Morra R, Ingiliz P, Imbert-Bismut F, Thabut D, Messous D, Munteanu M, Massard J, Benhamou Y, Ratziu V. Biomarkers of liver fibrosis. *Adv Clin Chem* 2008; **46**: 131-160

12 FI-BROCHURE: The FibroTest-ActiTest-HCV FIBROSURE Investigator's Brochure. Available from: URL: http://www.biopredictive.com

13 **Quinlan JR**. C4.5: Programs For Machine Learning. San Mateo: Morgan Kaufmann Publishers, 1993: 1-32

14 **Witten IH**, Frank E. Data Mining: Practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann Publishers, 2005: 62-69

15 **Ge G**, Wong GW. Classification of premalignant pancreatic cancer mass-spectrometry data using decision tree ensembles. *BMC Bioinformatics* 2008; **9**: 275

16 **Huang T**, Tu K, Shyr Y, Wei CC, Xie L, Li YX. The prediction of interferon treatment effects based on time series microarray gene expression profiles. *J Transl Med* 2008; **6**: 44

17 **Tanner L**, Schreiber M, Low JG, Ong A, Tolfvenstam T, Lai YL, Ng LC, Leo YS, Thi Puong L, Vasudevan SG, Simmons

CP, Hibberd ML, Ooi EE. Decision tree algorithms predict the diagnosis and outcome of dengue Fever in the early phase of illness. *PLoS Negl Trop Dis* 2008; **2**: e196

18 **Kothari R**, Dong M. Decision trees for classification: A review and some new results. In: Pal SK, Pal A, eds. Pattern recognition: from classical to modern approaches. Singapore: World Scientific Publishing, 2001: 169-184

19 **Fontanges T**, Bailly F, Trepo E, Chevallier M, Maynard-Muet M, Nalet B, Beorchia S, Pillon D, Moindrot H, Froissart B, Slaoui M, Tinel X, Pradat P, Trepo C. Discordance between biochemical markers of liver activity and fibrosis (Actitest((R))-Fibrotest((R))) and liver biopsy in patients with chronic hepatitis C. *Gastroenterol Clin Biol* 2008

20 **Massard J**, Ratziu V, Thabut D, Moussalli J, Lebray P, Benhamou Y, Poynard T. Natural history and predictors of disease severity in chronic hepatitis C. *J Hepatol* 2006; **44**: S19-S24

21 **Thein HH**, Yi Q, Dore GJ, Krahn MD. Estimation of stage-specific fibrosis progression rates in chronic hepatitis C virus infection: a meta-analysis and meta-regression. *Hepatology* 2008; **48**: 418-431

22 **Terrault NA**, Im K, Boylan R, Bacchetti P, Kleiner DE, Fontana RJ, Hoofnagle JH, Belle SH. Fibrosis progression in African Americans and Caucasian Americans with chronic hepatitis C. *Clin Gastroenterol Hepatol* 2008; **6**: 1403-1411

23 **Hissar SS**, Kumar M, Tyagi P, Goyal A, Suneetha PV, Agarwal S, Rastogi A, Sakhuja P, Sarin SK. Natural history of hepatic fibrosis progression in chronic hepatitis C virus infection in India. *J Gastroenterol Hepatol* 2009; **24**: 581-587

24 **Gressner OA**, Rizk MS, Kovalenko E, Weiskirchen R, Gressner AM. Changing the pathogenetic roadmap of liver fibrosis? Where did it start; where will it go? *J Gastroenterol Hepatol* 2008; **23**: 1024-1035

25 **Henderson NC**, Forbes SJ. Hepatic fibrogenesis: from within and outwith. *Toxicology* 2008; **254**: 130-135

26 **Roskams T**. Relationships among stellate cell activation, progenitor cells, and hepatic regeneration. *Clin Liver Dis* 2008; **12**: 853-860, ix

27 **Pinzani M**, Vizzutti F. Fibrosis and cirrhosis reversibility: clinical features and implications. *Clin Liver Dis* 2008; **12**: 901-913, x

28 **Naveau S**, Poynard T, Benattar C, Bedossa P, Chaput JC. Alpha-2-macroglobulin and hepatic fibrosis. Diagnostic interest. *Dig Dis Sci* 1994; **39**: 2426-2432

29 **Gangadharan B**, Antrobus R, Dwek RA, Zitzmann N. Novel serum biomarker candidates for liver fibrosis in hepatitis C patients. *Clin Chem* 2007; **53**: 1792-1799

30 **Strader DB**, Wright T, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C. *Hepatology* 2004; **39**: 1147-1171

S- Editor Li LF **L- Editor** Stewart GJ **E- Editor** Zheng XM