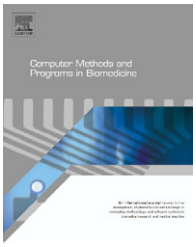
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 105 (2012) 194–209

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



# Single stage and multistage classification models for the prediction of liver fibrosis degree in patients with chronic hepatitis C infection

*Ahmed M. Hashem*[a,*], *M. Emad M. Rasmy*[b], *Khaled M. Wahba*[b], *Olfat G. Shaker*[c]

[a] *Department of Systems and Biomedical Engineering, Faculty of Engineering, Minia University, Minia, Egypt*
[b] *Department of Systems and Biomedical Engineering, Faculty of Engineering, Cairo University, Giza, Egypt*
[c] *Department of Medical Biochemistry and Molecular Biology, Faculty of Medicine, Cairo University, Cairo, Egypt*

ARTICLE INFO

*Article history:*
Received 29 July 2011
Received in revised form
14 October 2011
Accepted 14 October 2011

*Keywords:*
Hepatitis C virus
Liver fibrosis
Multistage classification
Pattern recognitions techniques

ABSTRACT

Predicting significant fibrosis or cirrhosis in patients with hepatitis C virus has persistently preoccupied the research agenda of many specialized research centers. Many studies have been conducted to evaluate the use of readily available laboratory tests to predict significant fibrosis or cirrhosis with the purpose to substantially reduce the number of biopsies performed. Although many of them reported significant predictive values of several serum markers for the diagnosis of cirrhosis, none of these diagnostic techniques was successful in accurately predicting early stages of liver fibrosis. Therefore, in this study a single stage classification model and a multistage stepwise classification model based on Neural Network, Decision Tree, Logistic Regression, and Nearest Neighborhood clustering, have been developed to predict individual's liver fibrosis degree. Results showed that the area under the receiver operator curve (AUROC) values of the multistage model ranged from 0.874 to 0.974 which is a higher range than what is reported in current researches with similar conditions.



## 1. Introduction

EGYPT has the highest prevalence of hepatitis C virus (HCV) in the world, reaching 14.7% of the population [1,2] equating to an estimated 11 million anti-HCV-positive persons. HCV is a major cause of chronic liver diseases and liver cirrhosis. The current gold standard for determining the extent of liver fibrosis is liver biopsy [3], but it is occasionally prone to limitations. These limitations include highly invasive nature and a risk of complications with morbidity between 0.3% and 0.6% and mortality of 0.05% [4]. Moreover, the interpretation of liver biopsy is prone to sampling error result due to the heterogeneous distribution of pathological changes in the liver [5]. Liver biopsy is 80% accurate in staging fibrosis, and may miss advanced fibrosis in 30% of patients [6]. Therefore, the tendency is to substitute the liver biopsy with non-invasive method for diagnosing and grading of liver fibrosis using serum markers assay and imaging techniques [7,8]. Many studies have been conducted to evaluate the use of readily available laboratory tests to predict significant fibrosis or cirrhosis in patients with HCV with the aim of substantially reducing the number of biopsies performed for the management of HCV infection [7,9–13]. Table 1 shows some studies of indirect serum markers of hepatic fibrosis and their possible

* *Corresponding author*. Tel.: +20 100 5109263; fax: +20 2 25292781.
E-mail addresses: ahmed.hashem@vodafone.com (A.M. Hashem), erasmy@gmail.com (M.E.M. Rasmy), kwahba@kwahba.com (K.M. Wahba), olfatshaker@yahoo.com (O.G. Shaker).
0169-2607/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2011.10.005

**Table 1 – Indirect serum markers of hepatic fibrosis and their possible interpretation.**

| Index | Parameters | CLD and number of patients | Calculation | Interpretation[a] | PPV/NPV (%) | AUROC |
|---|---|---|---|---|---|---|
| Forns | Age, plt, γGT, cholesterol | HCV; t = 351<br>v = 125 | 7.811 − 3.131 × ln(plt) + 0.781 × ln(γGT) + 3.467 × ln(age) − 0.014 (cholesterol). | >6.9 ≈ Scheuer 2–4<br><4.2 ≈ Scheuer 0–1 | PPV = 66<br>NPV = 96 | t = 0.86<br>v = 0.81 |
| APRI | AST, plt | HCV;<br>t = 192<br>v = 78 | ([AST/ULN]/plt [×10$^9$/l]) × 100. | >1.5 ≈ Ishak 3–6<br>≤0.5 ≈ Ishak 0–2 | PPV = 91<br>NPV = 90 | t = 0.80<br>v = 0.88 |
| FT, FS | Haptoglobin, α2-MC, apo-A1, γGT, bilirubin, γ-globulin | HCV, HBV;<br>t = 205<br>v = 134 | Logistic regression index (proprietary). | 0.75–1.00 ≈ F4<br>0.73–0.74 ≈ F3–F4<br>0.59–0.72 ≈ F3<br>0.49–0.58 ≈ F2<br>0.32–0.48 ≈ F1–F2<br>0.28–0.31 ≈ F1<br>0.22–0.27 ≈ F0–F1<br>0.00–0.21 ≈ F0 | PPV = 78<br>PPV = 76<br>PPV = 76<br>PPV = 67<br>PPV = 61<br>NPV = 91<br>NPV = 92<br>NPV = 94 | ≥F2–F4<br>t = 0.83<br>v = 0.87 |
| Fibroindex | Plt, AST, γGT | HCV; t = 240<br>v = 120 | 1.738 − 0.064 (plt [×10$^4$/mm$^3$]) + 0.005 (AST [IU/L]) + 0.463 × (γGT [g/dl]). | ≤1.25 ≈ F0–F1<br>≥2.25 ≈ F2–F3 | NPV = 61.7<br>PPV = 90 | t = 0.83<br>v = 0.82 |
| FPI | AST, cholesterol, past alcohol intake, HOMA, age | HCV; t = 176<br>v = 126 | $E^*/1 + e^*$, where $^* = -10.929 + (1.827 \times \ln[AST]) + (0.081 \times age) + (0.768 \times$ [past alcohol use graded as 0–2]) + (0.385 × HOMA). | <0.2 ≈ F0–F1<br>≥0.8 ≈ F2–F4 | NPV = 77.4<br>PPV = 87 | t = 0.84<br>v = 0.77 |
| FIB-4 | Plt, AST, ALT, age | HCV or HIV;<br>t = 555<br>v = 277 | (Age × AST)/(plt count × ALT$^{1/2}$). | <1.45 ≈ Ishak <4–6<br>>3.25 ≈ Ishak ≥4–6 | NPV = 90<br>PPV = 65 | 0.76 |
| Bonacini | ALT, AST, INR, plt | HCV; 79 | Sum (range 0–11) of (plt score) + (ALT/AST score) + (INR score). plt (×10$^9$/l): >340 = 0; 280–339 = 1; 220–279 = 2; 160–219 = 3; 100–159 = 4; 40–99 = 5; <40 = 6. ALT/AST ratio: >1.7 = 0; 1.2–1.7 = 1; 0.6–1.19 = 2; <0.6 = 3. INR: \1.4 = 2. | >8 ≈ Knodell 3–4 | PPV = 92.9 | NR |
| Pohl | AST, ALT, plt | HCV; 211 | Positive if: AST/ALT ≥1 and platelet count <150 × 10$^9$/l. | Positive ≈ F3–F4 | PPV = 93 | NR |
| AP | Plt, age | HCV; t = 500<br>v = 120 | Age score + plt score (0–10 possible score) age: <30 = 0; 30–39 = 1; 40–49 = 2; 50–59 = 3; 60–69 = 4; ≥70 = 5. Plt (×10$^9$/l): ≥225 = 0; 200–224 = 1; 175–199 = 2; 150–174 = 3; 125–149 = 4; <125 = 5. | ≥6 ≈ F2–F4 | PPV = 96 | t = 0.76<br>v = 0.69 |

ALT, alanine aminotransferase; AP, age-platelet; apo-A1, apolipoprotein A1; APRI, AST-platelet ratio index; AST, aspartate aminotransferase; AUROC, area under the receiver operator curve; CLD, chronic liver disease; FPI, fibrosis probability index; FS, Fibrosure® (Laboratory Corporation of America, Burlington, NC); FT, Fibrotest; GT, γ-glutamyltransferase; HOMA, homeostatic model assessment; INR, international normalized ratio; α2-MC, α2-macroglobulin; NPV, negative predictive value; NR, not reported; plt, platelet count; PPV, positive predictive value; t, training group; ULN, upper limit of normal; v, validation group.

[a] Fibrosis stages refer to the METAVIR system (F0–F4) unless otherwise indicated.