*International Journal of Engineering Research and Development*
*eISSN : 2278-067X, pISSN : 2278-800X, www.ijerd.com*
*Volume 2, Issue 2 (July 2012), PP. 08-15*

# Biochemical Markers of Fibrosis for Chronic Liver Disease: Data mining-based Approach

Torky I. Sultan[1], Ayman Khedr[2], Samir Sabry[3]
[1]*Information System Department, Faculty of Computers and Information Helwan University, Cairo , Egypt*
[2]*Information System Department, Faculty of Computers and Information Helwan University, Cairo, Egypt*
[3] *Laboratories Department, Electricity Hospital, Ministry of Electricity & Energy, Cairo, Egypt*

***Abstract*—*Staging of liver fibrosis has important implications for disease prognosis and treatment decisions. Knowing the degree of a disease allows a physician to provide specific treatment to the patient leading to better care. The use of serum markers becomes important to predict liver fibrosis. The objective of this work is to build a classification model in the form of a tree structure to predict liver fibrosis stage in patients with chronic hepatitis C genotype 4 in Egypt. In addition, to evaluate whether laboratory examinations can be used to determine the rate of liver fibrosis progression in patients chronically infected with hepatitis C virus (HCV). This paper focuses on the FibroTest which used to stage liver fibrosis to compare the results of the non-invasive markers using histology as reference method, and tried to determine if they could also be used in Egyptian patients.***

***Keywords*—*Data mining, Decision Trees, Serum Markers, FibroTest; Liver fibrosis, Chronic Liver Disease***

## I. INTRODUCTION

Chronic liver diseases (CLDs) represent a major cause of morbidity and mortality worldwide. Chronic infection with hepatitis C virus (HCV) is the most common causes of cirrhosis in the world today. Assessment of fibrosis is important in chronic hepatitis C for a number of reasons includingdecision-making regarding treatment and predicting prognosis.

Liver Biopsy (LB) is important diagnostic tool in CLD, cirrhosis and hepatocellular carcinoma but had its limitation and risks [1], [2]. Unfortunately, LB is often painful, requires bed rest for at least six hours, and is associated with a small but definite mortality. Among the complications of percutaneous LB are pain (10%-30%), bleeding (which may be severe and necessitate blood transfusion or emergency surgery). Accidental needle puncture of the lung, intestines, gallbladder, or kidney. Abdominal infection may accrue with pain. Furthermore, Subjective to many factors difference in the biopsy sampling size may miss the diseased part of the liver may require repetition of the samples (20%) is due to LB failure (small length).The risk of death from the biopsy is less than 1 in 1,000[3],[4].

The Arab Republic of Egypt has the highest prevalence of hepatitis C in the world. The national prevalence rate of hepatitis C virus (HCV) antibody positivity has been estimated to be between 10-13% according to a study published on August 2010 in the National Academy of Sciences. Chronic HCV is the main cause of liver cirrhosis and liver cancer in Egypt and, indeed, one of the top five leading causes of death. Genotype 4 represents over 90% of cases in Egypt. Non-invasive methods have been extensively developed in recent years as alternatives to liver biopsy for predicting liver fibrosis in patients with chronic hepatitis C, the most validated being FibroTest (FT) and ActiTest (AT) (Biopredictive, Paris, France)

FT measures the degree of fibrosis and combines five serum biochemical markers (Alpha2-macroglobulin, haptoglobin, gamma glutamyltranspeptidase (GGT), total bilirubin, and apolipoprotein A1) with patient age and sex.The outcomes describe the degree of fibrosis [FT unit range from 0-no fibrosis to 1-cirhosis]. AT measures the degree of necrosis and inflammation by combining the above measures with ALT [AT unit range from 0-no inflammation to 1-high degree of inflammation] [5].Data mining techniquesoften used to create and deploy successful business intelligence solutions. By applying data mining techniques, we can fully exploit data to discover previously unknown trends.

Decision trees are one of the most popular data mining algorithms and knowledgerepresentation means.Decision trees provide explicitrules to relate the range of values of the biomarkers withfibrosis scores, and they might help to gain a better graspof the importance and significance of the test.

## II. PATIENTS AND METHODS

### A. Study design and patients numbers

Fig.1 shows the flowchart of the study design and lists for patient. Patients were excluded if they had other causes of CLDs, including chronic hepatitis B, fatty liver disease, alcohol abuse or Autoimmune and genetic liver disease or if they had different HCV genotyping, or if they previously received interferon therapy. All of the patients were evaluated by abdominal ultrasound to exclude those with ascites.

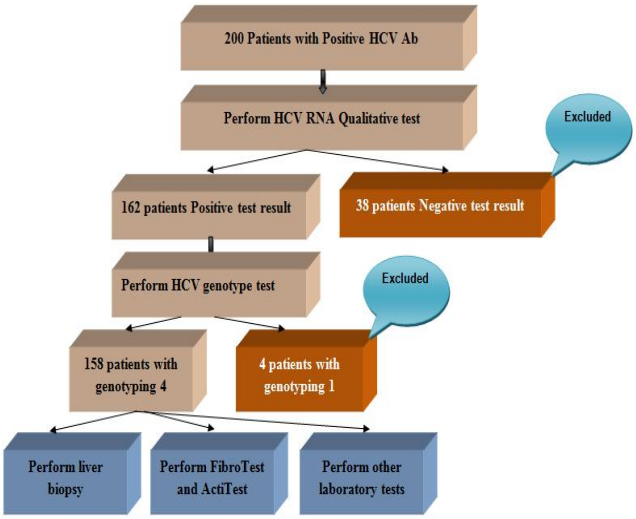



***Fig.1***Flowchart of the study design

**B. Patients**

A total number of 158 serum samples were collected from patients with chronic hepatitis C (CHC). HCV genotyping was done for all patients to assure that the patients are infected with HCV genotyping 4. Serum samples were obtained and liver needle biopsy was performed on the same day. Levels of fibrosis in FT and levels of activity in AT, both determined via serum biochemical markers, were compared with levels of fibrosis and activity in histopathological examination. The study group consisted of 158 patients (123 males and 35 females) with no prior antiviral treatment were included; All patients had positive HCV- RNA (genotype 4). The mean age of the patients was 49±9.74 years, ranging from 23 to 69 years.

**C. Biochemical Markers (FibroTest And ActiTest)**

Serum samples were taken on the day of biopsy from the patients in fasting state. Six serum biochemical markers were analyzed: α2-macroglobulin, haptoglobin, GGT, total bilirubin, apolipoprotein A1, and ALT on an automated analyzer (OLYMPUS AU640). All biochemical parameter and FT and AT determinations were done without knowledge of liver biopsy results. Fibrosis using FT was staged on a scale of 0–4 with respect to Metavir fibrosis staging. For FT score from 0 to 0.21 fibrosis was staged F0, from 0.22 to 0.27 F0–F1, from 0.28 to 0.31 F1, from 0.32 to 0.48 F1–F2, from 0.49 to 0.58 F2, from 0.59 to 0.72 F3, from 0.73 to 0.74 F3–F4, and from0.75 to 1 F4.

Necroinflammatory activity using AT was graded on a scale of 0–3 with respect to Metavir activity grading. For AT score from 0 to 0.17 activity was graded A0, from 0.18 to 0.29 A0–A1, from 0.30 to 0.36 A1, from 0.37 to 0.52 A1A2, from 0.53 to 0.60 A2, from 0.61 to 0.62 A2–A3, and from 0.63 to 1 A3 [6]. The FT score was computed on the Biopredictive website (www.biopredictive.com), by entering the patient's age, sex, and results for the five biochemical analyses listed.

**D. Laboratory Investigations**

All of the following 16 laboratory tests of routine work and fibrosis markers were identified in literature. Routine work tests are: platelet count (PLT), white blood cell (WBCs), red blood cells (RBCs), hemoglobin (HB) concentration, hematocrit (HCT), prothrombin time (PT), international normalized ratio (INR), serum albumin (ALB), Aspartate aminotransferase (AST), Alkaline phosphatase (ALP), creatinine (CRT), total cholesterol (CHO), high density lipoprotein (HDL), low density lipoprotein (LDL), triglycerides (TG) , and fasting blood sugar (FBS).

Biochemical assays are usually performed with fresh serum. Serum can be decanted and stored at –80°C, although freezing and thawing can only be done once. The following items, Hepatitis B Surface Antigen (HBsAg), Hepatitis C virus antibody (HCV-Ab), Human Immunodeficiency Virus (HIV), Hepatitis C virus RNA qualitative (HCV-RNA), HCV Strains (HCV genotype), are performed at the first time only to determine the eligible patients for the study.

**E. Histologic Staging**

Liver biopsies were obtained and fixed with formalin, embedded in paraffin, and stained with hematoxylin– eosin, and histology for fibrotic staging (F) and inflammatory process (A) was determined by the department of pathology according to the METAVIR score. Fibrosis was staged on a F0–F4 scale: F0, no fibrosis; F1, portal fibrosis without septa; F2 portal fibrosis with few septa; F3 septal fibrosis without cirrhosis; and F4 cirrhosis. And grade of activity on A0-A3; A0, no activity; A1 mild; A2 moderate; A3 severe activity[7].

Patients with scores of F0 or F1 were considered to have insignificant fibrosis, and those with scores of F2, F3, or F4 were considered to have clinically significant fibrosis that qualified for combination antiviral therapy. Liver biopsies were performed at the time of serum sampling and were reviewed and classified according to the Metavir scoring systems (Table I)[7].

***TableII:*** **Metavir Fibrosis Staging and Activity Grading**

| FibroTest | Metavir fibrosis Stage Estimate | ActiTest | Metavir Activity Grade Estimate |
|---|---|---|---|
| 0.75-1.00 | F4 | 0.63-1.00 | A3 |
| 0.73-0.74 | F3-F4 | 0.61-0.62 | A2-A3 |
| 0.59-0.72 | F3 | 0.53-0.60 | A2 |
| 0.49-0.58 | F2 | 0.37-0.52 | A1-A2 |
| 0.32-0.48 | F1-F2 | 0.30-0.36 | A1 |
| 0.28-0.31 | F1 | 0.18-0.29 | A0-A1 |
| 0.22-0.27 | F0-F1 | 0.0-0.17 | A0 |
| 0.00-0.21 | F0 | | |

## III. RESEARCH METHODOLOGY

### A. Approach Adopted

We proposed to evaluate whether the stage of liver fibrosis can be estimated based on laboratory tests. Our aim was to discover whether biopsies can be replaced by lab tests, since the former involve invasive procedures. The approach adopted here consisted of analyzing blood tests and biochemical markers together with biopsy results, seeking patterns that might indicate a correlation between the patients' exam results and the degree of their fibrosis. Fig.2 shows the Process of Building a Predictive Model.

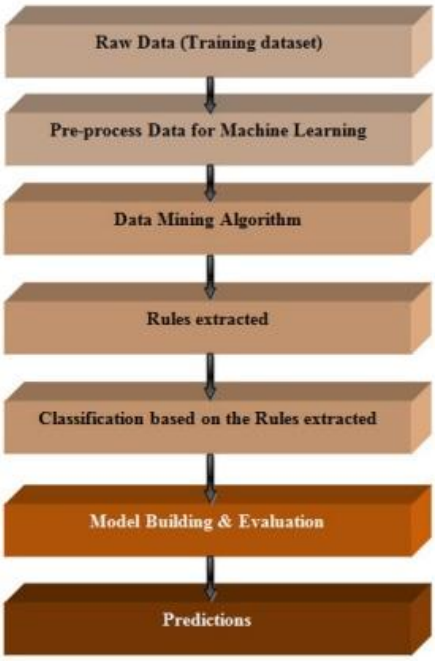



***Fig. 2*** The Process of Building a Predictive Model

### B. Data Mining

The main purpose of doing data mining and knowledge discovery on the medical database is to predict disease and disease classification. Classification and prediction are two forms of data analysis which can be used to describe the model of the important data type or predict the future trends of the data [8]. The aim of this study is to show that data mining can be applied to the laboratory databases, which will predict liver fibrosis stage by constructing decision trees in patients with chronic hepatitis C genotype 4 (HCV-4) in Egypt . For a good prediction or classification the learning algorithms must be provided with a good training set from which rules or patterns are extracted to help classify the testing dataset.

### C. Data Characteristics

The dataset contains data on laboratory examinations, which were collected on Electricity hospital in Egypt. The subjects are 158 patients of hepatitis C who took examinations between 2010 and 2012. The data were divided into four categories. The first data include patient's information (Id, age, gender, phone, City, and weight). Second data include pathological classification of the disease (result of biopsy, and result of activity). The third data include (six serum biochemical markers were analyzed: α2-macroglobulin, haptoglobin, GGT, total bilirubin, apolipoprotein A1, and ALT). The last data include 16 laboratory tests of routine work were performed PLT, WBC, RBC, HB, HCT, PT, INR, ALB, ALP, AST, CRT, CHO, HDL, LDL, TG, and FBS.

**D. Data Preprocessing for Machine Learning**

Data come in various formats, depending on their provider and without putting them into the correct shape, any machine learning tool would be useless [9]. Most data mining tools can use data in the CSV format for running the machine intelligent algorithms. We used Weka 3.6 software [10] which is a collection of machine learning algorithms for data mining tasks. We also used Microsoft Excel to collect laboratory data but the input data file needs to be organized in the form of ARFF in order to be processed in the Weka environment. The data that are used for WEKA should be made into the following format shown in Fig.3 and the file should have the extension dot ARFF (.arff). The last attribute where the classification of the patient is done into a categorical format, that is, the classification attribute 'BiopsyFibrosis' takes string values 'F0,F1,F2,F3,F4'.

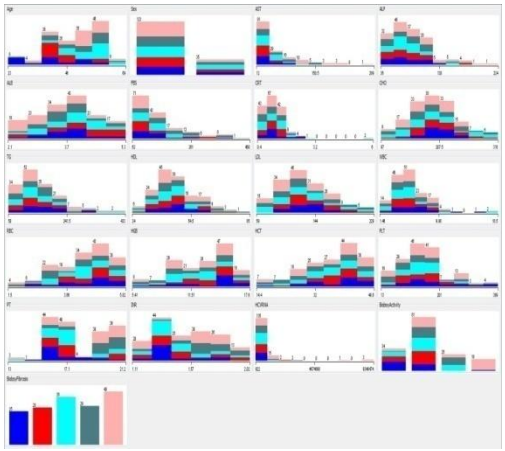
```
@relation 'all data'

@attribute Age numeric
@attribute Sex {MALE,FEMALE}
@attribute AST numeric
@attribute ALP numeric
@attribute ALB numeric
@attribute FBS numeric
@attribute CRT numeric
@attribute CHO numeric
@attribute TG numeric
@attribute HDL numeric
@attribute LDL numeric
@attribute WBC numeric
@attribute RBC numeric
@attribute HGB numeric
@attribute HCT numeric
@attribute PLT numeric
@attribute PT numeric
@attribute INR numeric
@attribute 'BiobsyActivity ' {A0,A1,A2,A3}
@attribute BiobsyFibrosis {F0,F1,F2,F3,F4}

@data
26,MALE,35.77.2.8.136.2.217.294.37.153.3.95.5.45.16.4.47.6.153.16.3.1.32.A0,F0
```

***Fig. 3***: The WEKA Training data file for Biopsy Fibrosis

**E. Building the Model**

Machine Learning is the area of Artificial Intelligence (AI) that examines how to write programs that can learn. Often used in classification, prediction and deals with small static datasets. All the attributes in this database are displayed in row format in the left half of the screen and on the right side of the screen the bar graphs represent the distributions of the different attributes that are considered for data mining.

Weka software was executed using the model training file. Fig.4shows histograms of each input parameter (laboratory tests)with the area color coded by fibrosis stage. AndFig.5shows histograms of the following parameter(predictive attributes)were used: sex, age, T.bilirubin, GGT, A2M, ALT, Apo A1, and haptoglobin with the area color coded by fibrosis stage, to predict FT and AT scores.



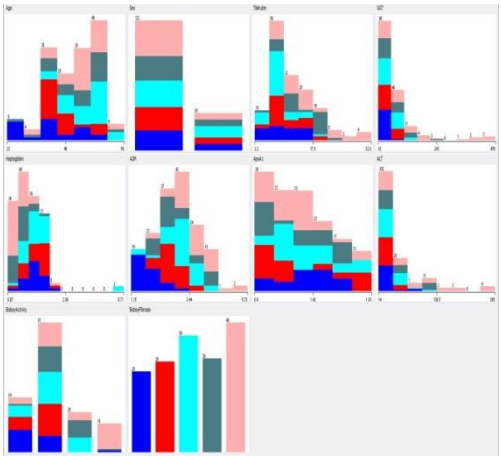


***Fig. 4***Visualization Histograms (Fibrosis Stage)

***Fig.5***Visualization Histograms (predictive attributes)

**F. Decision Tree**

Decision tree is one of the easier data structure to understand data mining. Rules from the training dataset are first extracted to form the decision tree which is then used for classification of the testing dataset. Decision trees were constructed with data from 158 patients with chronic hepatitis C using the C4.5 classification algorithm [12].

In order to enhance the confidence of the classifier, we used all the data for training and also for testing as follows, use 88% (139 patients) as train data and the remaining designated 12% (19 patients) as test-data and compute the classifier's performance. For the experiments, the following predictive attributes were used: sex, age, T.bilirubin, ALT, GGT, A2M, Apo A1, and haptoglobin. The targets were the FibroTest and ActiTest scores. Fig.6shows a graphical presentation of the decision tree model of all laboratory data as the data set.

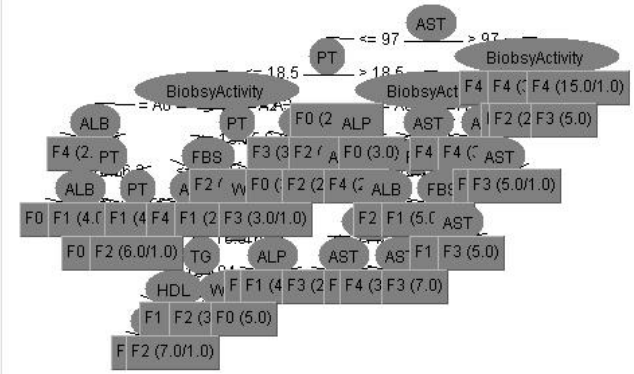



***Fig.6***Complete Decision Tree (Laboratory data)

## IV. RESULTS

**A. Laboratory Examination and Biochemical Markers**

One hundred fifty-eight patients with chronic hepatitis C genotype 4 were enrolled in this study. There were 123 men and 35. The mean values for laboratory examination data and biochemical markers of fibrosis are summarized in TablesII&III.

***Table II***:**laboratory Examination Data of the Study Population**

| Patients with CHC (n = 158) | Mean ± SD |
|---|---|
| AST | 55.3 ± 48 |
| ALB | 3.7 ± 0.78 |
| ALP | 84.6 ± 32.9 |
| GLU | 154.8 ± 83.4 |
| CRE | 1.40 ± 1.12 |
| CHO | 196.3 ± 45.5 |
| TG | 154.4 ± 71.6 |
| HDL | 46.6 ± 10.8 |
| LDL | 127.6 ± 35.4 |
| WBC | 5.8 ± 2.7 |
| RBC | 4.3 ± 0.96 |
| HGB | 12.6 ± 2.8 |
| HCT | 36.8 ± 8.1 |
| PLT | 151.5 ± 64.5 |
| PT | 17.9 ± 1.8 |
| INR | 1.5 ± 0.22 |

***Table III*: Biochemical Markers of Fibrosis**

| Patients with chronic hepatitis C (n = 158) | Mean ± SD |
|---|---|
| T-BIL | 12 ± 5.81 |
| ALT | 58 ± 48.6 |
| GGT | 74 ± 84.9 |
| Alpha2-macroglobulin | 2.9 ± 0.91 |
| Apolipoprotein A1 | 1.3 ± 0.26 |
| Haptoglobin | 1.1 ± 0.76 |

**B. FibroTest and ActiTest**

The reported FT scores indicate that 14% were F0, 4%F0-F1, 3% F1, 21% F1-F2, 15% F2, 17% F3, 1% F3-F4, and 25% F4. The reported AT score indicate that 27% were A0, 26% A0-A1, 12% A1, 11% A1-A2, 7% A2, 3% A2-A3, and 14% A3. (Table IV)

***Table IV*Results of FibroTest and ActiTest**

| FibroTest (n158) | | ActiTest (n158) | |
|---|---|---|---|
| F0 | 23 | A0 | 42 |
| F0-F1 | 6 | A0-A1 | 41 |
| F1 | 5 | A1 | 19 |
| F1-F2 | 33 | A1-A2 | 17 |
| F2 | 23 | A2 | 11 |
| F3 | 26 | A2-A3 | 5 |
| F3-F4 | 2 | A3 | 23 |
| F4 | 40 | | |

**C. Liver Biopsy**

The following distribution of METAVIR fibrosis stages was observed on liver biopsy: no fibrosis in 25 of 158 patients, portal fibrosis in 28 of 158, few septa in 36 of 158, numerous septa in 29 of 158 and cirrhosis in 40 of 158. Fig.7 shows Fibrosis stages and Fibrosis grading for 158 patients.

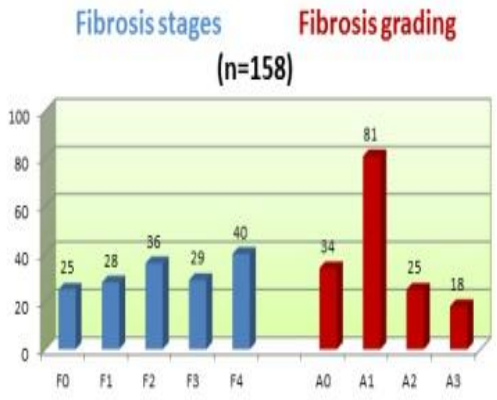



***Fig.7***Fibrosis stages and grading

## V. PREDICTION

In the present study, we tested the classification performance with compound rules and computed the confusion matrix in Fig.8.The overall classification error was 7.5% (accuracy 92.5%). FT cases with true scores of F0 and F4 were classified with very high accuracy (22/23 for F0 and 36/40 for F4), which indicated that in the extreme stages of fibrosis, decision trees produced the correct classification in approximately 94% of the cases. And AT cases with true scores of A0 and A3 were classified with very high accuracy (40/42 for A0 and 23/23 for A3), which indicated that in the extreme graded of fibrosis, decision trees produced the correct classification in approximately 95.5% of the cases.

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h   <-- classified as
 22  0  0  0  0  1  0  0 |  a = F0
  0 32  0  0  1  0  0  0 |  b = F1-F2
  0  1 22  0  0  0  0  0 |  c = F2
  0  0  1 24  1  0  0  0 |  d = F3
  1  1  1  1 36  0  0  0 |  e = F4
  1  0  1  0  0  4  0  0 |  f = F0-F1
  0  0  0  0  0  0  2  0 |  g = F3-F4
  0  1  0  0  0  0  0  4 |  h = F1
```

***Fig.8*** Confusion matrix FT

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 40  0  2  0  0  0  0 |  a = A0
  0 17  0  0  0  0  0 |  b = A1-A2
  0  0 40  0  1  0  0 |  c = A0-A1
  0  0  0 23  0  0  0 |  d = A3
  0  0  0  0 19  0  0 |  e = A1
  0  0  0  0  0 11  0 |  f = A2
  0  1  0  3  0  0  1 |  g = A2-A3
```

***Fig.9***Confusion matrix AT

## VI. EVALUATION AND DISCUSSION

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 [13] to the prediction of interferon treatment effects based on microarray gene expression profiles [14], and to the prediction of diagnosis and outcome of dengue fever based on clinical, hematological and Virological data [15].

Most of biochemical markers for predicting the stages of liver fibrosis, for examples, as indirect markers, AST, ALT, and g-GT reflect liver injury; prothrombin time, cholesterol, haptoglobin, and a2-macroglobulin reflect altered liver function caused by architectural changes; platelet counts are closely related with portal hypertension.

In this study, we included most previously known indirect potential markers of liver fibrosis and tried to determine if they could also be used in Egyptian patients, in whom chronic hepatitis C is the main cause of chronic liver disease.It is clinically important to assess the progression of liver fibrosis in patients with chronic liver disease. Furthermore, there is an increasing need for accurate noninvasive methods, such as biochemical markers, which enable repetitive measurement of the degree of liver fibrosis.

For the analysis 139 patient's information was used to develop the decision tree model and predict the other 19 patients' fibrosis stage. We used20 laboratory attributes Fig. 3to predict the result. The result of predicting the values with the constructed decision tree model is shown in Table 5. The decision tree model results showed an accuracy of 68.4% (13/19) of correct fibroid prediction.

***Table V***: **Decision Tree Prediction**

| **Known Fibrosis Stage (K)** | **Predicted Fibrosis Stage (P)** | **Difference=(Known-Predicted) D=K-P** |
|---|---|---|
| 3 | 3 | 0 |
| 4 | 4 | 0 |
| 0 | 0 | 0 |
| 4 | 4 | 0 |
| 2 | 2 | 0 |
| 4 | 3 | 1 |
| 3 | 3 | 0 |
| 3 | 3 | 0 |
| 3 | 1 | 2 |
| 4 | 4 | 0 |
| 1 | 1 | 0 |
| 4 | 4 | 0 |
| 3 | 2 | 1 |
| 2 | 2 | 0 |

| | | |
|---|---|---|
| 3 | 4 | -1 |
| 3 | 3 | 0 |
| 3 | 3 | 0 |
| 2 | 3 | -1 |
| 0 | 1 | -1 |

## VII. CONCLUSIONS

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.Serum markers are of great value not only in patients at risk for LB, but also as a part of the assessment of patients with chronic liver disease avoiding the invasive methods.

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

The results indicated that Decision trees model were useful and effective to predict liver fibrosis stage at least similar to biopsy and provide a qualitative and quantitative overview for physician to find the relations between FT and LB. Our great dream of having a test other than LB avoiding all the complications and is linear, reproducible and accurate became true by noninvasive biomarkers FibroTest-ActiTest.

## VIII. ACKNOWLEDGMENT

We would like to thank all the medical staff of Laboratory Unit and Anti-fibrosis Hepatology Unit at Electricity hospital for their guidance and collaboration. We are thankful to all staff of Al Borg Laboratories and Al Mokhtabar Laboratories for their help and cooperation. Finally, my deepest gratitude and appreciation goes to Eng. Mahmoud Shaaban for helping us throughout the study, we are truly appreciated all the time and suggestions given by him.

## REFERENCES

[1]. Fontana RJ, Goodman ZD, Dienstag JL, Bonkovsky HL, Naishadham D, Sterling RK, Su GL, et al. "Relationship of serum fibrosis markers with liver fibrosis stage and collagen content in patients with advanced chronic hepatitis C". Hepatology 2008; 47(3):789-798.

[2]. White IR, Patel K, Symonds WT, Dev A, Griffin P, Tsokanas N, Skehel M, et al. "Serum proteomic analysis focused on fibrosis in patients with hepatitis C virus infection" J Transl Med 2007; 5:33.

[3]. McGill DB, Rakela J, Zinsmeister AR, Ott BJ. "A 21-year experience with major hemorrhage after percutaneous liver biopsy".Gastroenterology 1990; 99: 1396-1400

[4]. Van Thiel DH, Gavaler JS, Wright H, Tzakis A. "Liver biopsy.Its safety and complications as seen at a liver transplant center". Transplantation 1993; 55: 1087-1090.

[5]. Colletta C, Smirne C, Fabris C, et al. Value of two noninvasive methods to detect progression of fibrosis among HCV carriers with normal aminotransferases. Hepatology 2005; 42(4): 838–45.

[6]. Poynard T. Diagnosis method of inflammatory, fibrotic or cancerous disease using the biochemical markers. United States: Patent No US 6,631,330 B1, 2003.

[7]. P. Bedossa, T. Poynard, An algorithm for the grading of activity in chronic hepatitis C, The METAVIR Cooperative Study Group, Hepatology 24 (August (2)) (1996) 289–293.

[8]. V. Špečkauskienė and A. Lukoševičius, "Methodology of Adaptation of Data Mining Methods for Medical Decision Support: Case Study", 2009.

[9]. Jiawei Han ans Micheline Kamber, "Data Mining Concepts and Techniques"

[10]. http://www.cs.waikato.ac.nz/ml/weka/. Weka Data mining Software.

[11]. Quinlan, J.R. 1993), C4.5: Programs for Machine Learning. California: Morgan Kaufmann Publishers, Inc.

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

[13]. Ge G, Wong GW. Classification of premalignant pancreaticcancer mass-spectrometry data using decision tree ensembles. BMC Bioinformatics 2008; 9: 275.

[14]. 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.

[15]. 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 earlyphase of illness. PLoS Negl Trop Dis 2008; 2: e196.