

# Business Intelligence framework to support Chronic Liver Disease Treatment

Ayman Khedr

Information System Department, Faculty of Computers and Information, Helwan University, Cairo, Egypt  
ayman\_khedr@Helwan.edu.org  
aymankhedr@gmail.com

## ABSTRACT

Business Intelligence (BI) framework designs the architecture of business intelligence information system which uses expert systems and Artificial Intelligence technology to support clinical decision and draw the strategy against chronic liver disease in Egypt. It makes integrated diagnostic and medical advice bases on the collected patient's information, providing reference for the clinical medical officers. This paper aims to support decision function and in particular utilization of historical data laboratory and outcome data processed through artificial intelligence tools. The combination of historical data and predictive tools provides valuable information in the hands of physicians as they develop a course of treatment for a patient.

**Keywords:** Business Intelligence, Data Mining, FibroTest, Liver fibrosis, Chronic Liver Diseases, Hepatitis C

## 1. INTRODUCTION

The framework of business intelligence system aims to support clinical decision for national strategy against chronic liver disease in Egypt. Liver fibrosis is a chronic disease that results from viral hepatitis, fatty liver disease, alcohol abuse or autoimmune and genetic liver disease. Chronic infection with hepatitis C virus (HCV) is one of the most common causes of cirrhosis in the world today. Assessment of fibrosis is important in chronic hepatitis C for a number of reasons including decision-making regarding treatment and predicting prognosis.

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). [1]

FT measures the degree of fibrosis and combines five serum biochemical markers (Alpha2-macroglobulin, haptoglobin, gamma glutamyltranspeptidase (GGT), total bilirubin, and apolipoprotein A1) with the patient's age and sex. The outcomes describe the degree of fibrosis [FT unit range from 0-no fibrosis to 1-cirrhosis]. 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] [2].

This paper is divided into six sections. Section 2 presents background and literature reviews for previous work in biomedical for chronic liver disease and business intelligence system. Problem definitions and Research methodology are presented in section 3. Section 4 presents the proposed business intelligence framework to treat the chronic liver disease. Section 5 presents the case study for deployment of business intelligence framework to treat the chronic liver. Finally, conclusions and future work are presented in section 6.

## 2. BACKGROUD AND PREVIOUS WORK

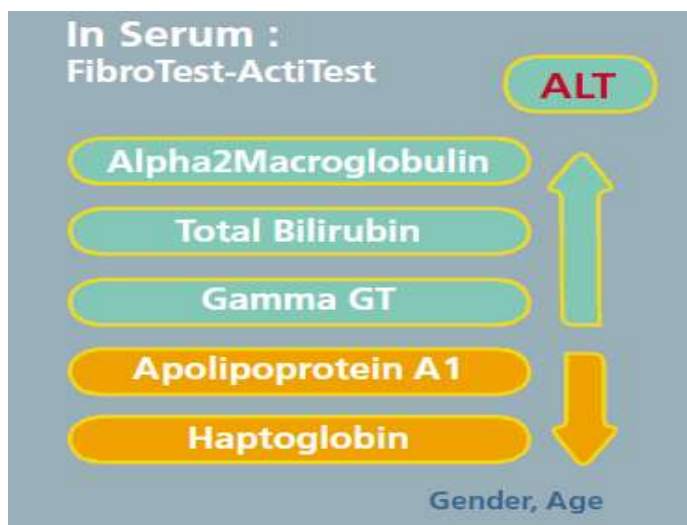
The "gold standard" for assessing fibrosis, liver biopsy (LB), is recommended prior to the initiation of antiviral therapy; in addition, it is vital for monitoring fibrosis progression. Unfortunately, this procedure is invasive, prone to complications, including hemorrhage and death, and has a high risk of sampling error. Biochemical markers for liver fibrosis (FT) and necroinflammatory features ActiTest (AT) are an alternative to LB, in patients with chronic hepatitis C. Since September 2002, FT and AT has been used in several countries as an alternative to liver biopsy in order to estimate liver fibrosis and necroinflammatory activity in chronic viral hepatitis C. Several prospective studies have validated these panels of tests in chronic viral hepatitis C and demonstrated its predictive value and the better benefit: risk ratio than biopsy [3].

### 2.1 FibroTest and ActiTest

The goal of this section is to introduce and to briefly review the area of biomedical technology for chronic liver disease and business intelligence features. First, Serum samples were taken on the day of biopsy from the patients in fasting state. Six serum biochemical markers were analyzed:  $\alpha$ 2-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 from 0.75 to 1 F4.

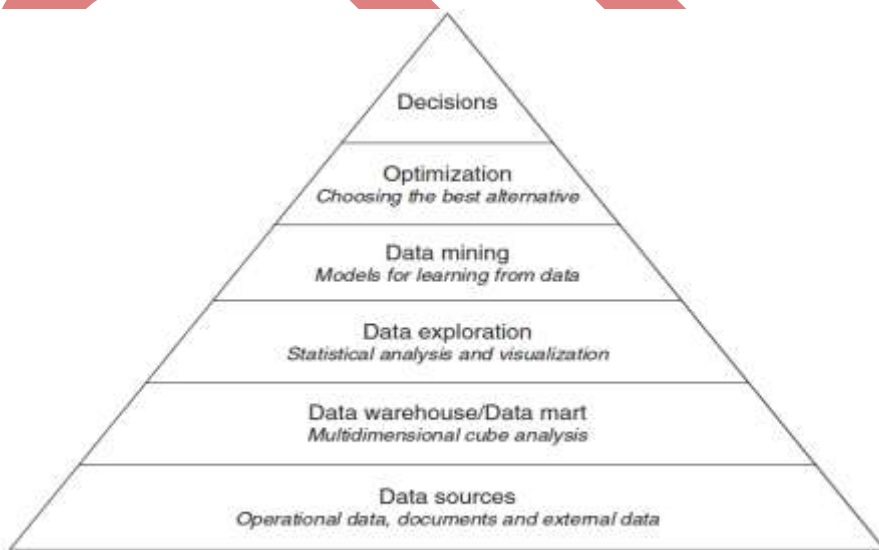
Necroinflammatory activity using AT was graded on a scale of 0–3 with respect to Metavir activity grading. For AT score from 0 to 0.17 activity was graded A0, from 0.18 to 0.29 A0–A1, from 0.30 to 0.36 A1, from 0.37 to 0.52 A1–A2, from 0.53

to 0.60 A2, from 0.61 to 0.62 A2–A3, and from 0.63 to 1 A3 [4]. 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 (see Figure 1).



**Fig 1: FibroTest Markers**  
**Source: Poynard, 2003.**

Secondly, business intelligence has the following definitions: Turban et al. (2007) define BI as “applications and technologies to help users make better business decisions” [5]. Vercellis (2009) defines BI as “a set of mathematical models and analysis methodologies that systematically exploit the available data to retrieve information and knowledge useful in supporting complex decision-making (see Figure 2). [6].



**Fig 2: The main components of a BI application**  
**Source: Vercellis, 2009.**

Bräutigam et al. (2006) defined BI as “getting the right information to the right people at the right time” [7]. Multi-criteria business intelligence approach (MCBI) is defined as “a set of models, descriptions and analysis methodologies that systematically exploit the available data to retrieve information, knowledge and advice useful in supporting complex decision making processes according to many criteria and alternatives”[8]. (See Figure 3) illustrates the architecture of multi-criteria business intelligence approach.

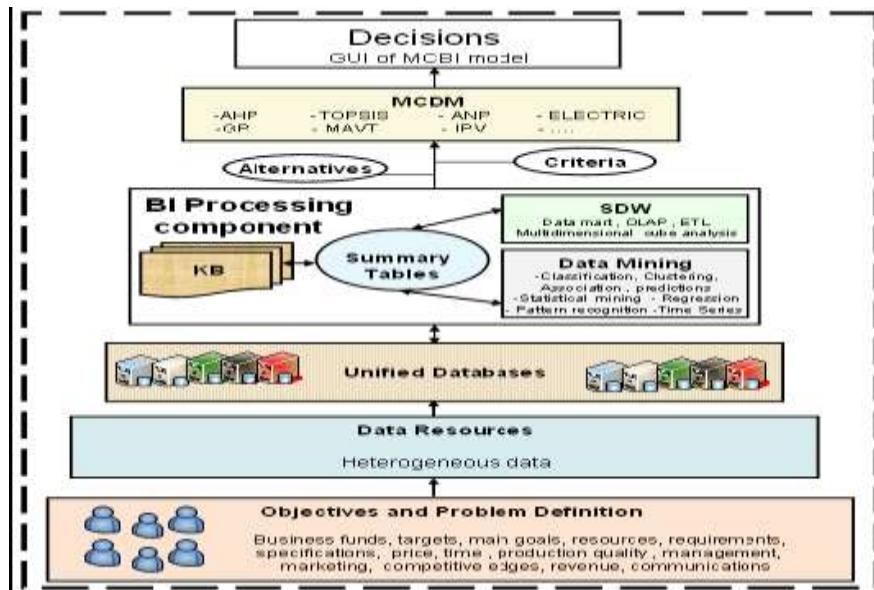


Fig 3: The architecture of multi-criteria business intelligence approach.

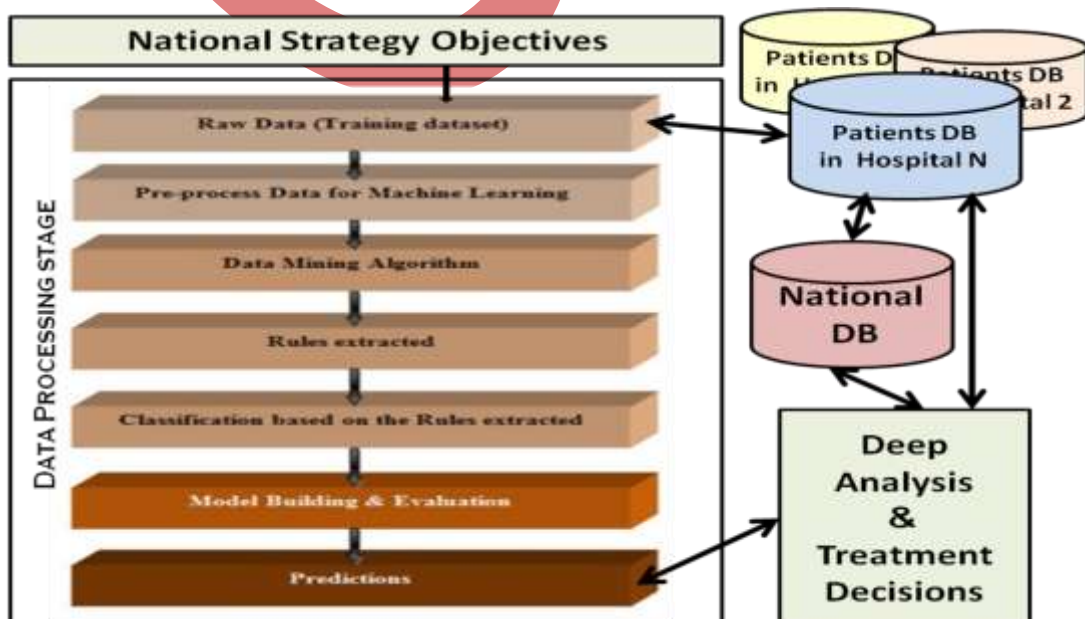
Source: Ali, 2012.

### 3. PROBLEM DEFINITION AND RESEARCH METHODOLOGY

This study uses both qualitative and quantitative approaches to support clinical decision and draw the strategy against chronic liver disease in Egypt. The qualitative approach will analyze data of livers samples to make the quantitative levels according to the analytical results for liver diseases. These results have many patients' levels and each level has a group of patients that has same analytical result level. The proposed framework for business intelligence information system will use these analytical results to support the strategic decisions to treat the chronic liver disease in Egypt.

### 4. Proposed Business Intelligence framework to treat the chronic liver disease

The Proposed framework explored the national strategy objectives, patients' databases in different hospitals, decision tree as a classification data-mining tool to extract useful information contained in large data sets and patients' historical reports to make deep analysis for the extracted facts to help decision-making processes to treat liver diseases in Egypt through the following architecture (see figure 4).



**Fig 3 Business Intelligence framework architecture to support the treatment of chronic liver disease**

**Source: Sabry, 2012.**

A decision tree is made up of a set of nodes that classify the past realizations of an objective variable. Each classification is achieved by separation rules according to the numerical or categorical values of the explanatory variables. The classification rules of each node are derived from a mathematical process that minimizes the impurity of the resulting nodes, using the available learning set [9]. Our aim was to discover whether biopsies can be replaced by blood tests, since the former involve invasive procedures. The approach adopted here consisted of analyzing FibroTest with biopsy results, seeking patterns that might indicate a correlation between the patients' results and the degree of their fibrosis. Fig. 2 shows the Process of Building a Predictive Model. Two groups of patients are required: training set in which all candidate serum markers are measured, and a validation set in which the performance of the final model is assessed.

Sometimes the two sets are created by random selection from one pool of patients or alternatively the validation set can be entirely separate. An essential requirement is to establish the desired fibrosis stage endpoints. Normally, there is no attempt made to predict individual fibrosis stages, instead a binary 'presence' or 'absence' is used. For example the simplest variant would be a single endpoint of significant fibrosis defined using the METAVIR system as a grade of F2, F3 or F4.

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 [10]. 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 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. Classifying patients was based on these rules and the five stages of fibrosis (F0-F4) [11].

## 5. CASE STUDY

The case study was performed through the following sub steps: first, the experiment of the patient's data characteristics, second, FibroTest & ActiTest, and finally, the national strategy against chronic liver disease.

### 5.1 The Experiment Patients Data Characteristics

The number of serum samples collected from patients with chronic hepatitis C (CHC) is X. 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 X patients (M males and F females, where M is the number of males patients and F is the number females patients) with no prior antiviral treatment were included; All patients had positive HCV- RNA (genotype 4).

The study will determine the age interval for patients (i.e. the patients ages interval start from the age of the youngest patient to the oldest patient). It will then determine the mean age of the patients. Patients were excluded from the study 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.

The dataset contains data on laboratory examinations, which were collected on Electricity Hospital in Egypt. The subjects are x patients of hepatitis C who took examinations between 2010 and 2012. The data was divided into three categories. The first data entry includes patient's information. Second data entry includes pathological classification of the disease. The last data entry includes (six serum biochemical markers were analyzed:  $\alpha$ 2-macroglobulin, haptoglobin, GGT, total bilirubin, apolipoprotein A1, and ALT). A key attribute called MID (masked ID) is included in each of the tables. A patient ID can be used to gather information of a patient in the different tables.

**Table 1: Basic information of patients**

Attribute	Meaning	Type
MID	Identification of the patient (masked)	unsigned integer
PT_SEX	Sex of the patient	Categorical
PT_AGE	Age of the patient	Integer

**Table 2: Results of Liver biopsy**

Attribute	Meaning	Type
MID	(masked)	unsigned integer
BIO_No	Number of the specimen	Real number
BIO_Fibrosis	Fibrosis result	Categorical
BIO_Activity	Activity result	Categorical

**Table 3: Biochemical markers (FibroTest and ActiTest) data**

Attribute	Meaning	Type
MID	Identification of the patient (masked)	unsigned integer
T-BIL	Total bilirubin	Real number
ALT	Alanine aminotransferase	Integer
GGT	Gamma glutamyltranspeptidase	Integer
A2M	Alpha2-macroglobulin	Real number
APOA1	Apolipoprotein A1	Real number
HAPT	Haptoglobin	Real number

## 5.2 FibroTest And ActiTest

The study will compute the FT score based on the patient's age, sex, and results for the five biochemical analyses listed by using the Biopredictive website ([www.biopredictive.com](http://www.biopredictive.com)),

## 5.3 The national strategy for treating chronic liver disease

The Strategy uses histologic staging of liver biopsies 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. 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 4).

**Table 4: Metavir Fibrosis Staging and Activity Grading**

Source: P. Bedossa, 1996.

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		

X patients with chronic hepatitis C genotype 4 were enrolled in this study. There were M men and F females. The reported FT scores indication was equal to 100% for (F0, 0-F1, F1, F1-F2, F2, F3, F3-F4, and F4). The reported AT score indications are 100% for A0, A0-A1, A1, A1-A2, A2, A2-A3, A3 ( i.e. suppose that 22% were A0, 29% A0-A1, 15% A1, 14% A1-A2, 6% A2, 7% A2-A3, and 7% A3. The result of FibroTest and ActiTest will be providing the following characteristics:

- Highly sensitive and specific to identify different stages of fibrosis.
- Readily available, safe, inexpensive and reproducible.
- Applicable to the monitoring of disease progression or regression as a part of natural history of liver disease or treatment regimens.
- Not susceptible to false positive results, for example, in individuals with inflammation related to other diseases.
- Although no single ideal marker exists, several markers have been identified as possible useful indicators of fibrosis when used in conjunction with each other.

## 6. CONCLUSIONS AND FUTURE WORK

The frame work for Business intelligence framework to treat liver diseases presents an automated classification and rank range for random sample of liver disease patients according to different criteria. The framework helps the decision makers to perform an appropriate solution for the treatment of the chronic liver disease in Egypt according to the national strategy to treat liver diseases in Egypt based on the extracted information and facts. Also, 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 construction of decision trees using the FibroTest attributes provided 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. In the future, we will work to deploy and implement the Business intelligence framework to treat liver diseases on the random samples of chronic liver disease.

## 7. ACKNOWLEDGMENT

I would like to thank all the patients who were generous enough to participate in this study. My deepest gratitude and appreciation goes to Dr. Samir Sabry and Mr. Mohamed Reda for helping me throughout the study, I truly appreciate all the time and suggestions given by them.

## 8. REFERENCES

- [1] Wwww.biopredictive.com
- [2] Colletta, C., Smirne, C., Fabris, C., and et al., "Value of two noninvasive methods to detect progression of fibrosis among HCV carriers with normal aminotransferases", *Hepatology* 2005; 42(4): 838–45.
- [3] Grigorescu M, Rusu M, Neculoiu D, et al. The FibroTest value in discriminating between insignificant and significant fibrosis in chronic hepatitis C patients. The Romanian experience. *J Gastrointestin Liver Dis* 2007; 16: 31-37.
- [4] Poynard, T., "Diagnosis method of inflammatory, fibrotic or cancerous disease using the biochemical markers. United States: Patent No US 6, 631, 330 B1, 2003.
- [5] 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., and 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.
- [6] Turban, E., Sharda, R., Aronson, J. and King, D., "Business Intelligence: A Managerial Approach", 1st ed., Prentice Hall, 2007.
- [7] Vercellis, C., "Business intelligence: data mining and optimization for decision making", John Wiley & Sons, Ltd., 2009.
- [8] Brautigam, D., Miller, G. and Gerlach, S., "Business Intelligence Competency Centers", John Wiley and Sons, Inc., Hoboken, New Jersey, Canada, 2006.
- [9] Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, August 2000.
- [10] Špečkauskienė, V. and Lukoševičius, A., "Methodology of Adaptation of Data Mining Methods for Medical Decision Support: Case Study", *Electronics and Electrical Engineering*, 2009.
- [11] Bedossa, P., Poynard, T., "An algorithm for the grading of activity in chronic hepatitis C", The METAVIR Cooperative Study Group, *Hepatology* 24 (August (2)) (1996) 289–293.
- [12] Ali, M., Khedr, A. and Sultan, T., "Enhancement Business Intelligence Applications by applying Multi-Criteria Decision Making", Master Thesis, Faculty of Computers and Information, Helwan University, Egypt, 2012.
- [13] Sabry, S., Sultan, T., and Khedr, A., "Biochemical Markers of Fibrosis for Chronic Liver Disease: Data mining-based Approach", Master Thesis, Faculty of Computers and Information, Helwan University, Egypt, 2012.