# Medical Predictions System

## Doina Drăgulescu

Politehnica University of Timisoara
RO-300222 Mihai Viteazu Blvd.
ddrag@cmpicsu.upt.ro


## Adriana Albu

Politehnica University of Timisoara
RO-300223 Vasile Parvan Blvd.
adriana.albu@aut.upt.ro

*Abstract: Health has a strong impact upon all activities and human experts must have the ability to decide, in any circumstances, what is the illness level of a patient, which is the adequate treatment and which will be the evolution of the patient during the treatment. But medical decision making may be a very difficult activity. There are a lot of applications in artificial intelligence domain that try to help human experts offering solutions for a problem. This paper describes an expert system developed in order to make some predictions regarding the hepatitis infection.*

*Keywords: hepatitis infection, expert system, logical inference, statistical inference, artificial neural networks*

# 1    Introduction

Medical domain is characterized, like many other domains, by an exponential evolution of the knowledge. There are a lot of tools which try to reduce the risk of error apparition in medical life. Diagnosis has a very important role here. It is the first step from a set of therapeutic actions; an error at this level can have dramatic consequences.

The presence of technology in diagnosis phase is welcome because of its advantages: pragmatism, repeatability, efficiency, immunity toward perturbation factors that are specific to human beings (fatigue, stress, diminished attention). The technology doesn't replace human experts in this point of medical assistance; it only tries to help them, implementing systems that are able to select or to generate data which are relevant for the physicians.

The system presented here belongs to this context. It is made using the main two branches of artificial intelligence:

— the traditional one, represented by expert systems (based on logical and statistical inference);

— the connexionist one, where the most common forms used are artificial neural networks.

The goal of the system is to offer predictions about patients infected with hepatitis virus. Hepatitis is one of the principal causes for liver cancer. A correct diagnosis and an adequate treatment could reduce the risks of liver cancer apparition.

The first step is to decide, using logical inference, what type of hepatitis virus is present. There are three possibilities:
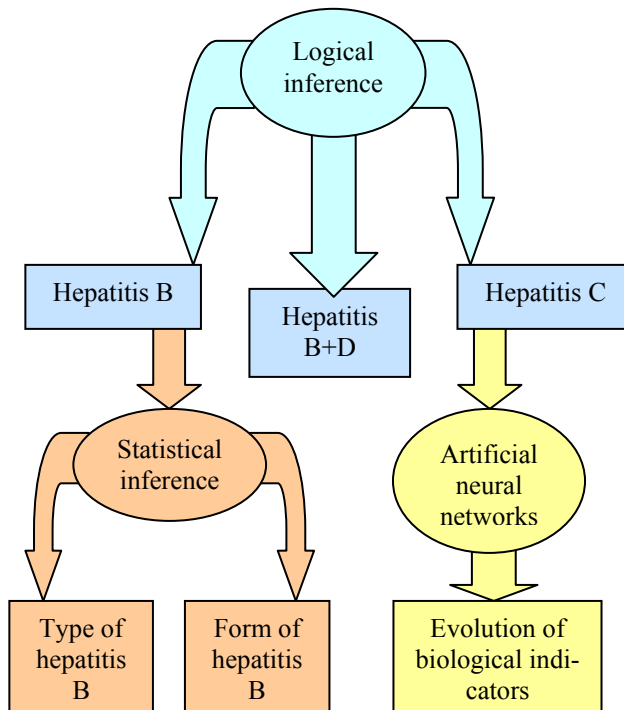
— hepatitis B

— hepatitis B+D

— hepatitis C.

Figure 1
The structure of the system

After the type of hepatitis was set, it is necessary to find some more information about this. There are many forms of hepatitis B. The system described in this paper will decide, based on statistical inference, which one is possible to appear for a patient. If the disease is hepatitis C, it will be important to predict the treatment response and the evolution of laboratory analysis during the treatment, because hepatitis C has a very expensive treatment and severe side effects can often appear. Artificial neural networks will be used in order to do the predictions regarding hepatitis C. Fig. 1 is a schematic description of the system presented in this paper. The application is implemented in MATLAB 7.0, which is a high-performance language and integrates computation, visualization, and programming in a very attractive environment.

Of course such a system will never replace human experts. A tool made to suggest a decision is able to extract information from other solved cases so it can obtain experience and can also take into consideration the results of the last researches, but won't be able to replace the most important factor in decision making: human judgment [1]. Therefore, the final decision has to be made by a human expert. These systems are created only to suggest a solution.

## 2    Methods

*A  Expert Systems*

There are two main possibilities of implementing expert systems: by logical inference and by statistical inference. Both of them were used in this system, in order to make some predictions regarding the hepatitis diagnosis and the evolution of an infected patient.

*Logical Inference*

The logical inference could be used in medicine to build expert systems that will produce a diagnosis starting from a set of premises. An expert system implements human reasoning and it needs some rules to make it possible. This type of system is also called rules based expert system and it is the most used system for implementing medical diagnosis [2]. It has a graph structure and a chain logical evaluation is applied on this structure. Such an expert system could be easy to implement and also very easy to use for a non-engineer because its rules are similarly with the natural medical language.

For hepatitis diagnosis it is necessary to specify which are the factors that define different types of hepatitis. After that, the rules for the expert system can be drawn.

There is a set of markers that have to be analyzed in order to decide what type of hepatitis is present in a patient organism. These markers are described in Table I.

Table I

The markers for the hepatitis diagnosis

| Marker | Value | Name |
|---|---|---|
| AgHBs | Positive | M1 |
| AgHBs | Negative | M2 |
| anti – VHD | Negative | M3 |
| anti – VHD | Positive | M4 |
| anti – VHC | Positive | M5 |

There are considered three possibilities: hepatitis B virus, hepatitis B+D virus and hepatitis C virus. The logical model consists of the following rules, which are created using the markers that appear in Table I:

*R1: If M1 and M3 then B*

*R2: If M1 and M4 then B+D*

*R3: If M2 and M5 then C*

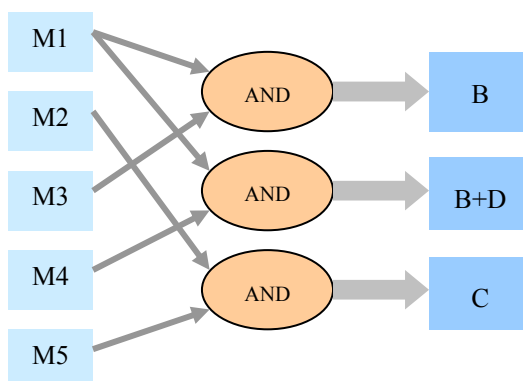Fig. 2 presents the expert system built using these rules.



Figure 2

Rules based expert system for hepatitis diagnosis

This type of expert system is easy to be implemented for simple rules like 'logical premises → conclusion', but it is not suitable to use logical inference for huge amounts of connected knowledge because the graph becomes too complex. Frequently, it is hard to express the rules for the system and also the translation of implicit knowledge into explicit rules would lead to loss and distortion of information content [3]. On the other hand, the tree structure of rule-based relationships becomes too complex if new levels of knowledge are added. For example, there are many types of hepatitis B and if the system described before has to decide between these types, it will be difficult to implement it.

*Statistical Inference*

The statistical inference is an alternative to logical inference and offers a lot of methods that use information of a sample, to learn about population characteristics and to provide some conclusions or decisions. A problem that must be taken into consideration is linked to the fact that inferences are done based on the information contained in a sample, which is only a part of the whole population. From this point of view it is necessary to indicate the precision of the results. The probability plays an important role, being used to define the quality of an affirmation, to measure the uncertainty or to describe the chance for an event to happen.

In this area, the most frequently used method is the *Bayes's theorem*, which sets a probabilistic value for each considered output (disease, if the system is applied in medical diagnosis). Bayesian networks have an important area of applicability in the entire field of artificial intelligence, setting a posterior probability when prior probability is known [4].

Bayes's theorem suggests that probabilities can be improved with new information (Fig. 3). The analysis starts with the prior probabilities (preceding the experience) for the interesting events. Then it is used a supplementary information from a sample, a test, a report or from other sources, information that affects the probability of the events. The prior probability will be revised using this new information and the result will be the posterior probability (after the experience and based on the experience). Bayes's theorem is an easy way to find the posterior probability.
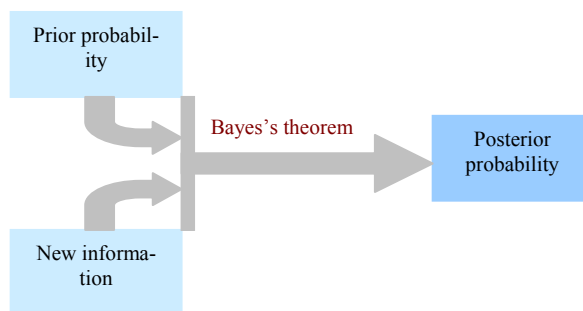


Figure 3
Bayes's theorem

There are three evolutional types of hepatitis B (usual, with relapses and with decompensations) and six grades of disease (easy, medium, grave, prolonged, cholestatic and comatose). It is very useful to have an expert system that can predict, using symptoms and laboratory test results, what type and what form of hepatitis B is present for a new patient. Bayes's theorem [5] will be used to build such a system. It needs a database with symptoms for a number of patients ($\Omega$ - statistical population) that have associated a final diagnosis set. In this application was used a database with over 150 patients with hepatitis B virus infection.

Bayes's theorem is a formula with conditioned probabilities. If it is applied in medical diagnosis, its form is:

$$p(D_k \mid S) = \frac{p(S \mid D_k) \cdot p(D_k)}{p(S)} \tag{1}$$

where $D_k$ is a disease and $S$ a set of symptoms. Using the theorem it can be calculated, for a patient, the probability of appearance for each disease $D_k$ when the set of symptoms $S$ is present.

$p(D_k)$ is easy to find because the frequency of apparition of disease $\delta_k=1$ in statistical population $\Omega$ is known:

$$p(D_k) = \frac{card\{x \in \Omega \mid \delta_k(x) = 1\}}{card\Omega} = \frac{cardD_k}{card\Omega} \tag{2}$$

$p(S \mid D_k)$ can be calculated if the considered symptoms are conditioned independents for a disease $\delta_k$:

$$p(S \mid D_k) = \prod_{i=1}^{n} p(\sigma_i \mid D_k) \tag{3}$$

where $\sigma_i$ is a symptom,

$$p(\sigma_i \mid D_k) = \frac{p(\sigma_i, D_k)}{p(D_k)} \tag{4}$$

and:

$$p(\sigma_i, D_k) = \frac{card\{x \in \Omega \delta_k(x) = 1, \sigma_i(x) = 1\}}{card\Omega}$$
$$= \frac{card\{D_k \cap S_i\}}{card\Omega} \tag{5}$$

$p(S)$ is hard to be determined. If it is supposed that a patient suffers of only one disease at a moment, then the following formula could be used:

$$p(S) = \sum_{j=1}^{m} p(S \mid D_j) \cdot p(D_j) \tag{6}$$

where $j$ is an index of all investigated diseases $\delta_1, \delta_2, \ldots, \delta_m$.

The Bayes's theorem becomes:

$$p(D_k \mid S) = \frac{p(D_k) \cdot \prod_{i=1}^{n} p(\sigma_i \mid D_k)}{\sum_{j=1}^{m} \left[ p(D_j) \cdot \prod_{i=1}^{n} p(\sigma_i \mid D_j) \right]} \qquad (7)$$

with $k = 1,..., m$.

This formula will be applied for each evolutional type and each form of hepatitis B disease, offering for each one a plausibility score.

Such an expert system could be successfully used if it is developed for mutual exclusive diseases and independent symptoms. But sometimes these restrictions cannot be accomplished because there are situations when some symptoms have the same cause (being connected) and a patient can suffer of more than one disease. It was also observed that Bayes's theorem needs an excessive calculation time if statistical population $\Omega$ is very large. In order to avoid these problems, two other statistical algorithms were implemented: Aitken's formula and Logistic model.

_Aitken's formula_ [5] is an alternative for equation (3) (which is the most time consumer in Bayes's theorem). The probability $p(S \mid D_k)$ can be quickly found if this formula is used:

$$p(S \mid D_k) = \frac{1}{T} \sum_{t=1}^{T} \lambda_\delta^{n-st} \cdot (1 - \lambda_\delta)^{st}, \quad k = 1,...,m \qquad (8)$$

where: $m$ – the number of considered diseases;

   $T$ – total number of patients;

   $\lambda_\delta$ – smoothing factor for the disease $\delta$ ($0.5 \le \lambda_\delta \le 1$);

   s$t$ – Hamming distance between the vector of new patient's symptoms $S = (S_1, S_2, \ldots, S_n)$ and the vector of symptoms of the patient $t$ from the database $S^t = (S_1^t, S_2^t, \ldots, S_n^t)$.

The Hamming distance derives from the Minkovski formula. If all the elements $S_i$ and $S_i^t$ (i=1, …, n) are binary codified, than the Hamming distance is the number of elements that are different in S and $S^t$:

$$d_{Hamming} = \sum_{i=1}^{n} XOR(S_i, S_i^t) \qquad (9)$$

_Logistic model_ [5] is a solution to the problem of mutual exclusive diseases which appears in Bayes's theorem. It starts with the notion of anti-probability:

$$o( E ) = \frac{p( E )}{p( \overline{E} )} = \frac{p( E )}{1 - p( E )} \tag{10}$$

and conditioned anti-probability:

$$o( E \mid F ) = \frac{p( E \mid F )}{p( \overline{E} \mid F )} \tag{11}$$

From (10) and (11), where *E* and *F* are two events, can be written equations (12) and (13):

$$p( E ) = \frac{o( E )}{1 + o( E )} \tag{12}$$

$$p( E \mid F ) = \frac{o( E \mid F )}{1 + o( E \mid F )} \tag{13}$$

It is easier to calculate o(E|F) than p(E|F). Logistic discrimination will be used in order to find the logarithm of the anti-probability of disease $D_k$ conditioned by the vector *S*:

$$ln\, o( D_k \mid S = s ) = w_{0k} + \sum_{i=1}^{n} w_{ik} \cdot sign( \sigma_i ) \tag{14}$$

where: *n* – the number of symptoms;

   *m* – the number of diseases;

   *k* = 1, …, m;

   $w_i$ – are called 'weights' and they are calculated with the equations (15) and (16):

$$w_{0k} = ln\, o( D_k ) \tag{15}$$

$$w_{ik} = ln \frac{p( \sigma_i \mid D_k )}{p( \sigma_i \mid \overline{D_k} )} \tag{16}$$

For the patient that is diagnosed it is analyzed the list of symptoms and it is calculated for each symptom $\sigma_i$ the value of the function *signum*, using the expression (17):

$$sign( \sigma_i ) = \begin{cases} -1, if\ \sigma_i = 0 \\ 1, if\ \sigma_i = 1 \end{cases}, \quad i = 1,...,n \tag{17}$$

At the end, the probability of apparition for each disease $D_k$ when a set of symptoms *S* is present can be found:

$$p(D_k \mid S) = \frac{e^{ln\,o(D_k\mid S)}}{1 + e^{ln\,o(D_k\mid S)}} \tag{18}$$

This is exactly the desired result.

*B Artificial Neural Networks*

There are a lot of cases when is not possible to implement human intelligence with expert systems. This is the reason why artificial neural networks have been developed. The initial idea was that in order to reproduce human intelligence, it would be necessary to build systems with a similar architecture [6].

Artificial neural networks are developed based on brain structure, representing a simplified mathematical model of central nervous system. Like the brain, artificial neural networks can recognize patterns, manage data, and, most important, learn [7]. They are made by artificial neurons, which implement the essence of biological neuron.

In this system, artificial neural networks are used in order to make some predictions regarding the treatment response for a patient infected with hepatitis C virus. Hepatitis C is a serious and frequent disease and its evolution has to be carefully overseen during the treatment. Even the efficiency of the hepatitis C treatment improves continuously, the burden of this infection will remain a major issue for the next several decades.

The patients from this study (almost 200) have been kept under observation for 12 months to establish the treatment's influence on the evolution of four biological indicators (TGP, TGO, GGT, and ARN VHC). Three different treatment schemes have been instituted:

— Simple Interferon (IFN);

— Peg interferon α-2a;

— Peg interferon α-2b.

The system offers for each evaluated biological indicator predictions regarding the next 12 months evolution, indicating its growing tendency, its stabilizing or decreasing tendency. It was developed using feed-forward neural networks with back-propagation learning algorithm. Its architecture is in fact a network of neural networks. Each neural network has a layer of 10 hidden neurons, a single output unit and a variable number of inputs.

For each of the four biological indicators that have been studied, there are four layers of neural networks. The networks on the first layer receive as inputs: patient's age, sex, location (rural/urban), treatment scheme, Knodell score, hepatic fibrosis score and value of the parameter for which the prediction is made, at the initial moment (before the treatment starts). These networks have as output the value of the biological parameter at 3 months. On the following layers the net-

works have the same structure as the first layer ones, but they have in addition, as inputs, the outputs of the networks on the former layers; therefore, the networks on the last layer will have not 7 inputs (as the networks on the first layer) but 10 (the initial inputs and the values of biological indicators at 3, 6, and 9 months).

The advantage of this architecture is that the input data are processed separate for each biological indicator. The disadvantage is that the errors are propagated through the system because the results of the networks from the first level (together with their errors) are used in the following levels. But this disadvantage can be minimized by learning process.

# 3   The Expert System

The application has a complex structure, analyzing information connected to the apparition of the hepatitis infection, its evolution, the antecedents, the symptoms, the results of the laboratory tests, and the evolution of some specific biological indicators during the treatment. It develops a multifunctional database and implements an expert system used in order to diagnose different types of hepatitis and to realize some predictions regarding the evolution of the patient and the response to the treatment. The system uses two major components (an inference machine and an architecture of neural networks) that operate on the multifunctional database (Fig. 4). It has an interdisciplinary character and fulfils the requirements of a system used in medical diagnosis and prediction.
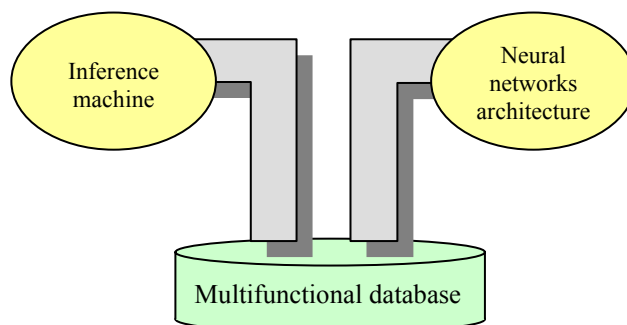


Figure 4
The configuration of the expert system

First of all, the system offers the possibility to diagnose the most frequent hepatitis types: B, B+D and C. Logical inference is used in order to do this. The result can be seen in one of the applications interfaces (Fig. 5). The user has to set the values of the markers that determine which is the hepatitis type. After that, on the bases of the rules described in section II, the result is displayed.
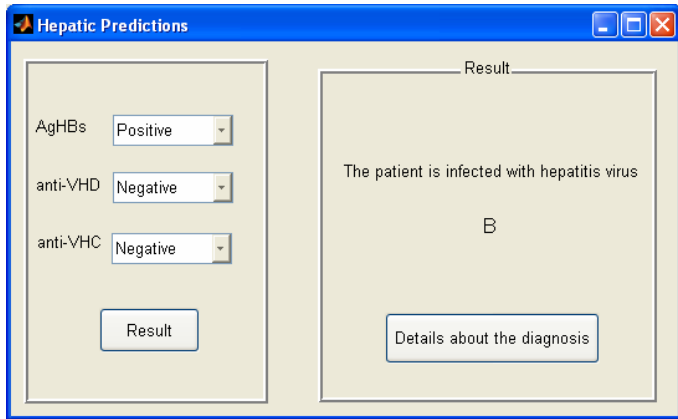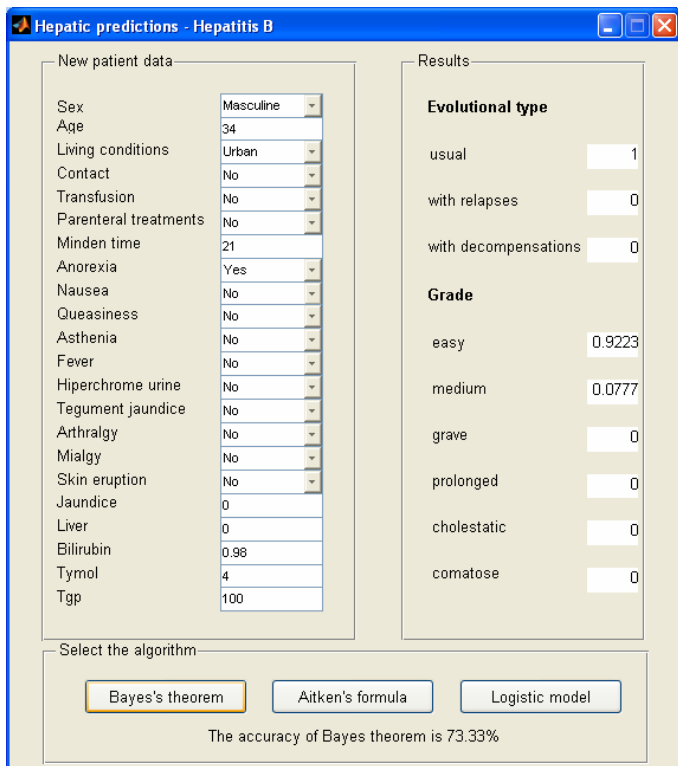
Figure 5
Hepatitis diagnosis



Figure 6
Hepatitis B – Predictions regarding evolutional type and form

For example, if AgHBs is positive, anti-VHD is negative and anti-VHC is also negative, then the patient is infected with hepatitis B virus. If the human expert needs more predictions regarding the diagnosis, than he can use the other two branches of the application. The button 'Details about the diagnosis' from Fig. 5 will lead to apparition of Fig. 6 if the diagnosis is hepatitis B or Fig. 7 if patient is infected with hepatitis C virus.

For hepatitis B is developed an expert system based on statistical inference. The user must set the characteristics of the patient: sex, age, living conditions, symptoms, and the results of laboratory tests. After that, he will choose one of the three implemented algorithms (Bayes's theorem, Aitken's formula, or Logistic model) and the plausibility scores for each evolutional type and grade of hepatitis B are calculated (as can be seen in the right part of Fig. 6).

These statistical algorithms are using a part of the multifunctional database: 165 patients infected with hepatitis B virus. The data which describe medical status of these patients were collected from Clinical Hospital of Infectious Diseases No. 4 'Victor BABES', Timisoara.

Fig. 7 is the user interface for hepatitis C predictions. The user has to choose a range regarding the age of the patient, the sex, the location where the patient lives (rural/urban), the treatment (IFN, Peg interferon α-2a or Peg interferon α-2b) and has to introduce the values of the Knodell score and of the fibrosis score. It is also necessary to introduce the values of the biological indicators before the treatment. The system will predict the evolution of the biological indicators depending on the treatment. Looking at the predicted tendency of the biological indicators during the treatment, a physician can estimate if the patient will respond to a treatment or not.



Figure 7
Hepatitis C - The prediction of biological indicators evolution

The artificial neural networks, which were used in order to do the predictions regarding the patient's treatment response, have the ability of learning and they need some data about a lot of patients (symptoms, laboratory tests, characteristics, etc.). All these are stored in another part of multifunctional database, which contain almost 200 patients infected with hepatitis C virus. These real data were collected from Country Clinical Emergency Hospital, Timisoara.

**Conclusions**

This paper tried to evidence some important aspects connected to medical decision making. Therefore, the system presented here is made from three important parts. First of all, logical inference is used to decide what type of hepatitis virus is present for a new patient. The possibilities are B, B+D and C. After that, the second part of the system will be used to see what will be the type and the grade of hepatitis B (if the patient is infected with hepatitis B virus). This branch of the system is developed using methods from statistical inference. The third part of the system is made for the patients infected with hepatitis C virus and it predicts the biological parameters evolution during the treatment using artificial neural networks.

The hepatitis is a serious disease, its treatment is expensive and severe side effects can appear very often. Therefore, it is important to set a correct diagnosis and to identify those patients who most probably can react to the treatment, so that the others can be protected from a treatment with no benefits. That's for what the use of such a system can support the physicians' decisions.

**References**

[1]     V. Podgorelec, P. Kokol: Self-Adapting Evolutionary Decision Support Model, in *Proceedings of the IEEE International Symposium on Industrial Electronics*, 1999, Vol. 3, pp. 1484-1489

[2]     D. Petrică: Structure of Models for Medical Knowledge Processing, in *Scientific Bulletin of "Politehnica" University Timişoara, Romania, Transactions on Automatic Control and Computer Science*, 2004, Vol. 49 (63), No. 2

[3]     J. M. Zurada: Introduction to Artificial Neural Systems, *West Publishing Company*, United States of America, 1992

[4]     D. Niedermayer: An Introduction to Bayesian Networks and their Contemporary Applications, www.niedermayer.ca/papers/bayesian/bayes.html, 1998

[5]     N. Szirbik, D. Pescaru: Expert Systems, *Politehnica University Timisoara*, 2000

[6]     The        Statistics        Homepage:        Neural        Networks, www.statoftinc.com/textbook/stathome.html

[7]     P. A. Maiellaro, R. Cozzolongo, P. Marino: Artificial Neural Networks for the Prediction of Response to Interferon plus Ribavirin Treatment in Patients with Chronic Hepatitis C, in *Current Pharmaceutical Design*, Vol. 10, pp. 2001-2009, 2004